BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF PROCESS ENGINEERING

ÚSTAV PROCESNÍHO INŽENÝRSTVÍ

Intelligent Energy-Savings and Process Improvement Strategies in Energy-Intensive Industries

MODERNIZACE A OPTIMALIZACE ENERGETICKY NÁROČNÝCH PRŮMYSLOVÝCH PROCESŮ

DOCTORAL THESIS

DIZERTAČNÍ PRÁCE

Author: Ing. Teng Sin Yong, MEng (Chemical Engineering), AMIChemE, G.Eng (BEM)

Supervisor: doc. Ing. Dr. Vítězslav Máša, PhD (Design and Process Engineering)

Co-Supervisor: Prof. DDr. Lam Hon Loong, PhD (Information Technology), PhD (Chemical Engineering)

BRNO 2020

DECLARATION

It is declared that the work presented in this thesis is the candidate's own work. All the sources including candidate's own publication are quoted in this thesis. All the related references are provided at the end of each chapter of this thesis.

Brno : Sin Yong TENG

ABSTRACT

As core processing technologies in energy-intensive industries improve leaps and bounds, existing facilities gradually fall behind in terms of efficiency and productivity. Ultimately, harsh market competition and environmental legislation will force these traditional facilities to stop operations and decommission. Process improvement and retrofit projects are critical in maintaining the operational performance of these traditional facilities. Current approaches for process improvement are mainly Process Integration, Process Optimization and Process Intensification. From a high-level context, analysis from these fields consists of mathematical optimization, accumulated experience, and operational heuristics. These approaches serve good as a basis for process improvement. However, their performance can be further improved with up-to-date computational intelligence. Therefore, **the purpose of this work is to apply advanced artificial intelligence and machine learning techniques into process improvement projects for energy-intensive industrial systems**. The approach taken by this work is a multi-directional approach which tackles this problem from simulation to industrial systems with the following contributions:

(i) Application of machine learning technique, which includes one-shot learning and neuro-evolution for data-driven single unit modelling and optimization.

(ii) Application of dimension reduction (e.g. principle component analysis, deep *autoencoder*) for multiple-unit multiple-objective process optimization.

(iii) Proposition of novel bottleneck tree analysis (BOTA) tool for the purpose of process capacity debottlenecking. An extended BOTA was also proposed to incorporate multi-dimensional problems via data-driven approach.

(iv) Demonstrated effectiveness of Monte-Carlo simulations, neural network and decision trees for decision-making when integrating new process technology in existing processes.

(v) Benchmarked Hierarchical Temporal Memory (HTM) and a dual-mode optimization with multiple predictive tools for real-time operational management.

(vi) Implemented artificial neural networks in the conventional process graph (P-graph) framework.

(vii) Highlight the future of AI and process engineering in biosystems via a commercial-based multi-omics paradigm.

KEYWORDS

Industrial Process Improvement, Data-driven Modelling, Process Optimization, Machine Learning, Industrial Systems, Energy-Intensive Industries, Artificial Intelligence

CITATION

Sin Yong, Teng (2020). Intelligent Energy-Savings and Process Improvement Strategies in Energy-Intensive Industries. Short PhD thesis, Brno: the Brno University of Technology, Faculty of Mechanical Engineering, Institute of Process Engineering.

ACKNOWLEDGEMENT

I would like to express my special gratitude to my supervisor, doc. Ing. Vítězslav Máša and co-supervisor, Prof Hon Loong Lam. They provided me with guidance in both research and in life. My PhD studies would not been possible without assistance from Prof Petr Stehlík, whom was a caring figure to me. I would also like to thank all my colleagues, particularly Dr. Marek Vondra, Dr. Michal Touš, Ing. Eva Konečná, Dr. Vladimír Brummer and Dr. Tomáš Juřena. Without them, my everyday research life would be much duller. Also, I would like to thank my close collaborators Dr. Bing Shen How, Dr. Ir. Wei Dong Leong, Mr. Adrian Chun Minh Loy, and Dr. Sue Lin Ngan. I also thank the reviewers and the committee of my PhD defence for spending valuable time in assessing my work.

Next, I would like to thank my parents, Mr. Koh Jye Teng and Ms. Fong Hun Kok for their unwavering support during my PhD studies. I would also like to thank my sister, Ms. Sin Hui Teng for making many delicious foods when I visit home. Furthermore, I would like to extend my gratitude towards my 5 paternal aunties and my maternal uncle for their care towards me. I would also like to take this opportunity to express my love and thanks to my girlfriend, Dr. Xixia Zhang who stood by me in many hard times throughout my studies.

Lastly, I would like to thank all my friends in Brno who played many sports and fun activities with me, especially Dr. Hua Tan, Ms. Changli Wu, Ms. Lixia Pan, Dr. Tomáš Luks, Mr. Sheng Zuo, Dr. Yee Van Fan, Mr. Bo Hong, Mr. Hon Huin Chin, Mr. Xuechao Wang, Ms. Xuexiu Jia, Ms Limei Gai. Without them, I would probably have no social life. Also, I would also like to thank all people whose name are not listed here due to page constraints, however played a significant role in my life.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

In the era of Big Data, data is becoming increasing important to science, industry, and our everyday life. The defining properties for Big Data are on the larger magnitude of volume, variety, velocity, value, veracity (5V) of data that is collected in today's system. UK famous mathematician and architect of Tesco's club card, Clive (2006), coined the phrase "Data is the new oil", showing the importance of data in this era. McKinsey group (Manyika et al., 2013) estimated that this data transformation would generate 5 to 7 trillion USD of economic impact by year 2025. Lueth et al. (2016) highlighted that advanced applications of data analytics and data sensors in the manufacturing industry can achieve improved revenue, customer satisfaction, product quality, resource planning, insights on customer needs, supply chain optimization, demand forecasting, cost base and compliance with regulation.

Babiceanu and Seker (2016) pointed out the promising outlook of Big Data on the manufacturing industry due to the availability of more sensors within manufacturing facilities. For practical adaptation, Kang et al. (2016) discussed that the global trend for smart manufacturing is rapidly adapted by countries such as Germany, United States, Korea etc. The applications for data-driven manufacturing is shown to be effective in grasping national policies, technology selections, prioritization of target manufacturing articles and forecasting within manufacturing. Although Big Data can be applied into many fields and industries, the focus of this work is specifically for operational phase for energy-intensive processes (See Figure 1.1). The main contribution of this specific scope for smart manufacturing is on the data-driven process improvement by a continuous loop of data extraction, data analysis, calibration and testing, and objective evaluation.

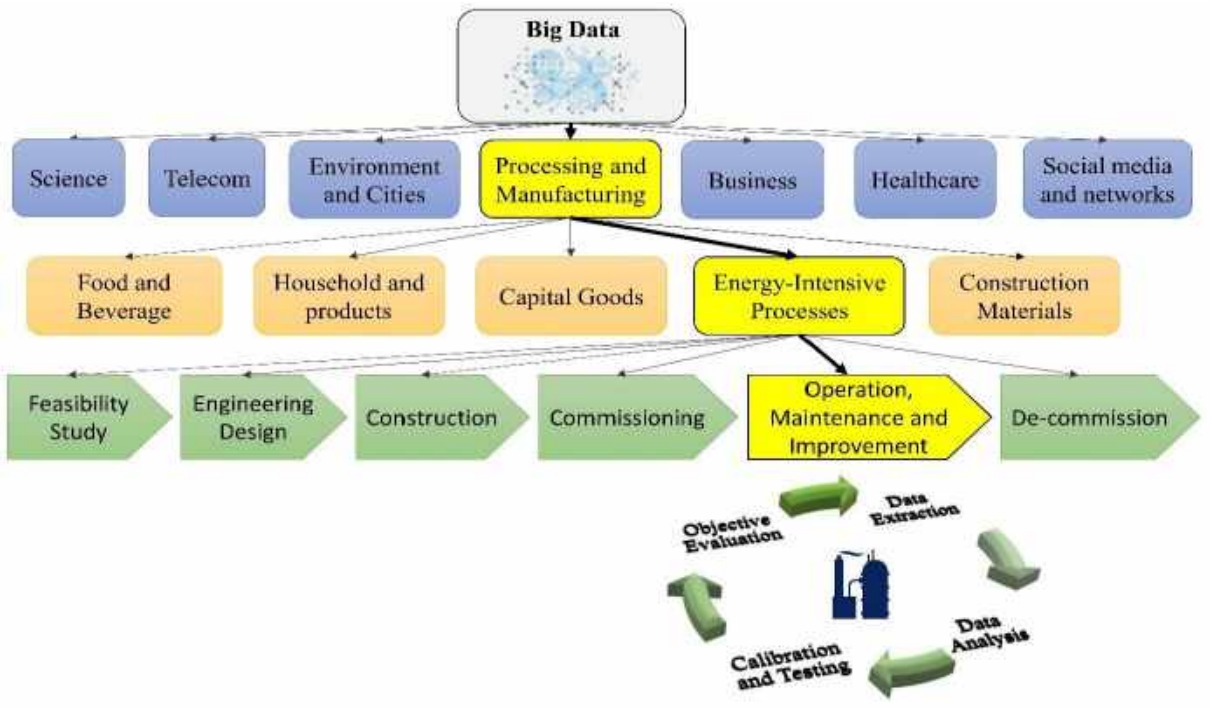


Figure 1.1: Specific industrial category of research

To highlight the importance of smart manufacturing and analytics for the industry, the keywords: "(TITLE-ABS-KEY(("Industrial" OR "Manufacturing" OR "Energy System") AND ("Analytics" OR "Big Data" OR "Machine Learning")))" was queried in SCOPUS database (Swoger, 2019). The growth of this research field is highly promising with exponential increasing number of documents with each year (Refer to Fig 1.2(a)). Moreover, from Figure 1.2(b), the top countries of publication are the world's superpower country, such as United States, China Germany, United Kingdom. This shows that this research direction is very prominent, even on a global scale.

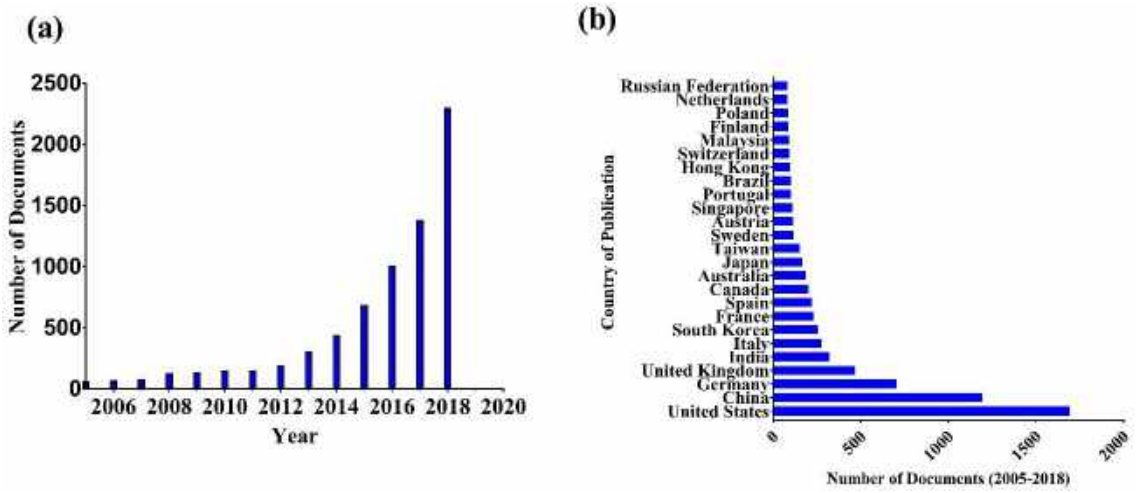


Figure 1.2: Growth of smart manufacturing and analytics from 2005 to 2018: (a) Number of citable documents on SCOPUS (b) Country of publications and number of documents

1.1 Brief Problem Statement and Algorithmic Challenges

Data is widely abundant within a manufacturing facility. Lueth et al. (2016) pointed out that current facilities commonly deploy automatic data collection systems such as Input/Output system (I/O), Programmable Logic Controller (PLC), Supervisory Control and Data Acquisition System (SCADA), Manufacturing Execution System (MES) and Enterprise Resource Planning (ERP). The work also pointed out that the future would only generate a higher velocity and variety of data due to the possibility of cheap and efficient Industrial Internet of Things (IIoT) technologies. Nevertheless, Ćwikła (2014) discussed that at the current situation, industrial processes data acquisition method can be categorised as manual methods, semi-automatic stationary, semi-automatic mobile and automatic (see Figure 1.3). From author's first-handed experience in an oil refinery (Teng et al., 2019) data transfer was also through the form of both automatic and manual, confirming this industrial phenomena. Harding et al. (2006) discussed that data collected within process facilities can be used to improve manufacturing systems, fault detection, engineering design, quality, decision support, Customer Relationship Management (CRM), maintenance, scheduling, layout design, concurrent engineering, plant operation, resource planning and material properties. At this point of time, the industrial data collection systems are well-established for data-driven analysis. However, the data transfer methods are usually not 100 % automatic. Therefore, the volume and velocity of the data highly depends on its variety, making analysis non-straightforward. Hence, the real challenge is at the roots of analysis, the algorithm.

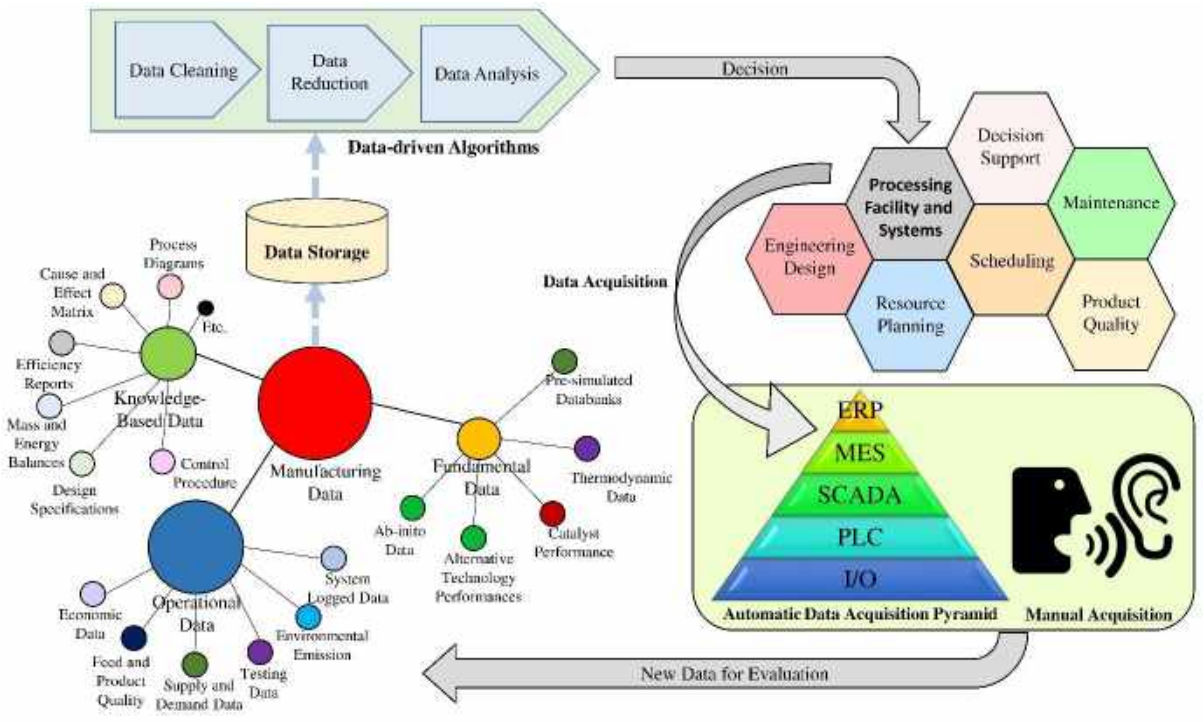


Figure 1.3: Data and means of transfer within a processing facility.

The application of Artificial Intelligence is one of the most prominent breakthroughs in this era of *Big Data*. Artificial Intelligence is commonly defined as the simulation of natural intelligence demonstrated by a machine, also commonly referred to as computational intelligence or machine intelligence. From a high-level perspective, there are three main field within Artificial Intelligence, which are expert systems, metaheuristics, and machine learning (Refer to Figure 4).

Expert systems are computer systems that simulates the decision-making ability of an expert human (Leo Kumar, 2018). The algorithms for expert systems are commonly procedurally coded, however since they are knowledge-based, the inferences were always desirable. These systems and algorithms were particularly popular in the 1980s, however dropped out in 1990s because using hand-coded rules meant that expert systems were severely bottlenecked in achieving new heights (Wagner, 2017). Bell (1985) discussed that there are various chronological problems with expert systems which are shown in Table 1.1.

Table 1.1: A compilation of problems in expert systems according to Bell (1985)

Problem Category	Problems
Human Expert	• The expert is not available • The expert is unable to communicate his idea • The expert is unwilling to communicate his idea • There is no expert
Knowledge Representation	• The knowledge representation is not rich enough • The knowledge representation is too rich • The knowledge representation is incompatible
Testing	• The system is being designed for extraordinary events • The system covers too wide an area • It is too expensive to test the system • It takes too long to test the system • The indicators for correct performance are not known
User Acceptance	• Lack of trust • Too much trust • Legal liability • Job satisfaction
Maintenance	• The system cannot be maintained • Additional knowledge causes major re-writes • Unable to adapt to unexpected changes

Nevertheless, expert systems today (Leo Kumar, 2018) are still very prominent and useful in more direct manufacturing tasks such as interpretation, prediction, control and monitor (Leung et al., 1992). Kumara *et al.* (1986) stated that expert systems consist of 6 functional components, which can be classified as the following:

1. Domain-related facts
2. Domain-related rules for drawing inferences
3. Rules interpretation and application.
4. Rule-ordering mechanism
5. Consistency mechanism
6. Explanation mechanism

Liao (2005) highlighted the importance of combining expert systems with other methodologies with focus on method integration to cope with changes during inference. These research pointed out that although expert system is flawed for some particular tasks, they can be useful when combined with other algorithms for real world applications.

Next, **metaheuristics** are a class of stochastic black box optimization algorithms that have their procedure derived from high-level phenomena in nature or physics. The term “metaheuristics” was introduced by Glover (1986) which means transferring the heuristic from a phenomena to be superimposed on a different problem. A famous example is the Genetic Algorithm (Holland J.H., 1984), where the procedure in evolution of biological genes can be leveraged as an algorithm to optimize black box problems (Whitley, 1994). Metaheuristics can be classified as trajectory or population methods, nature or non-nature-inspired methods, one or multiple neighbourhood methods (Gogna and Tayal, 2013). Some classifications of metaheuristics optimization algorithms can be found in Table 1.2.

Table 1.2: Metaheuristic Algorithms and their classifications

Algorithm	Trajectory	Population	Nature	Non-Nature	One Neighbour	Multi-Neighbour
Genetic Algorithm (Holland J.H., 1984)		✓	✓		✓	
Particle Swarm Optimization (Shi and Eberhart, 2002)		✓	✓		✓	
Tabu Search (Glover, 1986)	✓			✓	✓	
Simulated Annealing (Metropolis et al., 1953)	✓		✓		✓	
Variable Neighbourhood Search (Hansen et al., 2003)	✓			✓		✓

Metaheuristics have been known to be able to obtain very high-quality optimal solutions from a black box function with short computation time. Additionally, metaheuristics algorithm are also very useful for engineering problems due to their scalability and ease of parallelization (Toimil and Gómez, 2017).

On the other hand, **machine learning** (ML), is one of the newer class of artificial intelligent algorithm today. In general, machine learning algorithms can be classified as *Supervised Learning*, *Unsupervised Learning*, and *Reinforcement Learning* (Das et al., 2015). *Supervised learning* is an algorithmic classification in which where examples are shown to algorithms with expect outputs during learning stage. Contrarily, algorithms belong to *Unsupervised Learning* are only shown examples without expect outputs (Love, 2002). *Reinforcement learning* is a

type of machine learning where the agent(s) is placed in an environment with rewards to learn the optimal policy of the environment (Kaelbling et al., 1997). More specifically, the methodology of machine learning can be classified into three distinct directions, which are *Statistical Machine Learning*, *Deep Learning*, and *Reinforcement Learning*. *Statistical Machine Learning* focuses on learning predictions and decision from observed data with focus in statistical methods (Hastie et al., 2005). *Deep learning* is a subset of machine where the primary algorithm of interest is the deep artificial neural network (Lecun et al., 2015). As aforementioned, *Reinforcement Learning* focuses on learning from an environment instead of from observed data (Hessel et al., 2017).

It is worth mentioning that during real-world applications, it is common that a combination of multiple algorithms from different categories algorithms can be used. Namely, the combination of expert systems and machine learning can be classified as Human-In-The-Loop (HITL), tuning of hyperparameters for ML can be done by *Neuroevolution* etc. Figure 1.4 shows some of the major (but not exhaustive) cross-fields that might be implemented for engineering applications.

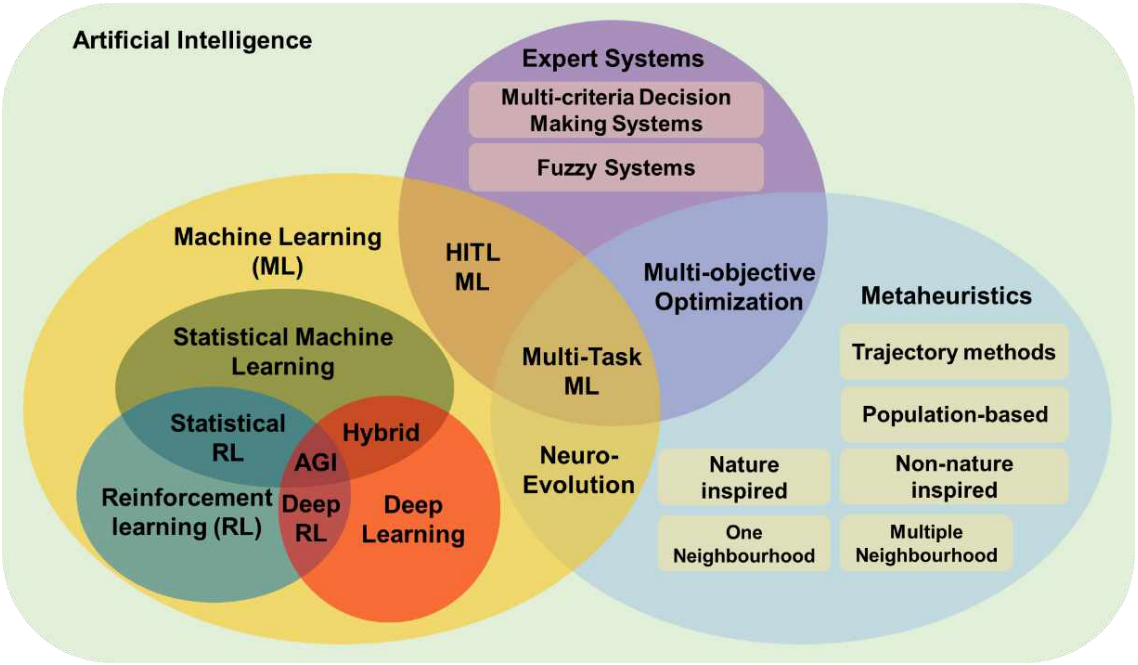


Figure 1.4: Sub-fields and categories for Artificial Intelligence.

Recent state-of-art breakthroughs in various fields (such as face recognition, image reconstruction, protein folding, molecular property predictions etc.) are all using *Artificial*

Intelligent Algorithms (Deng and Yu, 2013). One method that is particularly prominent is the use of *Artificial Neural Networks* (ANN) (Rumelhart et al., 1986). Figure 1.5 provides the illustration for a brief introduction to a commonly deployed *Neural Network*, which is the fully connect neural network.

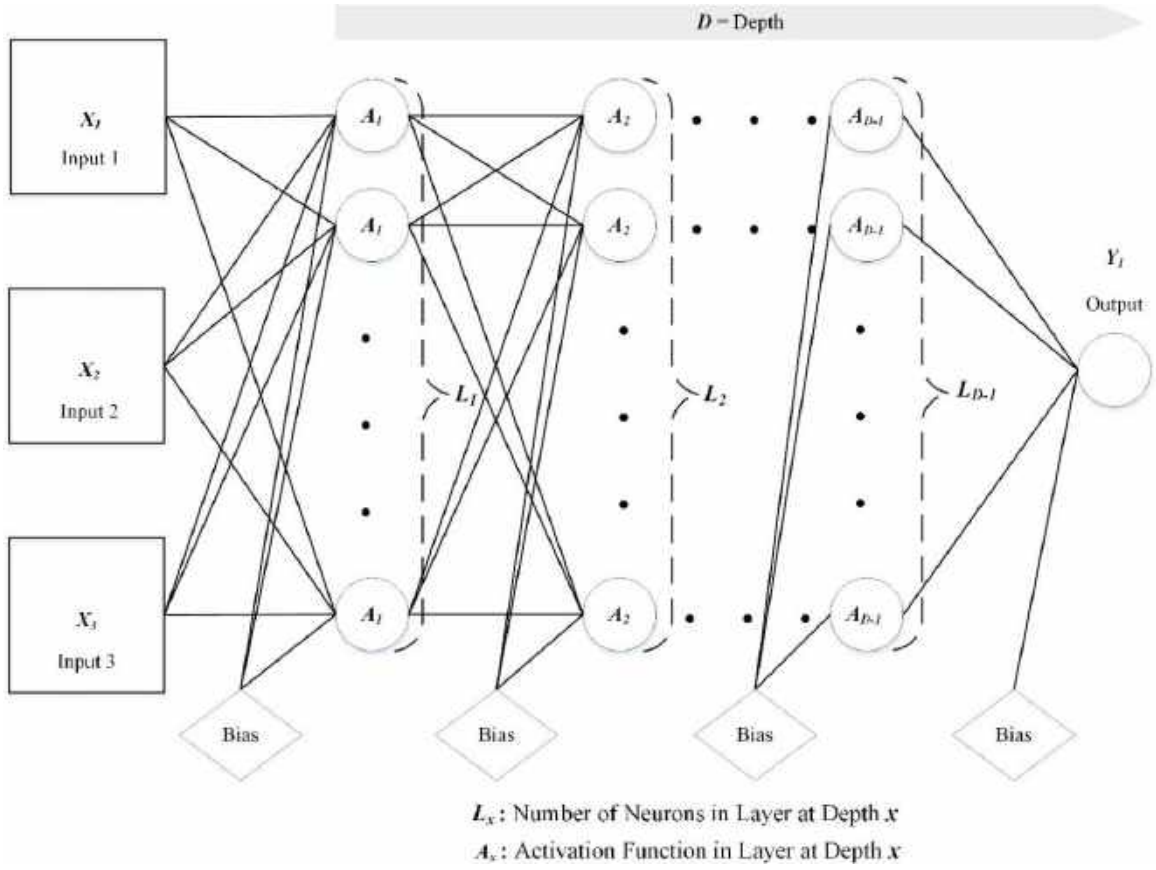


Figure 1.5: Typical structure of a fully connect neural network (modified from Teng et al. (2019))

From Figure 2, a dataset with inputs and output are notated as ***X*** and ***Y*** respectively. The neural network multiplies the data with a weight value (in the line) and sends the data into a layer of activation functions (notated as ***A***), which is a non-linear function (typically sigmoid). Before undergoing the non-linear transformation, an optional bias value can be added to the weighted input. Subsequently, the values from the activation of layer 1 (***L₁***) will undergo the same procedure when moving to layer 2 (***L₂***). Finally, at the last layer, these output values from activation functions are summed and added with a bias value to give the output. If the output is inaccurate, the error is backpropagated into the network, and all the weights and biases will be updated to incrementally reduce the error (Rumelhart et al., 1986). This is the basic concept of how a *Neural Network* operates and learns. For low data availability problems, one-shot

learning methods can be used via Bayesian approaches (Fei-Fei et al., 2006) and more advanced neural networks (Vinyals et al., 2016). Bayesian neural networks also had grown in popularity for the prediction of nonlinear time series (Kocadağlı and Aşıkgil, 2014). A promising direction for Bayesian neural network is from the biological perspective, where Hawkin and Ahmad (2016) proposed a Hierarchical Temporal Memory (HTM) to mimic the neocortex in the brain. The novel biological Bayesian neural network, HTM was then shown to be promising for time-series learning and anomaly detection (Wu et al., 2018).

This concludes this chapter which have discussed the specific field and industry of interest (i.e. energy-intensive industry in general), the potential of data-driven analytics in the industry, the current data acquisition methods, and a brief introduction to terminology of Artificial Intelligence.

1.2 Aims and Scopes

Process improvement (also known as continuous process improvement) is continuous task of identifying, analysing and improving products, systems and processes to meet better standards of quality (Appian, 2019). In this era in which global warming and energy development are of greatest interest, energy-intensive systems and processes are critically important in providing a sustainable future (Lin and Wang, 2015). Process improvement can effectively refine the Key Performance Index (KPI) in energy-intensive industries from sub-process level to the enterprise level (Han and Kang, 2007). Within the context of process system engineering, the approaches for process improvement can be technical classified as debottlenecking (Voudourjs and Consulting, 1996) and optimization (Backx et al., 2000).

Optimization is the procedure for improving the systems or process in order to achieve the best or most efficient use of situation or resource. Mathematically, optimization is to find an extreme point of a certain objective function by manipulating a set of variables, subjected to a set of constraints. These combination of manipulating variables, constraints and objective function is called an optimization model (Azadivar, 1992). Within the context of process improvement, optimization refers to varying process conditions or operational procedures in order to improve certain performances. On the other hand, **debottlenecking** is the procedure of improving the system or process by physically replacing a certain part of the system (Voudourjs and Consulting, 1996). The line between optimization and debottlenecking is often unclear, however the main difference is that debottlenecking requires physical system changes.

Mathematically, debottlenecking can be defined as a condition for the allowance of a sub-problem that deals with physical system changes to a main optimization problem that deals with the changes in operational conditions.

In order to achieve debottlenecking and optimization, the conventional methods include *Process Optimization* (Navalertporn and Afzulpurkar, 2011), *Process Integration* (Linnhoff and Flower, 1985) and *Process Intensification* (Stankiewicz and Moulijn, 2000). As aforementioned, *Process Optimization* refers to optimization of the operational procedures or conditions of processes, commonly by incremental implementation or "once-off" implementation (Sequeira et al., 2002). *Process Integration* is a holistic approach to process design which emphasizes the unity of the process and considers the interactions between different units from the outset (Dunn and El-Halwagi, 2003). Similar perspectives of process integration were also well described in book of Smith (2005) and Klemeš et al. (2014). Process Intensification is the study to improve process units in terms of their unit performance density (Olujić et al., 2003).

The main problem is that most optimization and debottlenecking today are dependent on pure mathematical programming, which highly depend on the accuracy of the mathematical model and are sometimes impractical to be carried out in the real world (Teng et al., 2019). This situation highlights the importance to have scientific methods, case studies, new methods and tools as a combination for effective real world implementation (Máša et al., 2018). In this area, Máša et al. (2016) has provided novel contribution by proposing grey-box approach to combine model-based approach and data-driven approach. Nevertheless, such approach can still be further improved by the incorporation of *Artificial Intelligence* to achieve (i) Better prediction accuracy (ii) Inference under imperfect information; and (iii) Self adaptability of systems.

To provide contribution to this field of research, a preliminary hierarchy of modern process improvement is demonstrated in Figure 1.6. *Artificial Intelligent* algorithms will be used to fortify the performance of data-driven and model driven approaches where the implementation will be carried out using directions of *Process Optimization*, *Process Integration* and *Process Intensification*. The purpose of these methods will ultimately provide solutions to either debottleneck or optimize the process system.

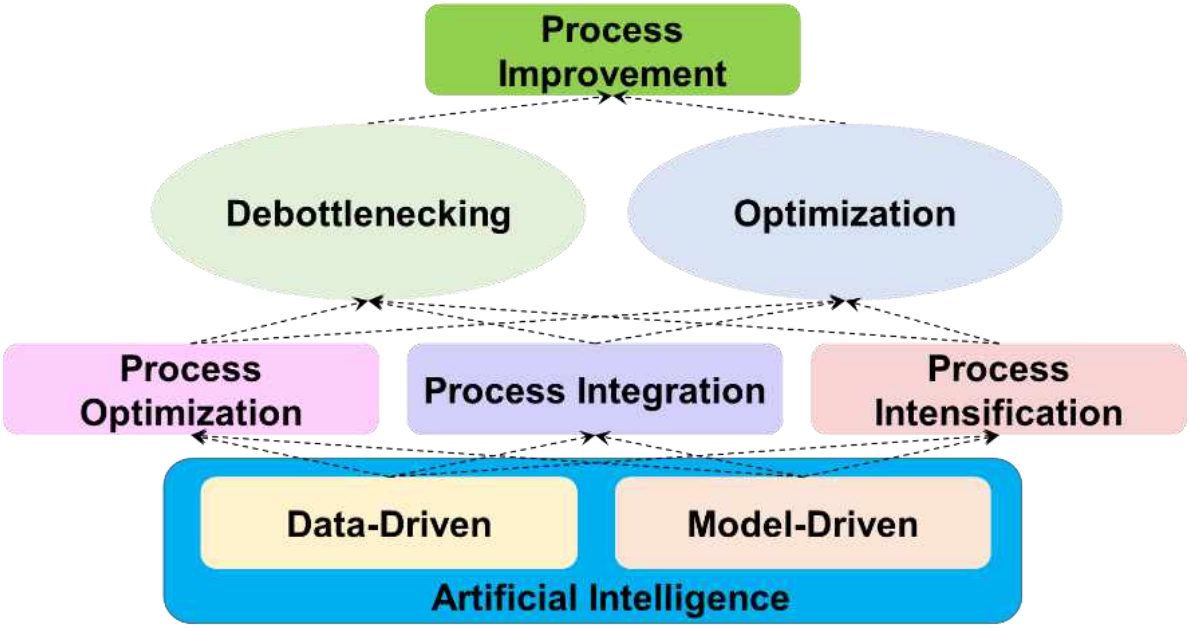


Figure 1.6: Proposed hierarchy of modern process improvement

Author proposes a data-driven approach, which will be studied from three different perspectives of data classifications. These data consist of fundamental data, knowledge-based data, and operational data. Operational data will be obtained from the Supervisory Control and Data Acquisition (SCADA) system of the process system while economic data, equipment data and models contribute to knowledge-based data. The fundamental data consist of data from thermodynamic simulations and *Ab initio* simulations, which may be pre-simulated. From a high-level perspective, the knowledge-based data represents the processing units, fundamental data bank represents the behaviour of materials, and the operational data represents the calibration between real system and simulation models (See Figure 1.7).

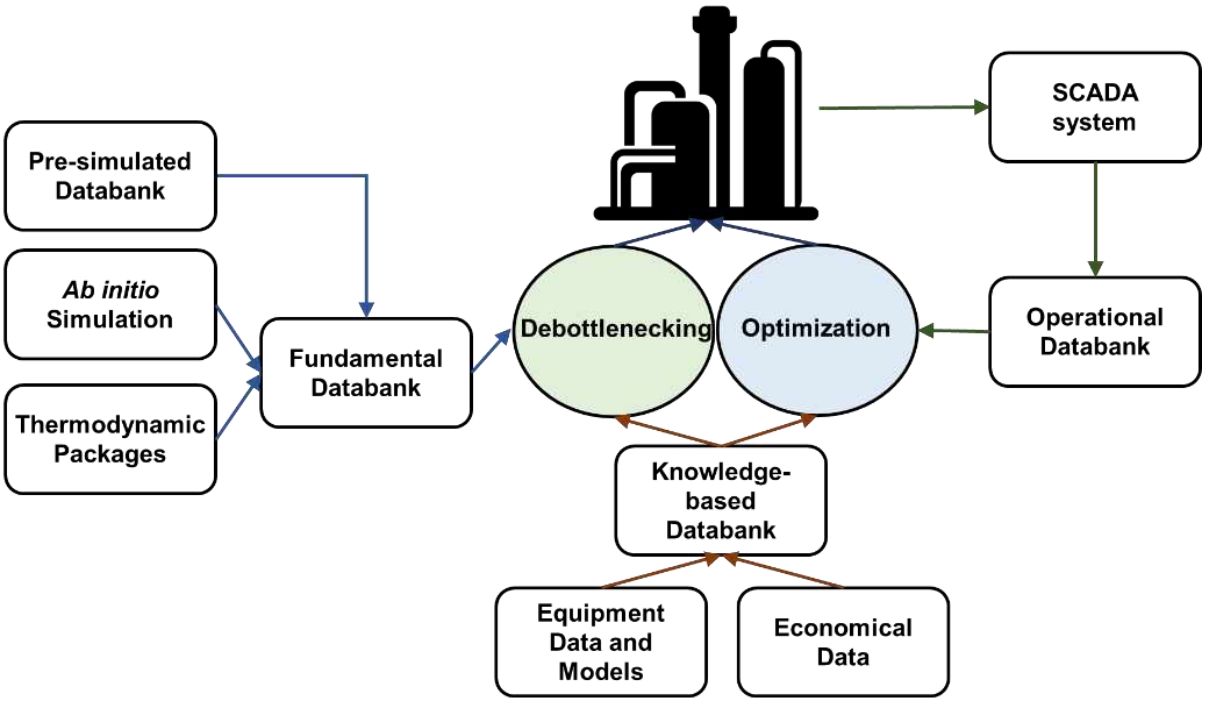


Figure 1.7: Data-driven approach for process improvement

The subsequent chapters will be presented as below:

- Chapter 2 will be literature review on AI, data-driven approaches, and modern infrastructure for process improvement
- Chapter 3 will present a guiding framework for using AI and data-driven integrated approach for process improvement.
- Chapter 4 will present a one-shot learning method for modelling process systems under low data availability.
- Chapter 5 will demonstrate using a neuro-evolution method for single unit process optimization.
- Chapter 6 will discuss and present dimension reduction-based process optimization for practical multi-unit process optimization.
- Chapter 7 will explore and benchmark multiple dimension reduction methods for processing systems.
- Chapter 8 will discuss a novel Bottleneck Tree Analysis (BOTA) method for process capacity debottlenecking.
- Chapter 9 will demonstrate the extension of Bottleneck Tree Analysis via ensemble neural network models to deal with multi-dimensional problems.

- Chapter 10 will demonstrate the use of Monte Carlo simulations, neural network and decision trees for decision-making when considering a new technology.
- Chapter 11 will discuss using predictive analytics, particularly Hierarchical Temporal Memory and a dual mode optimization procedure for real-time management.
- Chapter 12 will demonstrate the novel implementation of an artificial neural network in a traditional process optimization platform – the P-graph framework.
- Chapter 13 will discuss future application of AI and data-driven approaches on microalgae technologies for process improvement.

1.3 Conclusion

In conclusion, this chapter provides a brief introduction to the scope of study that this work is focused on utilizing AI and data-driven analytics for the fields of energy-intensive industries from a process engineering perspective. It is also highlighted that the broad definition of AI consists of expert systems, machine learning and metaheuristics algorithms. Such methods need to be used in an integrative and complex manner to provide high usefulness towards the industry. The basis of this work is on AI and data-driven approaches which considers process optimization, process integration and process intensification for the purpose of optimization and debottlenecking in various industrial facilities. A preliminary approach that was taken into consideration is that the algorithms for optimization and debottlenecking should accept fundamental data, knowledge-based data, and operational data. A short summary of the content of subsequent chapters is also presented in this chapter.

References

Appian, 2019. What is Process Improvement in Organizational Development? [WWW Document]. Process Improv. Organ. Dev.

Azadivar, F., 1992. A tutorial on simulation optimization, in: Proceedings of the 1992 Winter Simulation Conference. pp. 198–204. https://doi.org/10.1145/167293.167332

Babiceanu, R.F., Seker, R., 2016. Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. Comput. Ind. 81, 128–137. https://doi.org/10.1016/j.compind.2016.02.004

Backx, T., Bosgra, O., Marquardt, W., 2000. Integration of Model Predictive Control and Optimization of Processes: Enabling Technology for Market Driven Process Operation. IFAC Proc. Vol. 33, 249–260. https://doi.org/10.1016/s1474-6670(17)38550-6

Bell, M.Z., 1985. Why expert systems fail. J. Oper. Res. Soc. 36, 613–619. https://doi.org/10.1057/jors.1985.106

Clive, H., 2006. Think Big: Britain's data opportunity, in: Proc. ANA Sr. Marketer's Summit. Evanston, IL, USA, p. 2.

Ćwikła, G., 2014. Methods of manufacturing data acquisition for production management - A review. Adv. Mater. Res. 837, 618–623. https://doi.org/10.4028/www.scientific.net/AMR.837.618

Das, S., Dey, A., Pal, A., Roy, N., 2015. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. Int. J. Comput. Appl. 115, 31–41. https://doi.org/10.5120/20182-2402

Deng, L., Yu, D., 2013. Deep Learning Methods and Applications. Found. Trends Signal Process. 7, 197–387. https://doi.org/10.1007/978-981-13-3459-7_3

Dunn, R.F., El-Halwagi, M.M., 2003. Process integration technology review: Background and applications in the chemical process industry. J. Chem. Technol. Biotechnol. 78, 1011–1021. https://doi.org/10.1002/jctb.738

Fei-Fei, L., Fergus, R., Perona, P., 2006. One-shot learning of object categories. IEEE transactions on pattern analysis and machine intelligence, 28(4), 594-611. https://doi.org/10.1109/TPAMI.2006.79

Glover, F., 1986. Future paths for integer programming and links to artificial intelligence. Comput. Oper. Res. 13, 533–549. https://doi.org/10.1016/0305-0548(86)90048-1

Gogna, A., Tayal, A., 2013. Metaheuristics: Review and application. J. Exp. Theor. Artif. Intell. https://doi.org/10.1080/0952813X.2013.782347

Han, K.H., Kang, J.G., 2007. A process-based performance measurement framework for continuous process improvement. Int. J. Ind. Eng. Appl. Pract. 14, 220–228.

Hansen, P., Mladenovic, N., Moreno, J.A., 2003. Variable Neighbourhood Search. Intel. Artif. 7, 1097–1100. https://doi.org/10.4114/ia.v7i19.717

Harding, J.A., Shahbaz, M., Srinivas, Kusiak, A., 2006. Data mining in manufacturing: A review. J. Manuf. Sci. Eng. Trans. ASME 128, 969–976. https://doi.org/10.1115/1.2194554

Hastie, T., Tibshirani, R., Friedman, J., 2005. The Elements of Statistical Learning : Data Mining , Inference and Prediction Probability Theory : The Logic of Science The Fundamentals of Risk Measurement Mathematicians , pure and applied , think there is something weirdly different about. Math. Intell. 27, 83–85. https://doi.org/10.1007/BF02985802

Hawkins, J., Ahmad, S., 2016. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. Frontiers in neural circuits, 10, 23. https://doi.org/10.3389/fncir.2016.00023

Hessel, M., Modayil, J., van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., Silver, D., 2017. Rainbow: Combining Improvements in Deep Reinforcement Learning 3215–3222.

Holland J.H., 1984. Genetic Algorithms and Adaptation. In: Selfridge O.G., Rissland E.L., Arbib M.A. (eds) Adaptive Control of Ill-Defined Systems. NATO Conf. Ser. (II Syst. Sci. 16, 317–333.

Kaelbling, L.P., Littman, M.L., Moore, A.W., 1997. Reinforcement Learning:A Survey. J. Artif. Intell. Res. 4, 237–285.

Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H., Noh, S. Do, 2016. Smart manufacturing: Past research, present findings, and future directions. Int. J. Precis. Eng. Manuf. - Green Technol. 3, 111–128. https://doi.org/10.1007/s40684-016-0015-5

Klemeš, J.J., Varbanov, P.S., Wan Alwi, S.R., Manan, Z.A., 2014. Sustainable Process Integration and Intensification, Walter de Gruyter GmbH & Co KG, Berlin, Germany. https://doi.org/10.1515/9783110535365

Kocadağlı, O., Aşıkgil, B., 2014. Nonlinear time series forecasting with Bayesian neural networks. Expert Systems with Applications 41(15), 6596-6610. https://doi.org/10.1016/j.eswa.2014.04.035

Kumara, S.R., Joshi, A., Kashyapg, R.L., Moodie, C.L., Chang, T.C., 1986. Expert systems in industrial engineering. Int. J. Prod. Res. 24, 1107–1125. https://doi.org/10.1080/00207548608919791

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Leo Kumar, S.P., 2018. Knowledge-based expert system in manufacturing planning: state-of-the-art review. Int. J. Prod. Res. 7543, 1–25. https://doi.org/10.1080/00207543.2018.1424372

Leung, L.C., Miller, W.A., Okogbaa, G., 1992. Evaluation of manufacturing expert systems: Framework and model. Eng. Econ. 37, 293–314. https://doi.org/10.1080/00137919208903076

Liao, S.H., 2005. Expert system methodologies and applications-a decade review from 1995 to 2004. Expert Syst. Appl. 28, 93–103. https://doi.org/10.1016/j.eswa.2004.08.003

Lin, B., Wang, X., 2015. Carbon emissions from energy intensive industry in China: Evidence from the iron & steel industry. Renew. Sustain. Energy Rev. 47, 746–754. https://doi.org/10.1016/j.rser.2015.03.056

Linnhoff, B., Flower, J.R., 1985. Synthesis of Heat Exchanger Networks: I. Systematic Generation of Energy Optimal Network. AIChE J. 13, 107–119. https://doi.org/10.1080/00986448208939960

Love, B.C., 2002. Comparing supervised and unsupervised category learning. Psychon. Bull. Rev. 9, 829–835.

Lueth, K.L., Patsioura, C., Williams, Z.D., Kermani, Z.Z., 2016a. Industrial Analytics 2016/2017: The current state of data analytics usage in industrial companies. IoT Anal. 58.

Lueth, K.L., Patsioura, C., Williams, Z.D., Kermani, Z.Z., 2016b. Industrial Analytics 2016/2017: The current state of data analytics usage in industrial companies, IoT Analytics.

Manyika, J., Chui, M., Bughin, J., 2013. Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Glob. … 163.

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018. Key pillars of successful energy saving projects in small and medium industrial enterprises. Energy 158, 293–304. https://doi.org/10.1016/j.energy.2018.06.018

Máša, V., Touš, M., Pavlas, M., 2016. Using a utility system grey-box model as a support tool for progressive energy management and automation of buildings. Clean Technol. Environ. Policy 18, 195–208. https://doi.org/10.1007/s10098-015-1006-x

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092. https://doi.org/10.1063/1.1699114

Navalertporn, T., Afzulpurkar, N. V., 2011. Optimization of tile manufacturing process using particle swarm optimization. Swarm Evol. Comput. 1, 97–109. https://doi.org/10.1016/j.swevo.2011.05.003

Olujić, Seibert, A.F., Kaibel, B., Jansen, H., Rietfort, T., Zich, E., 2003. Performance characteristics of a new high capacity structured packing. Chem. Eng. Process. 42, 55–60. https://doi.org/10.1016/S0255-2701(02)00019-3

Rumelhart, D.E., Hinton, G.E., Willians, R.J., 1986. Learning Represenations by Back Propogation Errors. Nature 323, 533–536.

Sequeira, S.E., Graells, M., Puigjaner, L., 2002. Real-time evolution for on-line optimization of continuous processes. Ind. Eng. Chem. Res. 41, 1815–1825. https://doi.org/10.1021/ie010464l

Shi, Y., Eberhart, R., 2002. A modified particle swarm optimizer 69–73. https://doi.org/10.1109/icec.1998.699146

Smith, R., 2005. Chemical Process Design and Integration, John Wiley & Sons, Ltd. https://doi.org/10.1529/biophysj.107.124164

Stankiewicz, A.I., Moulijn, J.A., 2000. Process Intensification : Transforming Chemical Engineering. Chem. Eng. Prog. 22–34.

Swoger, B., 2019. Scopus. Libr. J. https://doi.org/10.5260/chara.18.2.52

Teng, Sin Yong, How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019a. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. 225, 359–375. https://doi.org/10.1016/j.jclepro.2019.03.272

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. 225. https://doi.org/10.1016/j.jclepro.2019.03.272

Teng, Sin Yong, Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F., Máša, V., 2019b. Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresour. Technol. 292, 121971. https://doi.org/10.1016/j.biortech.2019.121971

Toimil, D., Gómez, A., 2017. Review of metaheuristics applied to heat exchanger network design. Int. Trans. Oper. Res. 24, 7–26. https://doi.org/10.1111/itor.12296

Vinyals, O., Blundell, C., Lillicrap, T. and Wierstra, D., 2016. Matching networks for one shot learning. In Advances in neural information processing systems, 3630-3638.

Voudourjs, V.T., Consulting, A., 1996. Mathematical programming techniques to debottleneck the supply chain of fine chemical industries. Comput. Chem. Eng. 20, 1269–1274.

Wagner, W.P., 2017. Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies. Expert Syst. Appl. 76, 85–96. https://doi.org/10.1016/j.eswa.2017.01.028

Whitley, D., 1994. A Genetic Algorithm Tutorial by Darrell Whitley. Stat. Comput. 65–85.

Wu, J., Zeng, W. and Yan, F., 2018. Hierarchical Temporal Memory method for time-series-based anomaly detection. Neurocomputing, 273, 535-546.
https://doi.org/10.1016/j.neucom.2017.08.026

CHAPTER 2 LITERATURE REVIEW

This chapter has been peer-reviewed and published in Renewable and Sustainable Energy Reviews.

Teng, S.Y., Touš, M., Leong, W.D., How, B.S., Lam, H.L. and Máša, V., Recent advances on industrial data-driven energy savings: Digital twins and infrastructures. Renewable and Sustainable Energy Reviews, 135, 110208.

Abstract: Data-driven models for industrial process improvement heavily rely on sensor data, experimentation data and knowledge-based data. This work reveals that too much research attention was invested in making data-driven models, as supposed to ensuring the quality of industrial data. Furthermore, the true challenge within the Industry 4.0 is with data communication and infrastructure problems, not so significantly on developing modelling techniques. Current methods and data infrastructures for industrial process improvement were comprehensively reviewed to showcase the potential for a more accurate and effective digital twin-based infrastructure for the industry. With a few more development in enabling technologies such as 5G developments, Internet of Things (IoT) standardization, Artificial Intelligence (AI) and blockchain 3.0 utilization, it is but a matter of time that the industry will transition towards the digital twin-based approach. Global government efforts and policies are already inclining towards leveraging better industrial energy efficiencies and energy savings. This provides a promising future for the development of a digital twin-based energy-saving system in the industry. Foreseeing some potential challenges, this chapter also discusses the importance of symbiosis between researchers and industrialists to transition from traditional industry towards a digital twin-based energy-saving industry. This work serves as a concise guideline for researchers and industrialists who are looking to implement advanced energy-saving systems.

Keywords

Digital Twins, Data-driven Energy Savings, Artificial Intelligence (AI), Blockchain, Internet of Things (IoT), Cyber-physical Production Systems (CPPS)

2.1 Introduction

Throughout the timeline of manufacturing and processing, there were certain incredibly novel technological breakthroughs which flourished the possibilities of a new system, and even, a new manufacturing era. The first occurrence of such ground-breaking technology happened in 1763 where James Watt invented a version of the steam engine which was incredibly fuel-efficient at that time (Wisniak, 2018). This key piece of technology opened the doors towards the first industrial revolution which relied on the principles of the thermodynamic engine and mechanical gear. A reoccurrence of this phenomenon of industrial revolution happened in the 19th centuries where many new technologies, especially electricity, were invented (Atkeson and Kehoe, 2001). The acquisition of electricity and electrical applicants in this time of history allowed for mass production of daily products, giving birth to the start of a new economy. This was the second industrial revolution. Further development of electronics in the next hundred years had advanced both the energy industry and utilization of electronics, leading to the third revolution. Most notably, the transition from oil and gas to utilizing nuclear and bioresources as energy sources (Jänicke and Jacob, 2009) has drastically changed the economy. In the industry, the extensive use of automation also emerged due to the invention of programmable logic control (PLC) and simple robots (Naboni and Paoletti, 2015). Today's enabling elements, which include Artificial Intelligence (AI), advanced robotics, cyber-physical production systems (CPPS), internet of things (IoT) and Big Data (Keliang Zhou et al., 2016), is leading to another new revolution. Industry 4.0, also known as the fourth industrial revolution is the idealization of an economy that produces materials or provide services with highly automatized procedures (Rüßmann et al., 2015).

In advanced countries, such as Germany, France, Japan (Federal Ministry for Economics Affair and Energy, 2019) and China (Keliang Zhou et al., 2016), Industry 4.0 projects related to autonomous processing, operational improvement and energy-savings have been increasing steadily. However, World Economic Forum (World Economic Forum, 2018) reported that only 25 countries within the world are ready to benefit from the changing nature of Industry 4.0, which are (in alphabetical order) Austria, Belgium, Canada, China, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Isreal, Italy, Japan, Korea Republic, Malaysia, Netherlands, Poland, Singapore, Slovenia, Spain, Sweden, Switzerland, United Kingdom and United States. Furthermore, researchers such as Rajnai and Kocsis (2017) had shown that there are significant labour market risks associated with Industry 4.0. Nevertheless, the values derived from transitioning to Industry 4.0 is too much attractive for both the micro- and macro-

perspectives in terms of resource conservation, asset utilization, labour allocation, inventory management, quality improvement, supply-demand matching, time-to-market management and aftersales service (Sung, 2018). In this case, much more developments are required to achieve Industry 4.0 in various parts of the world. In terms of research effort, Preuveneers and Ilie-Zudor (2017) pointed out that more works have to be done on end-to-end production transparency, information management in industrial systems, optimization of industrial processes with Big Data and cloud computing, production-aided with machine learning, human-computer and machine interaction, security threats and regulations. Keliang Zhou et al. (2016) also highlighted that the challenges of transitioning to the Industry 4.0 contain complex aspects of scientific challenges, technological challenges, economic challenges, social challenges, and political challenges. On the production floor level, Weyer et al. (2015) demonstrated the importance of infrastructure standardization with examples related to electro-mechanical standards, production lines, communication standards, control architectures, work stations and integration of superordinate IT systems.

Energy savings is one of the most attractive targets to achieve improved energy efficiencies with the technologies of Industry 4.0 (Bornschlegl et al., 2013). Song and Wang (2018) proposed a data-driven measuring method to consider technological progress for energy saving and emission reduction in the settings of Industry 4.0. An energy management system named Energy Cloud (Sequeira et al., 2014) was also deployed to monitor energy consumption in multiple industrial sites by utilizing Big Data and cloud computing. Lee et al. (2014) argued that deploying a self-aware machine within the context of Industry 4.0 could also reduce processing cost by saving energy consumption. From a project implementation perspective, Oses et al. (2016) proposed using a statistical learning approach to measure and verify energy savings within an industrial plant. The model was able to act as a baseline energy-saving model while reducing uncertainties in real-time. Furthermore, Wang et al. (2016) proposed a four-level architecture for energy-saving operations which includes a physical and sensible manufacturing system, system unit models in virtual space, system production model in virtual space and active energy-saving operation decision model. By utilizing an intelligent information processing approach, Yan et al. (2017) argued that the intelligent factory can improve system reliability from applying predictive maintenance and intelligent energy savings. To demonstrate the importance of industrial energy savings from a regional viewpoint, the European Council emphasizes to reduce projections of primary energy consumptions in 2020 by 20 % (European Commission, 2017). The European Commission also adopted a roadmap

up to 2050 to focus on low carbon economy with a focus on energy efficiency (European Union, 2012). United States Department of Energy also established the Advanced Manufacturing Office (AMO) to improve the energy and material efficiency, productivity, and competitiveness of manufacturers across the industrial sectors (Advanced Manufacturing Office, 2012). To date (as of 2019), AMO has provided more than 1300 industrial partnerships and projects related to energy savings in the United States. In China, the National Development and Reform Commission (NDRC) and the National Energy Administration (NEA) jointly release the 13th Five-Year Plan for Energy Development with focus on optimizing energy systems, reducing energy consumption, promote renewable energy supply, promote efficient energy technology, build fair energy market system, strengthen energy cooperation and achieve energy sharing (International Energy Charter, 2018). Moreover, the Ministry of Power in India (Bureau of Energy Efficiency, 2018) is reported to promote Perform Achieve and Trade Scheme (PAT) which enhances energy savings within energy-intensive industries. In the first PAT cycle, the overall energy saving achieved was 8.67 million TOE (tonnes of oil equivalent), exceeding targets by 30%. In South-East Asia, the Ministry of Energy, Green Technology and Water (Ministry of Energy, 2015) of Malaysia have also allocated an annual budget of MYR 54.3 million (approximately 13 million USD) to improve energy efficiencies of appliances within the country. Evidently, countries around the globe prioritize energy savings and energy efficiency of industrial systems heavily, as it is critical for sustainable development on a macro-perspective.

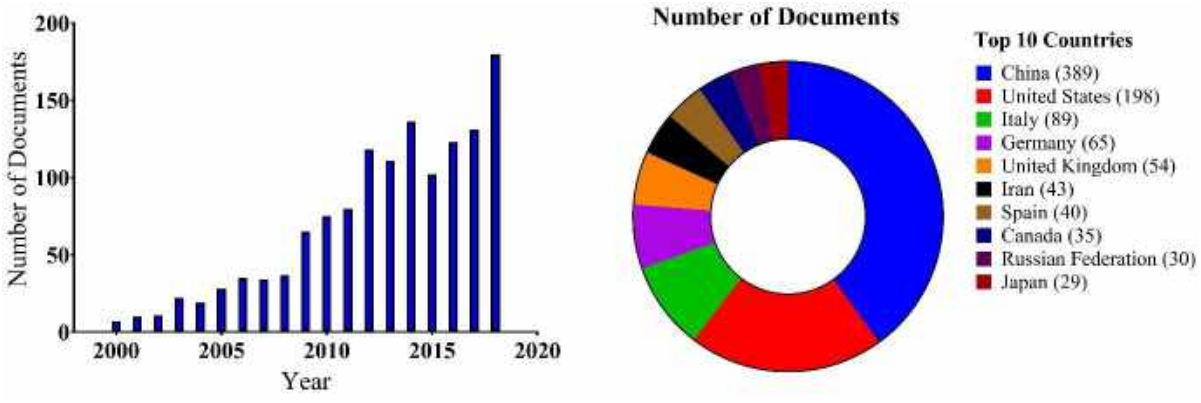
The transition to an energy-efficient Industry 4.0 is inevitable throughout the evolution of mankind. According to PricewaterhouseCoopers (PwC) reports (Koch et al., 2014), data-driven improvements to resource and energy efficiency are expected to have an 18 % increase. The report also mentioned that the data-driven industry is already generating more than 110 billion Euro of additional revenue in Europe. McKinsey & Company (Caylar et al., 2016) estimates that Industry 4.0 would improve productivity in the technical profession by 45 to 55 %. Nevertheless, in terms of realistic energy saving implementation, Máša et al. (2018) demonstrated that the real challenge is in dealing with the data and processing infrastructure in existing facilities. The work also highlighted the main problem within existing firms, which is, most of them have incomplete data acquisition system. Weyer et al. (2015) also agreed that data infrastructure is one of the most challenging aspects for firms to transition towards Industry 4.0. The work proposed to standardize and modularize the data infrastructure within smart production systems. Still, there is a research gap for the development of an efficient "one-

size-fits-all" approach for data-driven process analysis and data acquisition for non-experts or non-professionals. Furthermore, research in this field will be strongly motivated by the (i) significance, timeliness, and contribution towards Industry 4.0, (ii) international interest in providing a more energy-sustainable future, and (iii) government initiatives and policies.

In this thesis chapter, Chapter 2.2 will discuss the current state and challenges with the consideration of the conventional pipeline for data-driven energy saving. Chapter 2.3 will discuss enabling technologies that will accelerate the research field (such as Industrial Internet of Things (IIoT), digital twins and cyber-physical systems, cloud computing, and advanced blockchain technologies). Furthermore, Chapter 2.4 will discuss policies and government initiatives that will provide subsidies or benefits for data-driven energy savings in different countries or regions. Additionally, Chapter 2.5 discusses the difference between the modern and traditional implementation of industrial analytics for energy savings while giving a focus on industrial-academic collaboration. Authors have also provided future directions that may accelerate the research field in Chapter 2.6 while the concluding remark was provided in Chapter 2.7.

2.2 Current State and Challenges

The current situation of utilizing data-driven analytics for the purpose of industrial process improvement is steadily rising throughout the years (See Figure 2.1(a)). From the year 2000 to 2018, the top ten countries that contribute the most to data-driven energy-saving research (in terms of published research documents in SCOPUS) are China, United States, Italy, Germany, United Kingdom, Iran, Spain, Canada, Russian Federation and Japan (See Figure 2.1(b)). With modern manufacturing and production system having an increasing improvement in sensors and data acquisition, data-driven analytics has grown in interests throughout the years within the context of the Industry 4.0.



(a) (b)

Figure 2.1. Research interest of data-driven energy savings in the industry from 2000 to 2018 based on SCOPUS database: (a) Exponentially growing number of research articles (b) Top ten countries with contributing research articles. Search keyword on SCOPUS is "("Energy Savings") AND ("Industrial Process" OR "Manufacturing" OR "Production") AND ("Machine Learning" OR "Data" OR "Artificial Intelligence")".

There are still some undeniable gaps to be addressed by both academic researchers and industrial practitioners for data-driven analytics in the context of Industry 4.0. Kusiak (2017) discussed that these challenges arise in (i) adopting strategies for information management, (ii) improving data collection and utilization, (iii) designing predictive models, (iv) managing with model uncertainties, (v) connecting factories and control processes. Mittal et al. (2018) provided a concise review on the maturity model for aspects of Industry 4.0 showing that there are still many technologies that are not matured for the transition towards full Industry 4.0. In reality, Zhang et al. (2018) discussed that one of the main challenges for data-driven energy savings is the complexity that arises from the variety of energy use across thousands of processes. To deal with this complexity, Shrouf et al. (2017) proposed the use of multi-level energy awareness to process energy data which corresponds to process level, machine level, production line level and production level. In terms of improving the energy efficiency of existing systems, Grueneich (2015) laid out five critical challenges that includes: (i) the ability to support an increasing magnitude of energy efficiency savings, (ii) diversification of sources of energy, (iii) the measuring standards of energy savings must be established, (iv) energy-saving outcome must include carbon reduction framework, (v) variability of energy efficiency must be understood. To successfully overcome such challenges, strategies from technological innovation, energy market, policy framework and agency governance must be adopted.

2.2.1 Pipeline for Data-Driven Analytics

In all cases, data-driven energy-saving procedures can be classified into four steps: (i) data acquisition, (ii) data pre-processing, (iii) modelling and analysis, and (iv) industrial implementation (see Figure 2.2). Even though many researchers with actual industrial experiences suggested that the real challenge of data-driven energy savings is at the stage of data acquisition (Kusiak, 2017; Máša et al., 2018; Shrouf et al., 2017; Y. Zhang et al., 2018), academic researchers at the current state have more research attention towards modelling and analysis. Authors suspect that this is due to the academic ecosystem viewing research related to "modelling and analysis" as trendier and more applicable. With an overload of research interest in the fields of energy modelling and analysis, other parts of the data-driven energy-saving pipeline (i.e. data acquisition, pre-processing, implementation) becomes a significant bottleneck for the field of research. There is a strong need for a concise review of the development of each part of the pipeline to realign the researcher's interest towards the industry's demands.

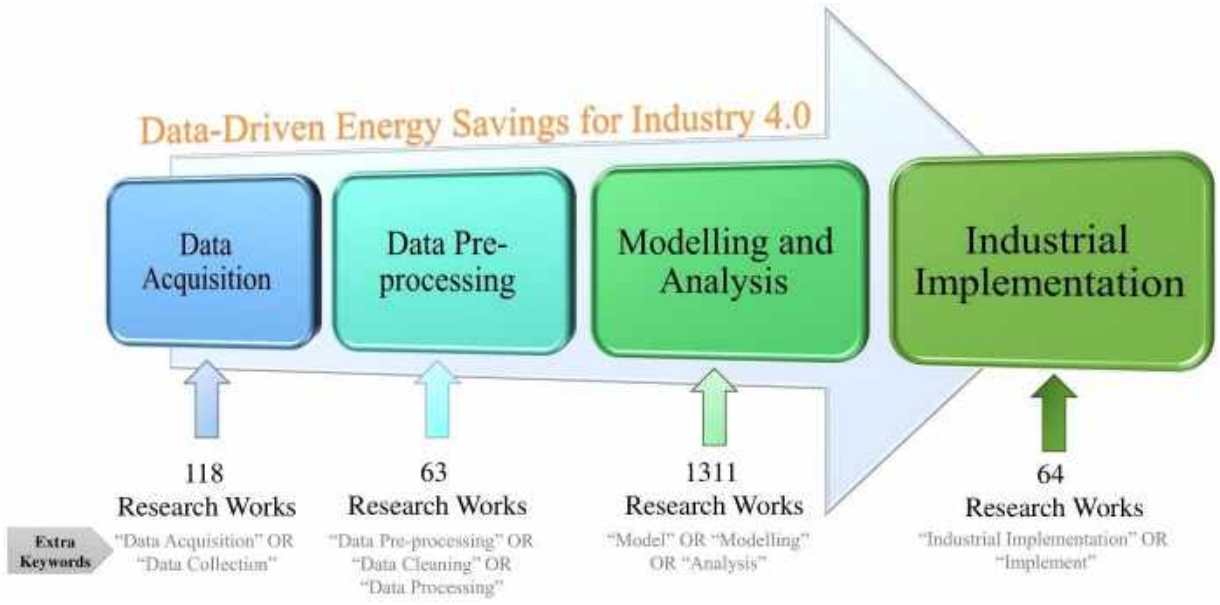


Figure 2.2. The attention of research works in the data-driven energy-saving pipeline

Data collection and acquisition is one of the most important procedure for data-driven energy savings. The quality of the energy-saving implementation is only as good as the quality of the data available. The task of data collection is a complex task which includes sensor selection, communication protocol, information systems, data warehousing, data prioritization and much more unforeseen engineering works (Lu et al., 2005). Work from Abdelaziz et al. (2011)

reviewed that even for audit-based energy-saving procedures, the importance of historical energy-related database cannot be avoided. Additional data is also proposed to be obtained via portable data acquisition tools such as fuel efficiency monitor, clamp-on power meter, thermocouple sensor and data loggers. By considering an enterprise energy-information system, Swords et al. (2008) reported that data collection is important for both energy data and enterprise data. Similarly, the work also emphasized the importance of data collection tools while promoting the use of energy management according to ISO 140001. For energy management aspect, data collection and acquisition are critical as data-driven practices can work continuously and seamlessly to detect inefficient and malfunctioning equipment, optimize energy usage and performances (Backlund et al., 2012). Some important works related to data acquisition can be found in Table 2.1.

Table 2.1: Significant works related to data acquisition and collection

Work	Contribution
Cordeau and Barrington (2010)	Evaluate the performance of data acquisition systems to conduct energy balance analysis in two commercial broiler barns. Discussed that better data acquisition procedure can give better results.
Januteniene et al. (2012)	Used an external data logger for data collection to carry out real-time process optimization using advanced process control (APC) controller to lower overall energy consumption.
Tian et al. (2012)	Proposed a company-level data collection methodology that relied on statistical data of local government, audit reports, and technical reports. Compared technical measures from 44 companies to assess the potential of energy-saving in a Chinese fine chemical industrial park.
Brundage et al. (2013)	Highlighted the importance of Energy Efficiency Performance Indicators that uses real-time production data to identify energy-saving opportunities.
Nunes et al. (2014)	Collected information related to facilities, equipment, technical operations and production processes from multiple refrigeration systems in the fruit and vegetable industry. Performed comparative analysis between various industrial plants of the sector.

Table 2.1 (Cont.): Significant works related to data acquisition and collection

Work	**Contribution**
Nunes et al. (2015)	Adapted a method of utilization of a combination of the industrial production database and separate technical data to gauge the potential energy savings in the dairy food sector.
Abele et al. (2015)	Emphasized the importance of information infrastructure for data collection to achieve energy-efficient production. The work refers to systems such as Programmable Logic Controller (PLC) and Human-Machine Interface (HMI) as prerequisites for energy Key Performance Indicator (KPI) monitoring.
Wei et al. (2016)	Implemented IoT-based communication framework for data collection within the processing facility and utility systems. Developed a complete energy management network to achieve energy savings
Cosgrove et al. (2017)	Collected data for energy savings in a holistic manner. Discussed that industrial energy can be classified as value-added energy, auxiliary energy and indirect energy. Highlighted the importance of lean energy management with consideration of value stream.
Tuo et al. (2018)	Propose the usage of real-time data collection to calculate energy efficiency index through "virtual part" of the machining systems.
Zhang et al. (2019)	Multi-attributed data related to total energy consumption were extracted from historical data to optimize machine scheduling.

Data pre-processing is also known as data cleaning as it is the operation of converting raw data from data collection devices to useful data for analysis. Many data engineers consider the work of this stage as 'data janitor work'. However, the industry proves that preventing "dirty data" usage is crucial for industrial process improvement (Huang et al., 2006). Cai and Zhu (2015) proposed 5 dimensions for the assessment of clean data which covers availability, usability, reliability, relevance, and presentation quality. The work proposed a conceptual idea of utilizing indicators, data quality elements and dimensions for the phase of data pre-processing. For the application of energy-saving, the energy consumption for computing the data cleaning algorithm is often studied by researchers (Deng et al., 2018). On the technical ground, Chu et al. (2016) reviewed that advanced data cleaning procedure should be able to carry out

qualitative error detection, error repairing, and adaptive data cleaning. Nevertheless, the most important purpose for data pre-processing for data-driven energy savings is on the removal of noise containing data, missing data removal and data structure unification (Lenz et al., 2018). Although there is lesser research work that relates to the "data janitor work", its importance is undeniable, and some useful works can be found in Table 2.2.

Table 2.2: Significant works related to data cleaning and pre-processing

Work	Contribution
Mikšovský et al. (2002)	This work proposed a data pre-processing tool called SumatraTT. The software was tested to be able to copy data, format data, calculate new attribute, filter data, report and visualize data on the case study of a water distribution company.
(Huang et al., 2006)	Shown the effectiveness of data pre-processing in providing a data-driven solution for a semiconductor chemical vapour deposition (CVD) process. The data pre-processing procedure focuses on data reduction, treating missing values, noise reduction and anomaly detection.
Deng et al. (2018)	Demonstrated that low energy in-network data cleansing algorithm can be used to pre-process data for Cyber-Physical Production System (CPPS) models.
Lenz et al. (2018)	Proposed a holistic approach to deal with manufacturing unit data as a whole system. Reduced the software and computational effort required to perform data cleaning to remove noise and deal with missing data.
Dai et al. (2019)	Pointed out that data pre-processing is crucial in noise reduction, missing value interpolation and inconsistency mitigation. Highlight the difficulties for data pre-processing, compression, and storage.

The core of the data-driven energy-saving pipeline is the procedure for modelling and analysis. Data-driven modelling has received very much attention in recent years due to its unparallel advantages in adaptability, accuracy, predictivity and simplicity (Leong et al., 2019b). Within this field, the rise of AI, digital twins and cyber-physical systems (CPS) have been pushing the boundaries of its possibilities (Ji et al., 2019; Kusiak, 2017) (this will be further discussed in

Chapter 2.3.2). Data-driven methods have proven to be much superior to traditional engineering correlations and mathematical modelling methods. For example, data-driven modelling can be used to reduce the number of processes of an oil refinery with minimal background studies while mathematical programming would require thousands of equations and variable sets for optimization (Teng et al., 2019b). Leong et al. (2019) also demonstrated that data-driven modelling methods can easily adapt to the uncertainties and changes within the real world, which traditional methods generally fail to achieve (Kaile Zhou et al., 2016a). Neural networks were also recently grown in popularity within the ecosystem of energy savings to provide predictive analysis and highly non-linear modelling accuracy (Durrani et al., 2018; Ronay and Bhinge, 2015). With rapid evolution within the field of data-driven modelling from various study disciplines, the future for data-driven energy-saving is promising. In this aspect, some of the significant works that contributed to utilizing industrial data for data-driven energy savings towards the perspective of Industry 4.0 are compiled in Table 2.3.

Table 2.3: Significant works related to data-driven modelling and analysis

Work	Contribution
Giacone et al. (2008)	Statistical process control approaches are proposed to be used as a modelling method for energy management in small and medium-sized companies.
Motlaghi et al. (2008)	Used a simple neural network model to model a crude oil distillation column and optimized the process
Errico et al. (2009)	Performed modelling of a crude distillation system to achieve energy saving. Utilizes real data from processing plant to construct process model in a commercial process simulator (i.e. Aspen Plus).
Le and Pang (2013)	Implemented an energy-saving decision system based on wavelet transformation, segment clustering and support vector machines (SVM).
Katchasuwanmanee et al. (2016)	Revealed an energy-smart production management system named “e-ProMan”. It uses real-time and historical data to perform correlation analysis on energy, work and data flow on industrial machines.

Table 2.3 (Cont.): Significant works related to data-driven modelling and analysis

Work	**Contribution**
Ronay and Bhinge (2015)	Performed energy prediction for the consumption of a machine tool using an ensemble of neural networks that were optimized by the NSGA-II (Non-dominated Sorting Genetic Algorithm II).
Kaile Zhou et al. (2016b)	Revealed that models for data-driven energy management systems are mainly based on evolutionary optimization, mathematical programming, machine learning techniques and statistics.
Zou et al. (2017)	Used a stochastic mathematical model to relate sensor data to state variables and disruption events. Carried out an energy-saving evaluation based on energy-saving opportunity window concept.
Durrani et al. (2018)	Data-driven optimization of a crude distillation unit for the consideration of energy efficiency using Taguchi method, Genetic Algorithm and Artificial Neural Networks.
Adenuga et al. (2019)	Demonstrated an energy efficiency analysis modelling system which considers aspects of energy consumed, operational energy costs, baseline power, production power and utilized total power.
Ji et al. (2019)	Used a Deep Belief Network with a combination of a Genetic Algorithm to reduce energy consumption in a simulated machine tool.

In today's world, training a data-driven model can be as easy as loading a dataset and pressing one button. Nevertheless, ensuring the model is truly useful in the real world is not presumption, the whole framework has to be properly designed meticulously for it to work (Gallagher et al., 2019). There are also many hidden problems during implementation that a theoretical study will not consider. For example, inconsistency of equipment wearing in a single process (Y. Zhang et al., 2018), uncertainties in operations (Touš et al., 2015a), information architecture (Gallagher et al., 2019), physical constraints (Teng et al., 2019) and even more unexpected problems. Researchers should favour actual industrial implementations of energy-saving

solution much more than elegant theoretical mathematics as they bring much more insights and experience to the problem itself. Diving too deep into theoretical problems will only lead to producing many "garbage-in-garbage-out" models that do no good to the industry. The needs to combine industrial implementation, knowledge and data-driven modelling is a critical aspect for successful projects (Máša et al., 2018). Tao and Qi (2019) pointed out that models without consideration of industrial implementation can fail to reflect in the real world, leading to poor decisions and catastrophic problems. The work provided an example of a Beijing-based company causing a steam turbine to overheat due to not including lubricant levels into its digital twin. Some works that carried-out industrial implementation for the purpose of energy savings can be found in Table 2.4.

Table 2.4: Significant works which include industrial implementation of energy-saving solutions

Work	Contribution
Gao (2013)	Google implemented a Neural network-based controller to predict and learn the power usage efficiency of its industrial data centre. The application was successful in catching erroneous meter reading and optimize plant operational parameters.
Touš et al. (2015)	Used a combination of Artificial Neural Network and regression analysis to achieve a stochastic Monte Carlo optimal planning decision in a waste-to-energy plant. The work implemented the simulation tool in a combined heat and power (CHP) plant within the Czech Republic and improved planning accuracy by 45%. This relates to 130 Euro increase in daily revenue.
Xu et al. (2017)	Developed a novel system architecture focused on distributed energy savings and big data analysis for industrial cloud manufacturing. The system was implemented in the form of a functional module.
Zhang et al. (2018)	Utilized a complete architecture of energy big data perception and acquisition which utilized IoT. Implemented a proof-of-concept application in the ceramic manufacturing industry. The energy monitoring system successfully identified potential energy-saving opportunities from a poorly maintained ball mill unit.

Table 2.4 (Cont.): Significant works which include industrial implementation of energy-saving solutions

Teng et al. (2019)	Proposed the combination of correlation-based principal component analysis-aided statistical process optimization (PASPO) to find optimal processing conditions from the SCADA system which simultaneously improved product quality, process energy and environmental impacts. The framework was deployed in a Malaysian oil refinery where process energy improved by 3.5 %, Acidification Potential improved by 90.89 % while main product yield and quality improved by 84.4 % and 46.5 % respectively.
Gallagher et al. (2019)	A cloud computing-based system called "IntelliMaV" was applied to verify the energy savings in near real-time. The system uses various machine learning models (such as ordinary least square, k-nearest neighbours, Artificial Neural Network and Support Vector Machines) for learning the data. The system identified various energy-saving potential from a large biomedical manufacturing facility in Limerick, Ireland.

Apart from operational difficulties during industrial implementation, the user interfaces and user experience (UI/UX) of the energy solution is also important. Zhang et al. (2018) demonstrated that a good user interface for data visualization is effective in identifying low energy efficiency ball mills in a Chinese milling factory. Good visualization at this stage can compress complex high dimensional data into simple bar charts that are directly indicative of the priority within the process (Teng et al., 2019). Furthermore, Lade et al. (2017) argued that a visualization platform is essential for the industrial big data that is gathered. The work discussed the importance of data visualization for aiding engineers to obtain a lucid and comprehensive overview of the whole process system. Some ready-made tools for this purpose are available in the market, such as AWS QuickSight, Google Data Studio, Tableau and Microsoft BI which makes implementation straightforward. The upcoming chapters will discuss the information infrastructure that will be hosting the data pipeline (Chapter 2.3), integration of advanced technologies, government policies and barrier between researchers and industrialists.

2.3 Industrial Infrastructures and Enabling Technologies

Industrial infrastructure for managing the processing and operational data is generally distributed and non-unified. Data infrastructure technologies may range from QR (Quick Response) codes and RFID (Radio-frequency Identification) tags to smart machines and devices (Shrouf et al., 2014). From a holistic viewpoint, the hierarchy of the data acquisition system (see Figure 2.3) can generalize the information infrastructure within a conventional industrial plant. As the basis for information infrastructure, sensors and machines are connected to input/output (I/O) to transfer physical and measured information to electronic input. For most processing system, there exists a programmable logic control (PLC) to deal with process actions which require fast responses. The supervisory control and data acquisition (SCADA) system is a monitor system for the overview of the manufacturing area. Control set-points and operational settings are mostly carried out at the SCADA level. Manufacturing execution system (MES) are higher-level computerized system to track material and energy flow and to aid with production management. Lastly, enterprise resource planning (ERP) is a system used in management-level for customer services, sales, procurement, production, distribution, accounting, human resource, corporate performance, and governance. Conventionally, the data is collected from the bottom of the hierarchy (Figure 2.3) and gradually moving up. Processing data is also stored in some cases and processing historian software and databases can be used. For the computation and optimization of system and operation, the data can either be sent to embedded devices, cloud services or a centralized machine to be processed. Alternatively, the decision made by the computation will flow in the reverse direction of the hierarchy, ultimately being implemented on the machine level.

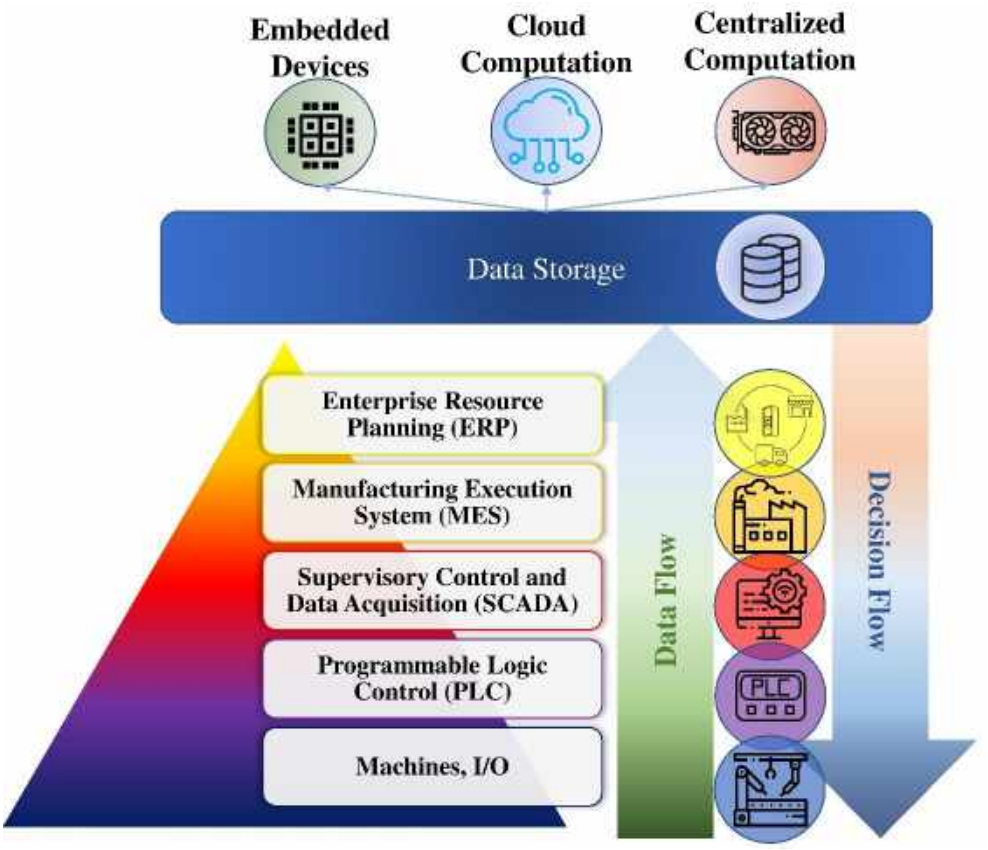


Figure 2.3: Typical Hierarchy of Data Acquisition Systems

In large industries, the *de facto* standard of data acquisition system is the SCADA system with Modbus RTU protocol (Kang and Robles, 2009). However, in small-and-medium enterprises, Máša et al. (Máša et al., 2018) discussed that many facilities have incomplete data acquisition system and the implementation of SCADA systems (at the very least) is crucial for energy-saving projects. Yuan et al. (2017) also observed the incompleteness of energy Big Data in the industry due to non-ideal data collection infrastructure. Additionally, Yan et al. (2017) discussed that industrial information is multisource, heterogeneous and sometimes unstructured. The existence of good information infrastructure is crucial for energy-saving projects. Nevertheless, the implementation of SCADA system in existing industrial facilities is a major challenge due to the expensive costs of the systems. For example, Stojkovic and Vujosevic (2002) demonstrated that even a compact and small experimental SCADA system would cost around 10,000 USD in Europe. OmniSite company (Omnisite, 2009) estimated that a typical industrial SCADA system (20 stations) in the United States would cost a total of 476,500 USD in 10 years. In many cases, the integrity of SCADA system should not be compromised for costs, as disastrous incidents can occur due to inadequate security within the system. Works of Miller and Rowe (2012) showed that poor implementation or operation of

SCADA systems can potentially give catastrophic accidents due to malware infiltration and misuse of resources. Furthermore, for industrial processes, the SCADA system should not fail under power shortage. The cost of mitigating power shortage is often overlooked. For instance, a single industrial-grade 10 kVA uninterruptable power supply (UPS) system that can acts as backup power for 18 minutes could easily cost more than 3500 Euros (approximately 3855 USD) (Eco Power Supplies Ltd, 2019). Zhu and Wu (2005) even pointed-out that data acquisition for intelligent analysis is a compromise between the cost for information infrastructure and the data quality. The work highlighted the need to process corrupted data under cost-constraints. Evidently, industrial data acquisition is not a simple task.

2.3.1 Industrial Internet of Things for Data Coverage

A recent direction to reduce the cost of acquiring industrial data is by implementing the industrial internet of things (IIoT) sensors and infrastructures (Lade et al., 2017). In terms of device pricing, Zheng et al. (2011) pointed out that IoT improves intelligent interconnection of data infrastructure by utilizing low-cost information gathering and dissemination devices. As the cost for IoT devices diminishes with technical progress, many processes can benefit from an increased awareness of data-driven analytics (Mattern and Floerkemeier, 2010). Works of Xiaojun et al. (2015) estimates that IoT system can reduce hardware costs by 1/10 for the purpose of industrial monitoring and forecasting. Moreover, the amount and variety of data that is enabled by industrial IoT infrastructure are remarkable. Ahmed et al. (2017) discussed that technologies from IoT have enabled explosive growth in the number of devices connected to the data infrastructure, allowing for Big Data analytics. The work discussed that IoT environment provides opportunities for intelligent decision-making, improved efficiencies, reduction of data silos and value-added applications. Additionally, Shrouf et al. (2014) also repositioned the importance of IoT in smart factories, highlighting that IoT can improve the factory sustainability, mass customization, flexibility, planning methods, proactive maintenance, connected supply chain and energy management. The perspective that industrial Big Data can be achieved as a result of IoT adaptation in manufacturing had also been proposed by Mourtzis et al. (2016). Evidently, the implementation of IoT in Industry 4.0 would greatly improve data collection frequency, data coverage and data variety.

The pathway towards industrial IoT is not all sunshine and rainbows as there are many major challenges. Kim and Kim (2016) surveyed that IoT for energy management faces challenges from a technological prospect, market potential and regulatory environment. One of the most

significant technical challenges is the robustness and reliability of industrial IoT (Koutsiamanis et al., 2018). Specifically, Duan et al. (2016) even discussed that the reliability of data transmission has become the major bottleneck of the IoT application in industry as data transmission failures can lead to production errors. One of the most recognized solutions for this reliability problem in industrial IoT is by applying 5G-enabled IoT (5G-IoT) technologies (Rao and Prasad, 2018). 5G is the fifth generation of mobile, cellular technologies and solution which features over 10 Gb/s of data rate and lesser than 1 *ms* of latency. Cheng et al. (2018) discussed that 5G would enable remote real-time monitoring, operation control and unmanned factory with its low packet loss rate (<1 x 10^{-12}), which the latest 4G transmission technology cannot achieve. The timing of when will 5G be industrially matured and applicable is also to be questioned. Although some researchers expect 5G to be rolling out between 2018 and 2020 (Mueck et al., 2016), with the recent global technological trade war revolving around 5G (Wolff, 2018), a delay in rollout is imminent (Peter Harrell, 2019). A conservative expected roll-out date for the 5G technology will be around 2025 (Supriadi and Haryadi, 2017). This work proposes an expected timeline for the evolution of industrial data-driven energy-saving timeline in Figure 2.4.

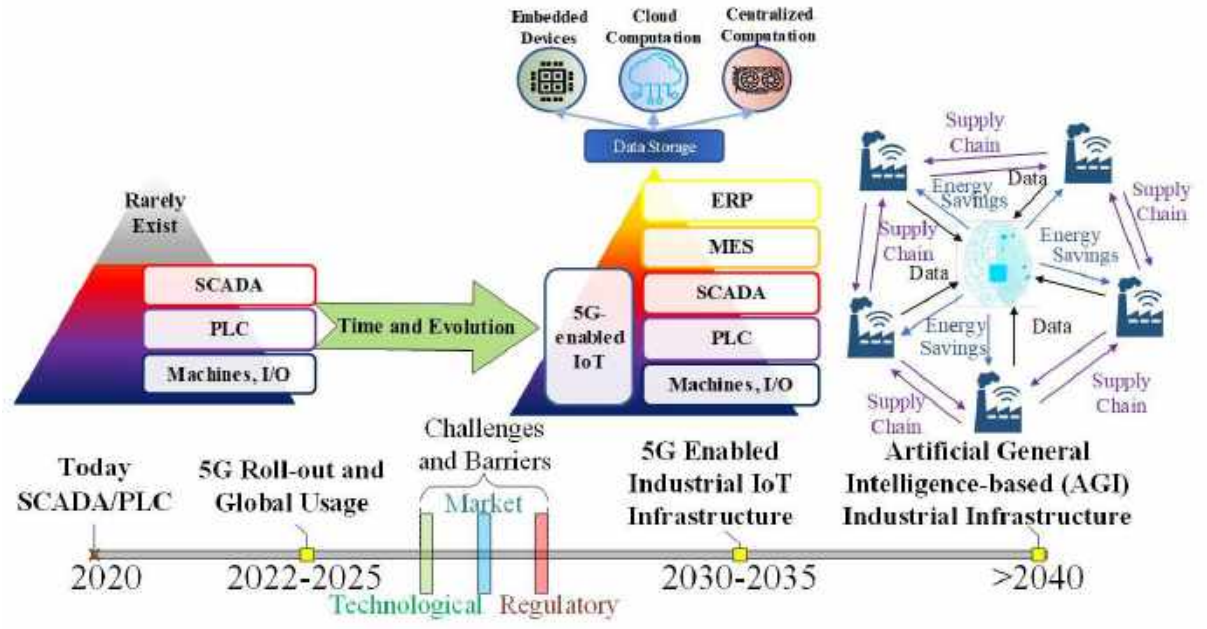


Figure 2.4: Expected timeline for the evolution of data-driven energy savings in Industry 4.0

Another significant challenge for the transition of IoT is on the security of the IoT network, software, and hardware. By rediscovering past experiences, researchers such as Sadeghi et al. (Sadeghi et al., 2015) pointed out the challenge in terms of network security for industrial IoT. They revealed that common security attacks on industrial IoT System can be in the form of

runtime-attacks, reverse engineering, malware, network eavesdropping, man-in-the-middle attack, denial of service attack, social engineering, and phishing. Within these security issues, Sajid et al. (2016) revealed that malware and malicious codes are statistically the most pressing security issue for IIoT system. Nevertheless, more research effort is being carried out in the fields of security technologies for IIoT (Lesjak et al., 2015) such as ARM TrustZone and variants of security controller. In terms of network and software security, researchers are also working to design trusted execution environments (TEEs) for industrial IoT applications (Pinto et al., 2017).

An obvious obstacle for the implementation of IoT devices in the industrial setting is on the standardization of the system to meet industrial regulations (Kim and Kim, 2016; Weyer et al., 2015). For most matured industrial processes, the IoT (and non-IoT) devices have to be up to International Electrotechnical Commission's (LEC) protection rating of IP69k (Müller et al., 2014) (i.e., dust-tight and able to sustain high-pressure cleaning or steam jet). Furthermore, in specialized industries, more difficult-to-achieve protection standards may apply. For example, in the oil and gas industry, there may be a regulatory requirement for devices to be explosion-proof (Chen, 2011), which can greatly incur costs of industrial IoT implementation. At this stage of development, IoT devices are far from being matured in terms of being standardized for industrial requirements (Weyer et al., 2015). Much more research and manufacturing effort must be carried out to advance the field.

2.3.2 Digital Twins and Cyber-physical Systems

The first concept of a digital twin was presented by Grieves (2014) to better understand production and design using a virtual factory replication (see Figure 2.5). Later, the famous work from Glaessgen and Stargel (2012) proposed a digital twin paradigm for NASA and US air force application. The work described digital twin as an integrated multi-physics, multiscale, probabilistic simulation of an as-built system that corresponds to the best available model, sensor updates, system history, etc. Subsequently, Grieves and Vickers (2016) extended the original digital twin concept towards mitigating undesirable and critical behaviours in complex systems. This crucial work also discussed the application of digital twins within product lifecycles. A concise guideline for digital twin-driven product design was also presented by Tao et al. (2019). Their work discussed that six important steps were required to build a functional digital twin, which includes (i) building a virtual representation of physical product using CAD or 3D modelling (ii) process data to facilitate design decision-making (iii)

simulation of physical systems in the virtual environments (iv) test the physical system to calibrate the virtual world (v) establish real-time bi-directional secure connections between physical and cyber system (vi) collect more related data for continuous system integration.

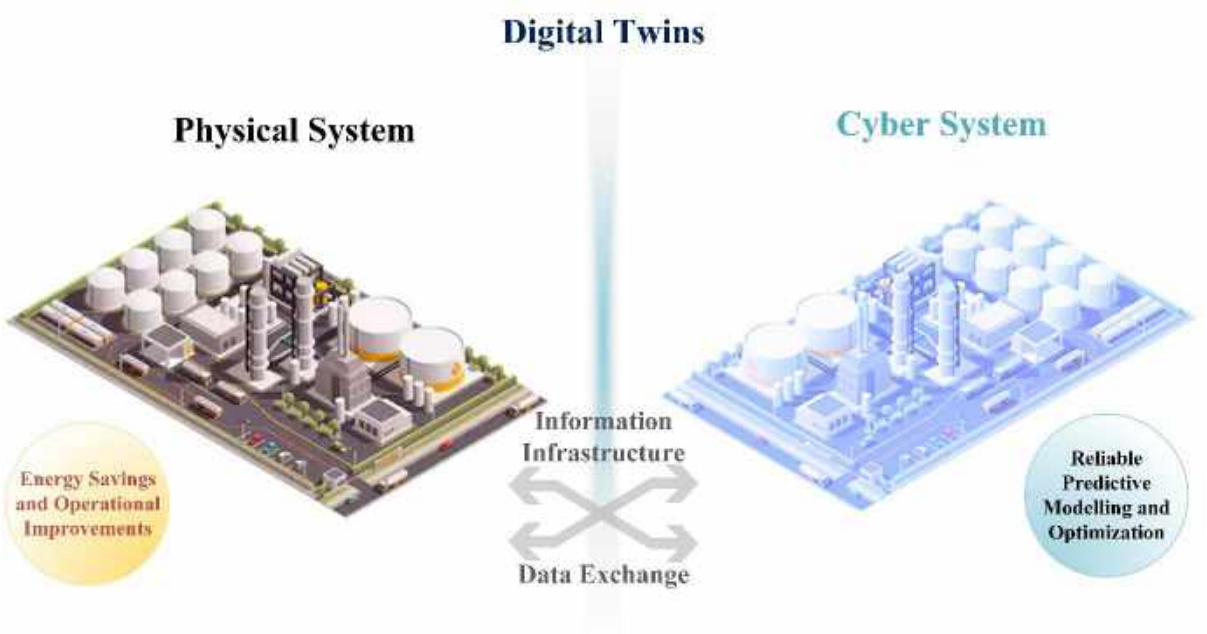


Figure 2.5: Illustration of a digital twin for a cyber-physical production system

Nevertheless, implementing digital twins within existing infrastructures is not an easy task. Uhlemann et al. (2017) discussed that challenges related to digital twins include difficulties in real-time data acquisition, requirements for information systems and infrastructure, improper implementation, standardization of data acquisition systems, costs of investment, weak integrity between the cyber and physical world and data security. Furthermore, the work also highlights the functional dilemma of the simulation aspect and optimization aspect within digital twins. It is to be questioned whether the accuracy of simulation or optimization should be prioritized. Moreover, it is required to consider and balance between fault-estimation (Chua et al., 2020) and state-estimation (Mishra et al., 2014) for dynamic systems of digital twins or cyber-physical systems. State- and fault-estimation possess computational dilemma, but can be utilized for effective condition monitoring (Migueláñez and Lane, 2010) and predictive maintenance in industrial systems via digital twins (Jain et al., 2020). Despite the difficulties, the infrastructure of digital twins has been established in various industry-leading companies such as General Electric, PTC, Siemens, Oracle, ANSYS, Dassault, SAP and Altair (Qi et al., 2018). Digital twins have also been implemented for the recovery, recycle and remanufacture of waste electrical and electronic equipment (Wang and Wang, 2019). J. Wang et al. (2019)

had also shown that digital twins are effective for the application of rotary equipment in manufacturing. Some preliminary software solution for the deployment of digital twins are already available in the market (see Table 2.5).

Table 2.5: Software solutions for digital twins available commercially

Software Solution	**Company**	**Features***
PREDIX (General Electric, 2019)	General Electric (GE)	Supports infrastructure with asset-centric communication, edge-to-cloud, distributed architecture, data management, integrated analytics and embedded cybersecurity. The digital twin is mainly dependant on asset models and knowledge base. Machine learning is also supported in the platform.
IoT Production Monitoring (Oracle, 2019)	Oracle	Supports real-time Key Performance Indicators-based (KPIs) analytics. A feature named 'Deep Dive' is available to gain operational visibility in multiple manufacturing levels. Able to diagnose production anomaly, act on prescriptive analytics and reduce inefficiencies.
Akselos (Akselos, 2019)	Akselos	Provides a high-fidelity multi-physics 3D-based digital twins framework that utilizes RB-FEA (reduced basis finite element analysis) technology. The technology uses a parametric component approach to pre-solve simulation building blocks to speed up the 3D digital twin. It also supports cloud computation and sensor integration.
Digital Twin Builder (ScaleOut Software, 2019)	ScaleOut Software	A fully customized digital twin that is constructed by Java or C# object-oriented programming codes. Supports cloud computation and data source integration. Good infrastructure for process event messages and real-time feedback.

*As of the time of writing

Table 2.5 (Cont.): Software solutions for digital twins available commercially

Software Solution	Company	Features*
Elements for IoT (CONTACT software, 2019)	**CONTACT Software**	**Provides an asset state-based digital twin for monitoring and predictive analytics. Supports sensor integration, 3D models, maintenance history, customer records, cloud computation and edge connectivity.**
Seebo Industry 4.0 Platform (Seebo, 2019)	Seebo Interactive	Specializes in process flow-based digital twins. The application focuses on process-based predictive analysis, automatic root cause analysis and predictive simulations. Artificial intelligence-enabled analytics and streamline IoT data integration are enabled through Microsoft Azure technology stack in the back end.

*As of the time of writing

The driving model of digital twins was initially based-on 3D finite element analysis (FEA) or computational fluid dynamics (CFD) (Glaessgen and Stargel, 2012). However, the time taken to simulate a few minutes of such analysis in the virtual world would take time in the scale of hours in the real world (Wriggers, 2008) due to the complexity of solving large amounts of particles within the model. In such cases, the simulation cannot provide real-time solutions for the digital twin's requirement. Although leading researchers are able to use finite element-based simulation as the digital twin for smaller cyber-physical systems (Tuegel et al., 2011), extending such approach to a full production or manufacturing system is generally difficult without clever simplifications. Due to this difficulty during simulation, some researcher classifies larger manufacturing systems as cyber-physical production systems (CPPS) to specifically address its problems and solutions.

A successful approach to tackle this problem was from using knowledge-based domains. For example, Miller et al. (2018) extended their 3D-CAD (Computer-Aided Design) digital twin with behavioural information. By taking in knowledge-based input from experts, they were able to improve the value of their digital twin for utilization. The knowledge-based domain also includes the utilization of data from experiments. Kraft (2016) demonstrated that digital

twins can benefit greatly from the integrative use of computational fluid dynamics and experimental fluid dynamics. Especially for cyber-physical production systems (CPPS), the existence of knowledge-based data cannot be exempted as there are many human decisions apart from pure physics. For the case of Liu et al. (2019), they used knowledge of discrete events, system dynamics and physical components to construct their digital twin for the application of shop-flow manufacturing systems.

Artificial Intelligence is a more advanced approach to implement seamless digital twins. For instance, C2PS was a digital twin architecture developed by Alam and El Saddik (2017) to analyse key properties of cloud-based digital twin such as computation, control and communication. The work utilized a Bayesian belief network which dynamically considers the digital twin's context. Recent work from Luo et al. (2019) utilized an artificial neural network to model and utilize the data streams from the digital twin of a CNC milling machine tool. The work proved that Artificial Intelligence approach can be used as a multi-domain unified modelling method for establishing a digital twin. Some more advanced work can be found from pioneering teams of Artificial Intelligence. Peng et al. (2018) from OpenAI demonstrated that a randomized initialization for Deep Reinforcement Learning acts as a method to reduce the "reality gap" to Sim-to-Real problems. Rusu et al. (2016) from DeepMind, also demonstrated that for a Sim-to-Real problem, the utilization of a progressive network to bridge the reality gap and policy transfer can perform more efficiently than Deep Reinforcement Learning. It is clear that this is still a progressing field and there are still many developments awaiting.

2.3.3 Cloud Infrastructures

Cloud infrastructure is a remote system that covers both hardware and software that support web frontend applications which are connected to the cloud storage (Wei and Blake, 2010). It is the core foundation for cloud computing (see Figure 6). The application of cloud computing can be back-dated to the early 1960s, where it was introduced as the "intergalactic computer network" by Joseph Carl Robnett Licklider (J.C.R., 2001). Nowadays, its utility is now being widely extended (e.g., Elastic Compute Cloud by Amazon (Amazon Web Services, 2019a), App Engine by Google (Google Cloud, 2019), etc.). Generally, the cloud services can be distinguished into Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Data as a Service (DaaS) (Kumar and Goudar, 2012).

The global market for private cloud service is anticipated to reach $ 262.4 billion by 2027 (Dawkins, 2019). The ever-expanding market is probably due to the unique advantages offered by cloud computing technology. In terms of the economic aspect, the deployment of cloud computing can avoid the need for high investment cost for constructing a data centre (Dar, 2018; Xue and Xin, 2016). Data security is another factor that caused a high adoption rate of this technology. With the aid of cloud infrastructure, data loss can be avoided. Up till 2018, about 94 % of organizations in Europe, the Middle East and Africa (EMEA) have integrated cloud services into their business model (TechNative, 2016). Scalability (or flexibility) is another key feature of cloud computing which allows users to conveniently scale up or scale down the resources (e.g., the need of cloud storage) based on the actual requirement (Xue and Xin, 2016). Aside from that, the high mobility of the cloud service (i.e., the service and data can be accessed anytime anywhere) allows a timely response to be taken. This is important for energy providers especially when the energy sector is getting more and more competitive (Marinakis et al., 2018). Last but not least, the implementation of cloud computing helps to enhance business capability. With the aid of the big data analytics which is embedded in most cloud infrastructure, insightful findings can be yielded from the massive data (Marinakis et al., 2018). For instance, by integrating the smart metering with cloud technology, energy suppliers are granted with a bird-eye's view on the electricity flows, starting from the origin to the destination (Sykes, 2019). In other words, outages can be efficiently and accurately identified, while correction actions can, therefore, be carried out earlier.

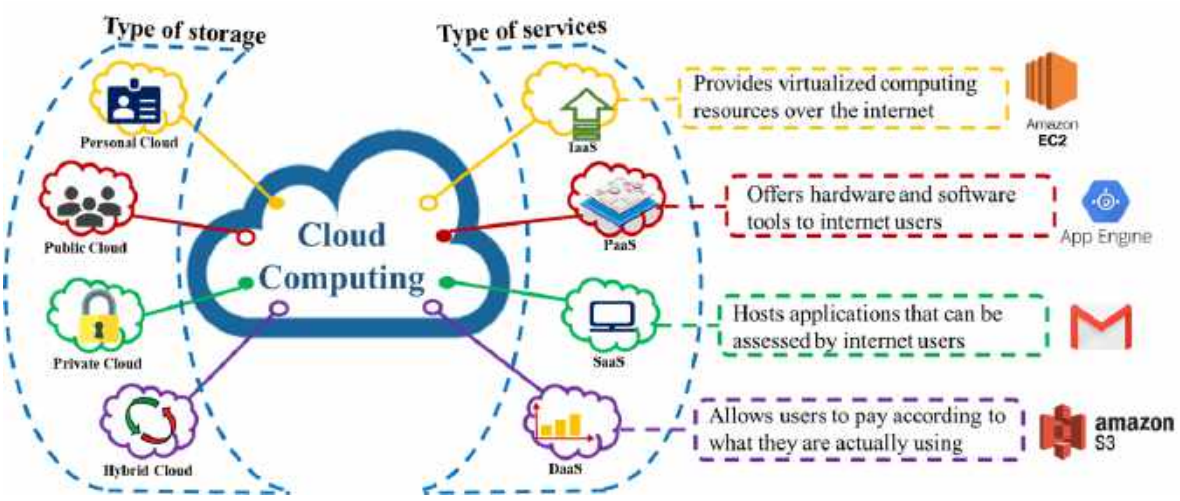


Figure 2.6: A holistic view of cloud computing options for industrial applications

It is estimated that the annual generation of data for a plant is about 72 TB per year (attributed to the application of IoT) (Mourtzis et al., 2016). Therefore, it is vital to have enough data storage capacity to store a tremendous amount of data. Cloud storage is a cost-effective alternative to conventional hardware storage. Based on Figure 2.6, cloud storage can be classified into four types, i.e., personal cloud, public cloud, private cloud, and hybrid cloud, where some details of these cloud storages are tabulated in Table 2.6. The deployment of cloud storage in the energy sector have been elucidated in numerous works. For instance, Chen et al. (2016) highlighted various roles of open innovation intermediaries (OIIs) (includes providing cloud storage to store and protect the clients' data), in promoting the smart grid industry in China. It can be used to store weather information, energy profile and other data that were generated in the smart grid (Mayilvaganan and Sabitha, 2013). Alonso et al. (2013) proposed to use cloud storage that was developed by Ingenia to store big data that were generated from the energy adapters in an integrated network. Despite the importance of having cloud storage, the huge energy consumption issue for data storage remains a major concern nowadays. It is anticipated that the use of cloud infrastructure will consume up to 20 % of the global power consumption by 2025 (Vidal, 2017). To address this issue, some researchers (e.g., Long et al. (2014) and Yang (2013)) are exploring ways to improve the energy efficiency of the cloud system.

Table 2.6: Four types of cloud storage available

Cloud Storage	**Description**	**Example and Pricing**
Personal	• Stores individual's data in the cloud environment which is accessible from anywhere at any time. • Syncs and shares data across multiple devices (Beal, 2019) • A subset of public cloud storage	• iCloud (free tier: 5 GB; 50 GB plan: $0.0198/GB; 200 GB plan: $0.01495/GB; 1 TB plan: $0.00999/GB (Statt, 2015)) • Google Drive (free tier: 15 GB; 100 GB plan: $0.0199/GB; 1 TB plan: $0.00999/GB; 10 TB plan: $0.00999/GB (Lardinois, 2014))

Table 2.6 (Cont.): Four types of cloud storage available

Cloud Storage	• Description	• Example and Pricing
Public	• Enterprise stores data in an external cloud storage provider, which results in lower data storage cost required. • Lesser flexibility to alter the cloud environment.	• IBM Cloud (below 500 TB: $0.022/GB; more than 500 TB: $0.020/GB (IBM, 2019)) • Amazon Elastic Compute Cloud (free tier: 5 GB; first 50 TB: $0.023/GB; subsequent 450 TB: $0.022/GB; over 500 TB: 0.021/GB (Amazon Web Services, 2019b))
Private	• Enterprise stores data in an internal data centre, which results in higher capital investment cost and maintenance cost needed. • Provide higher level of security as the enterprise owns the cloud environment.	• Phoenix co-lo (capital investment for a scenario of 100 virtual server instances: >$300 k (Singh, 2012)) • Private cloud will become more economically preferable when the monthly spend is more than $ 17 k (Suarez and Kirkwood, 2019).
Hybrid	• Combines the use of public and private cloud storages. • Critical and confidential data are stored in an internal data center, while other data are stored in an external cloud storage provider.	• Cantemo Portal case (Monthly cost for scenario when 100 % of the 2,000 TB data are stored in private cloud: $ 24,000 - 40,000; Monthly cost for scenario when 20 % of the 2,000 TB data are stored in private cloud and the remaining are stored in public storage: $ 12,800 - 48,000 (Klein, 2017)).

Aside from selecting the data storage system, an efficient computing system is needed to deal with the enormous amount of data sets. Apache Hadoop (The Apache Software Foundation, 2019a) and Apache Spark (The Apache Software Foundation, 2019b) are the two iconic computing systems that can be used to process large-scale data, where the latter offers in-memory cluster computing feature that causes it to be outperforming the other (Farhan et al., 2018). It is capable to run the same task 100 times and 10 times faster in memory and on disk respectively (Elshawi et al., 2018). In addition, unlike Apache Hadoop, Apache Spark is suitable for real-time and streaming data analysis (Tu et al., 2017). Apache Spark can be built on Java, Scala, Python, etc. Till-date, some works have reported the use of Apache Spark in the energy sector. Apache Spark machine learning tool (MLib) is utilized to perform big data analytics for well and reservoir management system (Bello et al., 2017). More recently, Krome and Sander (2018) had developed a hybrid computing system that integrates the use of Apache Spark and the R language, to perform time series analysis for energy price and load profiles forecasting.

Aside from cloud computing, edge computing which the data is processed at the source point rather than in a centralized cloud-based data centre, is being introduced. This computing technology is often used to deal with the time-sensitive data generated from the Industrial Internet of Things (IIoT), i.e. interconnected sensors, instruments, smart devices, and facilities (Figure 2.7). To achieve the edge computing services, the use of microcontroller (e.g. Arduino (Arduino, 2016), which serves as a simple computer that can run a single program repetitively) and general-purpose computer (e.g., Raspberry Pi (Newmarch and Newmarch, 2017) and NVIDIA Jetson Nano (NVIDIA, 2019) which is capable to run multiple complex programs; where the latter offers greater capability in machine learning and artificial intelligence applications). Note that the edge computing technology can be implemented in the energy sector to enhance production capability and improve process efficiency. For example, in a scenario where the windmills are located at remote areas which is not accessible to the internet, edge technology can be implemented to (i) collect and analyse data; and (ii) optimize and make decisions (e.g., adjust the opening angle of wind turbine blades) (General Electric, 2016). In one of the recent publications, an Edge-IoT platform is proposed to effectively reduce the energy consumption of an energy distribution network (Sittón-Candanedo et al., 2019).

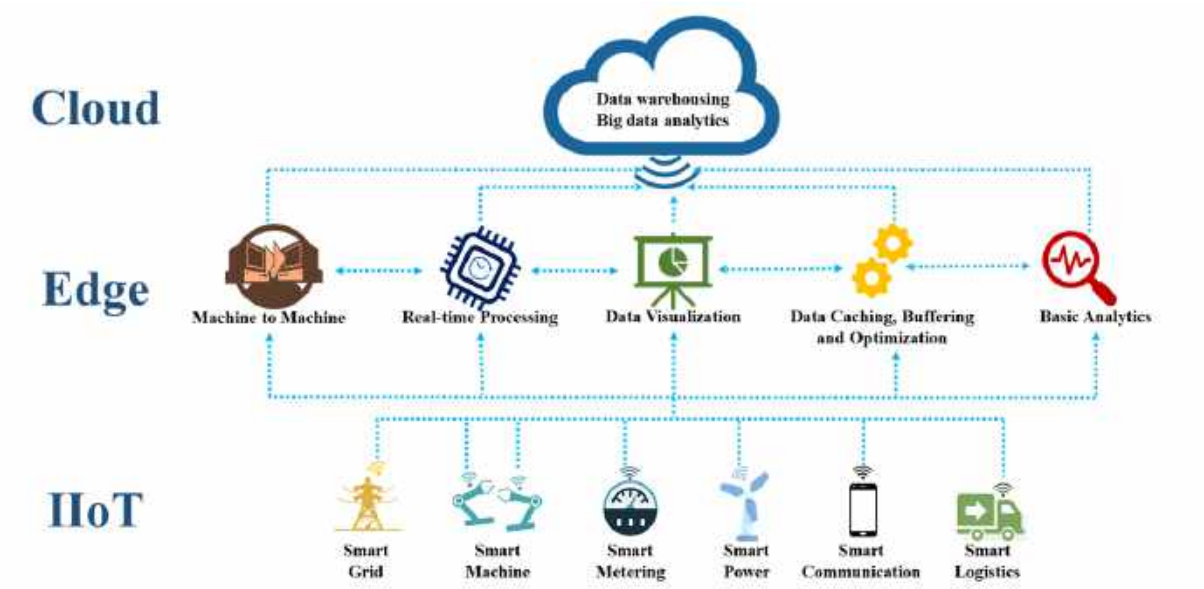


Figure 2.7: Linkage between cloud, edge, and Industrial Internet of Things (IIoT).

To efficiently process the massive amount of data, specialized hardware such as Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) are required (Li and Porter, 2019). In general, GPUs and TPUs are used as accelerators for the sub-portions of the model that can be decomposed into data-parallel computations (Wang, Yu Emma et al., 2019), where the latter is designed specifically for neural network machine learning. The overall comparisons of each type of processing unit are highlighted in Table 2.7.

Table 2.7: Comparisons between the three processing units (Jouppi et al., 2017; Reitsma, 2019; Sato et al., 2017)

Criteria	**CPU**[1]	**GPU**	**TPU**
Compute primitive	Scalar (1 x 1 data unit)	Vector (1 x N data unit)	Tensor (N x N data unit)
Application	General-purpose	Graphics Rendering	Machine Learning model
Operations per cycle	tens	tens of thousands	up to 128 thousand
Relative Performance to watt ratio (based on CPU)	1	2.9	83
Throughput per second[2]	5,482	13,194	225,000
Training cost[3]	$ 1.0507	$ 0.3995	$ 0.2410
Machine cost	$ 0.14/hour	$ 1.87/hour	$ 4.57/hour
Developers	Intel, NVIDIA, IBM, Samsung, etc.	NVIDIA, AWS[4], AMD, PowerVR, etc.	Google

Footnote: [1]Central Processing Unit; [2]Under 7ms latency limit; [3]Trained with Adam until 0.25 validation loss was reached for 5 runs; [4]Amazon Web Services

2.3.4 Potential for Blockchain

The blockchain technology was originally designed by Satoshi Nakamoto to serve as the main foundation for Bitcoin (Nakamoto, 2008). In general, it is a growing list of records (or blocks) that stored the information of all committed transactions (Bolt, 2017). As shown in Figure 2.8, The blockchain architecture can be decomposed into six layers, i.e., data layer, network layer, consensus layer, incentive layer, contract layer and application layer (Wu and Tran, 2018). This unique structure enables the automated execution of smart contracts in the peer-to-peer network which allows multiple users to make changes in the ledger simultaneously (Bolt, 2017). This technology enables four key features of persistency, anonymity, decentralization and auditability which further improve the overall cost-effectiveness and the efficiency of a data management system (Zheng et al., 2017).

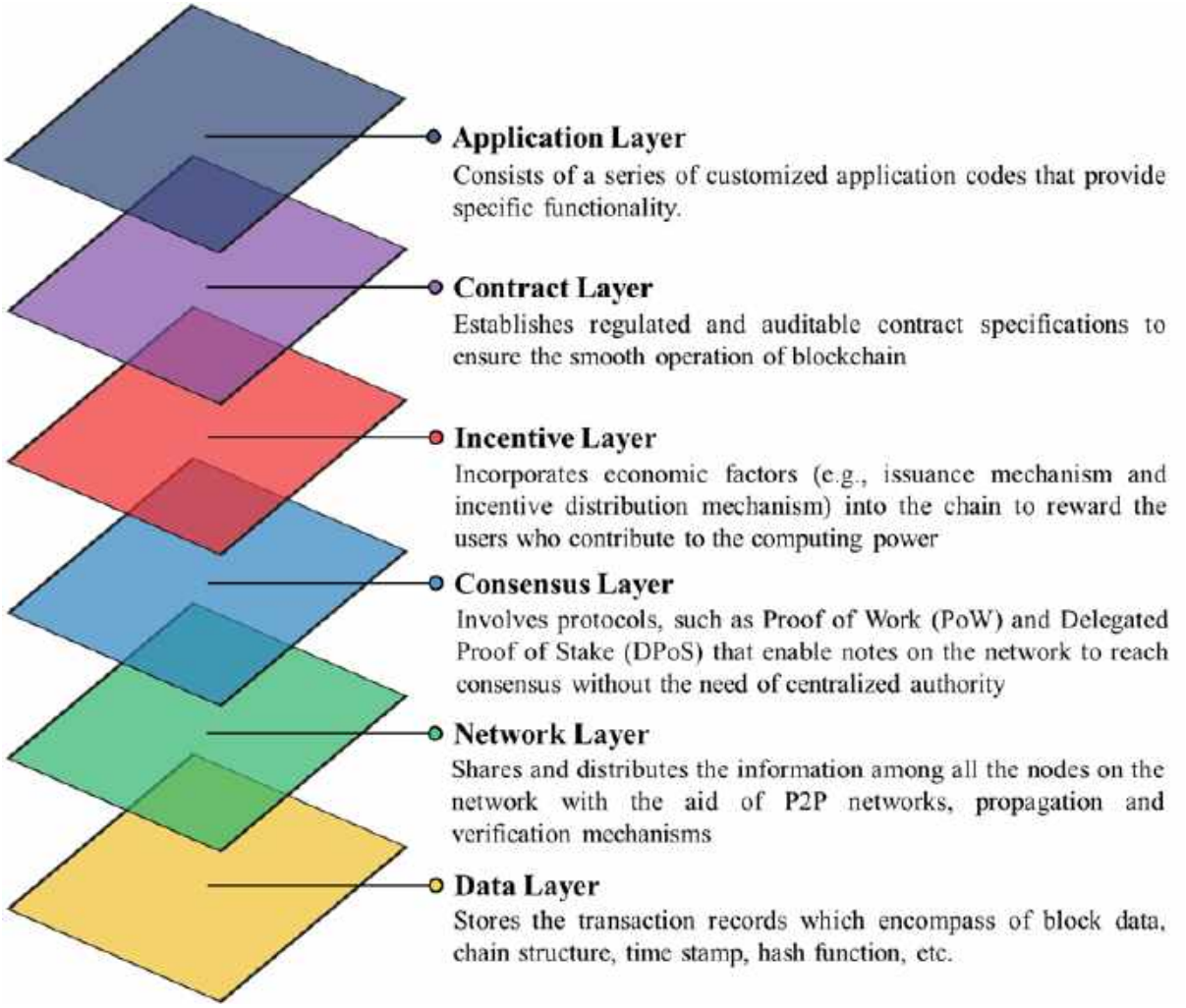


Figure 2.8: Generic blockchain architecture

To date, three generations of blockchains have been developed (see Figure 2.9). *Blockchain 1.0* refers to the origin utility of blockchain technology which generally used for trading cryptocurrencies (Zhao et al., 2016). Bitcoin (Nakamoto, 2008) system is one of the most well-recognized examples that utilized blockchain technology to allow peer-to-peer transactions that operate without the need for a centralized administrator (Angelis and Ribeiro da Silva, 2019). This first generation of blockchain is then evolved to *Blockchain 2.0*, where the values being transferred is no longer restricted to currency but in the form of smart contracts (i.e., programs or scripts that will self-operate after certain conditions and requirements are met without the need of manual commands) (Creydt and Fischer, 2019). This generation of blockchain focuses on the deployment of decentralized applications (dApps). There are three well-recognized features for dApp, i.e., (i) open-sourced (changes are based on the consensus of the users and developer where the code base is available and accessible for scrutiny), decentralized (all data are stored on decentralized platform to encourage transparency, trust and efficiency) and

incentivized (rewarding system to encourage the involvement of the validators). Ethereum, which was proposed by Vitalik Buterin (Buterin, 2014) in late 2013, is one of the most prominent examples for *Blockchain 2.0* applications. Unlike the Bitcoin system, Ethereum focuses on executing codes for any decentralized applications that are deployed in the network, instead of merely offering a peer-to-peer cryptocurrency transaction system. Under this context, Ether is the cryptocurrency that fuelled the blockchain, where it is paid to run the smart contract. Blockchain technology is currently applied in diverse applications (also known *Blockchain 3.0*) such as IoT (Wang et al., 2019), healthcare system (Tripathi et al., 2019), education (Alammary et al., 2019), digital government services (Hou, 2017), etc.

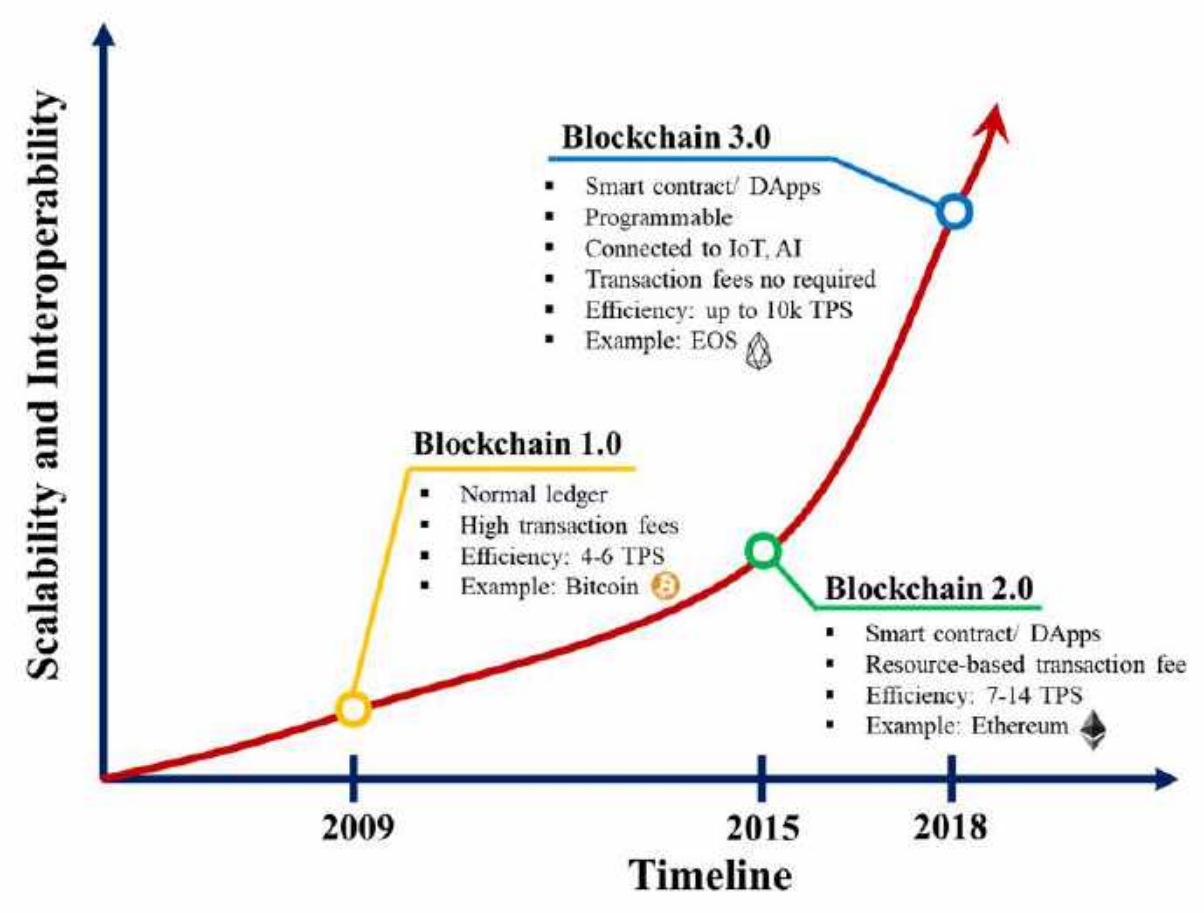


Figure 2.9: Evolution of blockchain technology

In other words, the functionalities of blockchain technology are no longer limited to the finance and asset transfer-related applications, and therefore, higher scalability is obtained (Gatteschi et al., 2018). There are numerous successful blockchain-based decentralized operating systems under the third generation of blockchain which non-exhaustively includes EOS (Block.one,

2019), Cardano (Cardano, 2019) and ICON (ICON Foundation, 2019). More recently, major companies such as IBM, Intel and Microsoft, are attempting to incorporate artificial intelligence (AI) and machine learning into the blockchain system (Salah et al., 2019). The unique ability of AI and machine learning which is capable to address and express uncertainty opens another possibility for blockchain application in solving complex problems (Angelis and Ribeiro da Silva, 2019). Under this generation, a novel data structure, Tangle (i.e., one of the distributed ledgers that is based on Directed Acyclic Graph (DAG)) is commonly being implemented as it can overcome the efficient issues of the conventional blockchain network (Popov, 2018). IOTA ledger is one of the notable DAG-based ledgers that was designed to track, record and execute decisions between machines and smart devices that are connected in the IoT network (Hu, 2017). It was founded by David Sønstebø and his colleagues back in 2015 (M and Biradar, 2018). Figure 2.10 represents the schematic diagrams for the structures of conventional blockchain and tangle. As shown, unlike the conventional blockchain structure, the data flow in a tangle is constrained to a single direction (Schueffel, 2018). This enables higher efficiency of data transfer, i.e., about 100 to 270 times faster than that of the conventional blockchain technology (Schueffel, 2018). As a trade-off to this increment in terms of efficiency, tangle provides a lower level of security due to its less robust nature of the structure (i.e., each device merely needs to validate two previous transactions).

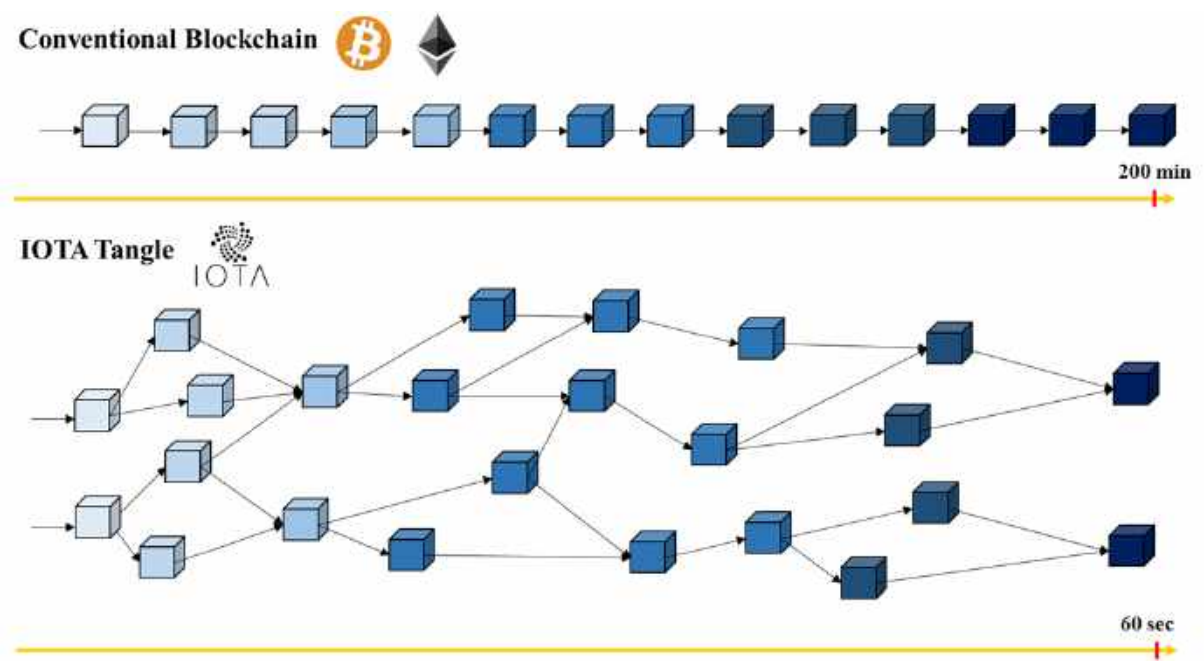


Figure 2.10: Schematic diagrams for the structures of conventional blockchain and tangle (DAG-based blockchain)

Till date, blockchain technologies have been gradually applied in the energy sector to fulfil a diverse range of objectives, including (i) bill payment with cryptocurrency; (ii) rewarding system using cryptocurrency; (iii) peer-to-peer energy trading market; (iv) imbalance settlement for energy market; (v) blockchain-enabled IoT platform; (vi) blockchain-enhanced smart metering; (vii) certification system; (viii) electric vehicle application; (ix) blockchain-based Life Cycle Assessment (LCA); (x) AI-enhanced blockchain for market forecasting; and (xi) blockchain-enhanced Intelligent Energy Storage (IES). These applications are then classified based on the corresponding blockchain technology used, while the remarks of each application are summarized in Table 2.8.

Despite the vast potential of blockchain technology, there are still various challenges that need to be addressed. For instance, the deployment of distributed ledger technology (DLT) solutions (e.g., integrating smart meters with blockchain network) can be costly (Higginson et al., 2019). Aside from this, the low throughput (i.e., the transactions which can be cleared per second) of the technology is another main concern that must be overcome. For instance, the conventional electronic payments that utilized a centralized network can clear thousands of transactions per second (e.g., Visa can support about 1,700 transactions per second (Li, 2019), whereas Bitcoin network which utilized distributed network can merely address about 7 transactions per second (Andoni et al., 2019; Vukolić, 2016). Finally, privacy concern is another key challenge that may hinder users from venturing into blockchain-based interventions. Despite the economic benefits attributed from the blockchain-enabled IoT platform, users might still be reluctant to share their personal information (e.g., consumer behaviour) into the open-source network (Radoglou Grammatikis et al., 2019). For this reason, numerous works have explored strategies for the development of privacy-friendly blockchain-based platform (Kshetri, 2017; Omar et al., 2019; Zhao et al., 2019)

Table 2.8: Energy saving-related application of each generation of blockchain generation

Blockchain Technology	**Application**	**Remark**
Blockchain 1.0	Bill payment with cryptocurrency	Recently, more utility offering companies accept bill payments with cryptocurrencies, e.g., Eva Energy (Eva Energy Romania, 2019) in Romania, NexGen Energy (NexGen Energy Ltd., 2019) in New Zealand, Elegant (Elegant Energy, n.d.) and Enercity AG (Enercity AG, n.d.) in Germany. Despite the analysis shows that the use of Bitcoin can benefit consumers from 4 to 6 % reduction in their electricity bill (Nagata, 2016), most users are still using fiat currencies to complete the transactions instead.
	Rewarding system using cryptocurrency	KWATT was used to reward energy supplier that had committed into waste-to-energy initiatives (i.e., 1 coin is awarded for every kW of energy generated from waste) (Sharpe, 2018). Whereas SolarChange had launched SolarCoin, a cryptocurrency which is designed to reward energy supplier that generate and supply solar energy (Solarcoin, 2019). On the other hand, customers who showed good behaviour, such as supporting carbon-neutral energy can also be incentivized with such virtual currency (e.g., GoodCoin (CarbonX Personal Carbon Trading Inc., 2019)). All the aforementioned tokens or coins can be used to mine other available virtual currencies or be sold to reclaim fiat currency (Andoni et al., 2019).

Table 2.8 (Cont. 1): Energy saving-related application of each generation of blockchain generation

Blockchain Technology	**Application**	**Remark**
Blockchain 2.0	Peer-to-peer energy trading market	Peer-to-peer energy trading market is an interconnected network that allows users to trade energy, i.e., users that have excess energy can either store or sell to other users who encounter energy deficit (Zhang et al., 2017). To date, few companies, including The Brooklyn Microgrid in the United States (Mengelkamp et al., 2018), Power Ledger in Australia (Power Ledger Pty Ltd, 2019), WePower (WePower, 2019) have contributed to the development of such trading ecosystem (some noted as microgrid system). Mihaylov et al. (Mihaylov et al., 2014) on the other hand, had proposed the use of virtual currency (i.e. NRGcoins) to represent the energy flow into and from the grid, where the rate of the NRGcoins varies according to the real-time supply demand situations.
	Imbalance settlement for energy market	Smart contract in blockchain technology enables the near-real-time tracking and confirmation of the billing (Andoni et al., 2019; Danzi et al., 2018). With the aid of the distributed ledger technologies, the issue of having a huge settlement period (up to 28 months (Andoni et al., 2019)) can, therefore, be resolved.

Table 2.8 (Cont. 2): Energy saving-related application of each generation of blockchain generation

Blockchain Technology	**Application**	**Remark**
Blockchain 3.0	Blockchain-enabled IoT platform	Blockchain can be utilized to facilitate the data exchanges between IoT devices (Andoni et al., 2019), where the interoperability of the IoT platform can be assured (Khan and Salah, 2018). Under such platform, smart devices can be programmed to achieve certain desired goals, such as minimizing the need for external energy. Therefore, users can now make more rational decisions about their respective energy usage (e.g., the Swedish housing society case study proposed by Mattila et al. (Brandon et al., 2016)). Grid+ (GridPlus Energy, 2019) is one of the blockchain-based energy companies that connect consumers to the grid via the integration of blockchain technology with IoT devices. This can further result in an energy bill reduction of about 40 % (Brock, 2018).
	Blockchain-enhanced smart metering	Due to the nature of blockchain that could offer higher traceability, the utility charges can become more transparent (Andoni et al., 2019). Companies are exploring applications of blockchain technologies for various metering systems, such as electricity (e.g., Klenergy Metron developed by Pylon Network (Pylon Network, 2019), blockchain-based pay-as-you-go solar services offered by M-PAYG (M-PAYG, 2017)), water (e.g., automated system that can make decisions on maintenance scheduling by Engie (Enegie, 2019)) and heat (e.g., the four Energy Innovation Projects launched by Blocklab (Blocklabs, 2017).

Table 2.8 (Cont. 3): Energy saving-related application of each generation of blockchain generation

Blockchain Technology	**Application**	**Remark**
Blockchain 3.0	Renewable energy certification (REC) systems	Conventionally, power is non-traceable as it will be mixed with others in the common pool (Sorrell et al., 2017). To address this issue, companies such as Acciona (Acciona, 2019) and SP group (Singapore Power Ltd., 2019) have adopted blockchain technology so that their clients can verify that the energies they consumed were sustainable. As a side note, smart contracts are used to perform automated tracking of REC (Andoni et al., 2019).
	Electric vehicle application	The use of blockchain technology offers electric vehicle users greater rights in selecting the source of power supply and provides greater transparency in the power charges (e.g., Share&Charge (The Share&Charge Foundation, 2019)).
	Blockchain-based Life Cycle Assessment (LCA)	The blockchain-based LCA enables a more reliable and trusted data collection correspond to the actual energy consumption throughout the entire product life cycle (Zuo et al., 2018). In a recent publication, Lu et al. (2019) had highlighted the potential of blockchain technology in tracking the life cycle performance of the oil and gas supply chain. This intervention is essentially useful for policymaking, product design and improvement, supply chain management, environmental assessment, and process debottlenecking.

Table 2.8 (Cont. 3): Energy saving-related application of each generation of blockchain generation

Blockchain Technology	Application	Remark
Blockchain 3.0	Artificial intelligence (AI) enhanced blockchain for market forecasting	Numerous companies aim to develop an AI-enhanced blockchain to achieve accurate and efficient forecasting of the energy production and consumption pattern (Mylrea, 2019). The Bittwatt platform (BittWatt Pte. Ltd, 2019) and Jouliette platform (Spectral Energy, 2019) are few of the notable examples that applied AI and blockchain technologies to conduct a successful market forecast.
	Blockchain-enhanced Intelligent Energy Storage (IES)	Blockchain-enhanced IES integrates the strength of the two technologies (blockchain and AI). Blockchain technology serves as a media to trace the energy information, such as charges, prices and carbon footprint data that can facilitate users in making energy storage management decisions; where AI, on the other hand, is capable to provide optimal energy storage management decision for users, based on collected data (Sun et al., 2018).

2.4 Policies and Government Initiatives

Energy efficiency is continually improving according to yearly evaluations. Industrial energy consumption in 2013 was 17 % below its 2000 level and represented 25 % of the energy used by final consumers. The chemical industry was the main consumer with 19 % of the total industrial consumption, followed by steel with 18 % in 2013. (EASME, 2015). The study (ICF Consulting Ltd, 2015) focuses on the evaluation of energy use in most demanding industrial sectors (iron and steel, chemical and pharmaceutical, petroleum refineries, pulp and paper, etc.). It presents that process heating was the most significant energy use (about 66 % of total energy consumption) followed by electricity use (26 % of total energy consumption) in 2013.

Policymakers whose motivation is mainly decreasing energy consumption try to introduce various rules and incentives. One of the tools applied is energy auditing. In the case of EU, this

is specified in the energy efficiency directive 2012/27/EU (European Union, 2012). Regular energy audits are mandatory for large enterprises and it is recommended to introduce programs encouraging SMEs to undergo energy audits too. The directive also emphasizes the necessity for high-quality audits which should be carried out by accredited experts or supervised by independent authorities in a cost-effective manner. Energy audits should also respect regional or international standards such as EN ISO 50001 on energy management. Some of the government initiatives from various countries are tabulated in Table 2.9.

Table 2.9: Initiatives and policies of countries and regions for energy improvements

Country or Region	Initiatives or Policies	Description
European Union	Revised Energy Efficiency Directive (European Comission, 2019; European Union, 2012)	The directive sets the target for an improvement in energy efficiency at EU level of at least 32.5% in 2030, following with an extra 20% target in 2020.
	National Energy and Climate Plant (NECPs) (European Commission, 2019)	Energy and climate plans are designed to meet EU's target for 2030. EU Member States are to establish a national energy and climate plant from 2021 to 2030. The plans address issues related to energy efficiency, renewables, emissions reductions, interconnections, research, and innovation.
United States	Advanced Manufacturing Office (AMO) Initiatives (Advanced Manufacturing Office, 2012)	Provided more than 1300 industrial partnerships and projects for energy saving and energy efficiency in US. Conducted numerous studies that related to energy utilization.

Table 2.9 (Cont. 1): Initiatives and policies of countries and regions for energy improvements

Country or Region	Initiatives or Policies	Description
United States	Federal Energy Management Program (FEMP) (US Department of Energy, 2016)	The program focuses on services to aid agencies in meeting water and energy reduction targets (such as audits, energy management, efficiencies, operations, and contracts). Actions from the program are mandated by laws which are included in title 42 of the United States Code.
	State Energy Program and Energy Efficiency and Conservation Block Grant Program (US Department of Energy, 2011)	Reward projects (according to American Recovery and Reinvestment Act) that reduces operating costs, develops manufacturing capacity, improve energy performance, produce clean energy, enhances environmental performance or a combination of the above.
	State & Local Energy Efficiency (SEE) Action Network (SEE Action, 2013)	An action network facilitated by the federal government to help states, utilities, and other stakeholders to take energy efficiency to scale and achieve targets by 2020. Provides policy design for energy audits and energy efficiencies.

Table 2.9 (Cont. 2): Initiatives and policies of countries and regions for energy improvements

Country or Region	Initiatives or Policies	Description
China	Plan for Energy Development (International Energy Charter, 2018)	National Development and Reform Commission (NDRC) and the National Energy Administration (NEA) provided five-year plans to improve energy systems, save energy, promote renewables, target energy efficiencies, develop energy market, strengthen energy cooperation, and achieve energy sharing.
	Chinese Laws and Regulations with the Energy Charter Treaty (Shan, 2015)	China adopted "One Belt One Road" imitative and various energy treaty practices. The new Silk Road Economic Belt extends to Asia, Africa, Pacific countries to construct an "Energy Silk Road" for global energy benefits.
	Mandatory minimum efficiency standards (Zhou, 2008)	Developed by China National Institute of Standardization to provide mandatory energy efficiency standards to residential and commercial appliances, lighting, heating, and cooling equipment.
India	Perform Achieve and Trade Scheme (PAT) (Bureau of Energy Efficiency, 2018)	Ministry of Power initiated the scheme to improve industrial energy efficiencies, strengthen energy security, lower emissions, improve renewables and contribute to the country's economic value. The scheme covered various energy-intensive sectors for benchmarking.

Table 2.9 (Cont. 3): Initiatives and policies of countries and regions for energy improvements

Country or Region	Initiatives or Policies	Description
India	The Energy Conservation Act (Ministry of Power, 2001)	The Act empowers the government to notify energy-intensive industries, establish energy standards, appoint energy audits and other regulations regarding energy conservations.
	Energy Efficiency in Small and Medium Enterprises Scheme (Bureau of Energy Efficiency, 2008)	Bureau of Energy Efficiency and designated state agencies have initiated diagnostic studies in various enterprises cluster. The task was to conduct energy audits, prepare project reports, enhance capacities of service, provision for financing, improve awareness and outreach.
Malaysia	National Energy Efficiency Action Plan (Ministry of Energy, 2015)	Ministry of Energy, Green Technology and Water prepared a budget (approximately 13 million USD) to promote low energy prices, finance for energy efficiency, national plans for energy efficiency, champion to drive energy efficiency and consistency in embarking on energy efficiency. The action plan covers energy targets up to the year 2025.

Table 2.9 (Cont. 4): Initiatives and policies of countries and regions for energy improvements

Country or Region	Initiatives or Policies	Description
Malaysia	Economic Transformation Programme (Prime Minister's Office, 2010)	The programme covers a large variety of entry point projects and allocated 14 million MYR (approximately 3.3 million USD) for the improvement of energy efficiency in the energy industry. The action on energy efficiency focuses on better energy-efficiency practices, simulate sales of energy-efficient appliances, provide co-generation economic support and other energy-efficient technologies.
	Green Technology Master Plan (Ministry of Energy Green Technology and Water, 2017)	The master plan was established by the Ministry of Energy, Green Technology and Water to focus on six key sectors which include energy, manufacturing, transportation, building, waste, and water. The purpose of the master plan is to provide policy directions towards green technology, elevate economics of energy-efficient technologies, improve cost efficiency, lower energy prices, and accelerate technological advancement.

The common method of energy efficiency assessment for action plans and regulations in various countries revolves around energy audits. The audits usually propose implementation of an energy management system which should be responsible for all the measures and other future activities (monitoring of energy consumption, evaluation, maintenance, etc.). Often, a large part of the energy audit is focused on improving buildings efficiency by use of insulating building elements, energy-efficient lights, energy-efficient air conditioning, rarely efficient

HVAC setpoints and scheduling, etc. Regarding the energy-saving measures for the process itself, the audits typically recommend replacing the low-efficiency units with new highly efficient ones, use of adjustable speed drivers, occasionally process integration, or waste heat recovery. Some tips and guidelines can be found, for example, in these documents: Worrell (2010), Hesselbach (2011) and ICF Consulting Ltd (2015). However, implementation of these energy-saving measures often faces many barriers from which the main barrier is the economic barrier as demonstrated in the case of Sweden (Johansson and Thollander, 2018). Two years' simple payback period (SPP) is mostly considered to be feasible for industrial enterprises. Few of them may consider payback period up to 5 years, while anything with a higher payback period is not feasible (ICF Consulting Ltd, 2015).

ICF Consulting Ltd (2015) evaluated over 230 energy-saving opportunities with respect to the payback period. In the category of opportunities with SPP period under 2 years, integrated control systems are supposed to have the greatest energy-saving potential followed by sub-metering and interval metering (it is possible to show more measures) for energy use monitoring. However, there might be other opportunities for process efficiency improvement. These can be uncovered by applying advanced modelling, simulation, and optimization methods. Digitalisation in Industry 4.0 provides new options thanks to the availability of large amounts of operational data. Nevertheless, a procedure utilizing data for energy efficiency improvement by advanced methods must be time and cost-effective, so it satisfies the 2 years payback period rule.

2.5 Traditional and Modern Industrial Implementations

The concept of energy savings was formed in the 1970's, when the energy crisis struck the American domestic oil supply. In 1973, American's support for Israeli in the Arab-Isreali War has triggered the Arab nation to stop supplying oil to the American. With that, the oil price has tripled. This has exposed the American with the risk of the energy crisis. The concern in the energy crisis has created the momentum to increase public awareness on energy conservation (Lifset, 2017). The concept first started with the mentality of "just use less". Delmes et al. (Delmas et al., 2013) highlighted that energy conservation is highly consumer behaviour dependent. The reluctance in conserving energy is mainly due to the lack of information, ease of convenience and lack of immediate result (Volyand et al., n.d.). However, the constant

increase in energy demand has led the global community to extend energy conservation to energy efficiency.

Even though energy generation has contributed up to 72% of the global greenhouse gas (GHG) emission (C2ES, 2019), the EIA has projected that the global energy demand will continue to increase between 2018 to 2050 (Kahan, n.d.). Borozan (2013) also highlighted the importance of energy contributing to the global economy. As such, the global community improves energy consumption by prioritizing energy efficiency instead of energy conservation. With that, Petrecca (1993) recommended an energy management strategy of reducing the energy consumption per unit of products with a constant or reduce cost from the operation.

In practice, the conventional approach to energy saving is highly dependent on policy and management approach through energy audits, management commitment and standard operating procedures (Abdelaziz et al., 2011). Many countries such as Malaysia (Abd Rahman et al., 2019), South Africa (Rosenberg and Winkler, 2011), United State (Nadel et al., 2015) and European Union (Ringel and Knodt, 2018) have prioritized energy efficiency approach in order to meet the energy demand without compromising the global warming. In the European context, an energy manager training program was implemented in the in early 2000 under the EUREM initiative to increase energy efficiency in companies (European Energy Manager, n.d.). Apart from European region, Malaysia also has its energy manager program under the Energy Commission Malaysia. According to Li et al. (2010), energy management for building can contribute to 20% energy saving which reflects as 60 billion Euros of saving.

The paradigm shift in industry 4.0 has introduced data analytic with big-data (Kaile Zhou et al., 2016b) and machine learning (Gallagher et al., 2019) to enhance the energy-saving strategy. Di Orio et al. (2013) reviewed that data-driven models such as machine learning techniques can be trained to predict and plan for energy saving. Basl (2018) added that the influence of industry 4.0 encourages industries to installation IoT devices to move towards smart factories. With that, many industry players are tapping on to the potential of big data to enhance and improve productivity through data analytics. Li et al. (2016) had also compared the performance of machine learning models with traditional human estimation. The outcome had proved that the machine learning model is far superior to the human being's ability and capability when dealing with complex energy prediction. The accuracy and prediction from the machine model have outperformed the human's performance extensively. The new era of digitalisation has empowered data-driven technology to provide optimized and cost-effective

solutions that the industry can leverage on. The differences between modern and conventional approaches for energy savings can be found in Table 2.10.

Table 2.10: Comparison between conventional and modern

Property	**Conventional Approach**	**Modern Approach**
Data collection method	Human and instrumentation.	IoT-based infrastructure.
Cost	Cost is dependent on human experience, qualification and expertise (Antos, 1983).	High fixed cost due to investment cost, servicing cost and upgrading of parts (Ansari and Seidenberg, 2016).
Flexibility on task fulfilment	Humans need to be trained to achieve the flexibility to perform multi-tasking task (Ansari and Seidenberg, 2016)	High flexibility depending on computation algorithm and model (Boelaert and Ollion, 2018).
Flexibility on availability	Humans are restricted by mental and physical limitation.	High, as the algorithm can operate on 24 hours mode (Ma and Zhai, 2019).
Capacity on information processing	• Time intensive (Bhatnagar, 2018). • The ability to detect errors and corrections may not be consistent.	• Moderate to high time effectiveness in data processing (Qiu et al., 2016). • Moderate to high ability in error detection and corrections.
Ability on problem-solving (Ansari and Seidenberg, 2016)	• Solving ill-structured problems. • Managing exception conditions. • Perform collaboration to troubleshoot heuristics-based problems.	• Learning and formalizing the troubleshooting process. • Able to detect and recommend corrections based on repeated problems. • Perform predictions on standard problems based on continuous monitoring.

Table 2.10 (Cont.): Comparison between conventional and modern

Property	Conventional Approach	Modern Approach
Performance variation	Performance variance is high as it is dependent on individual capacity (Antos, 1983).	Performance variance is low.
Quality variations on decision making	Decision making is highly dependent on individual experience, qualification, problem-solving ability and competency (Antos, 1983).	Decision-making performance is highly depending on the quality of the input data. However, the quality can be improved by training the system with larger datasets (Cagala, 2017).

2.5.1 Barriers and Gaps Between Evolving Academics and Industry

The era of the information society is transitioning towards a knowledge-based society (Gill, 2002). Knowledge has become the new economic resource where knowledge transfer is important and necessary in order to commercialize knowledge through proper channels (Mueller, 2006). Bekkers and Bodas Freitas (2008) found that about 10 % of new products or processes are the contributions from academics. The industry and academic collaboration have formed a critical relationship worldwide (Ankrah and AL-Tabbaa, 2015) especially in today's knowledge-based society (Pinheiro et al., 2015). Collaboration between an organisation with different expertise and perspective can be difficult. However, the outcome can be impactful.

Many research outcomes have highlighted the importance of collaboration between industry and academics. For instance, a healthy and strong collaboration can boost innovation performance, enhance product development (George et al., 2002) and improve the novelty of a product (MingJi and Ping, 2014). A study has found that collaborative research and informal contacts with the industry were more important than contract research to establish effective knowledge exchange (Meyer-Krahmer and Schmoch, 1998). In collaborative research, both academic and industry can gain mutual benefits, especially on knowledge transfer for innovative ideas. Table 2.11 lists the benefits that both parties can leverage to achieve greater output.

Table 2.11: Advantages of academics and industry from collaboration

Benefiting Party	Advantages	Source
Academics	Access to industry funding	OECD (2015)
	Access industry equipment and patent	Barnes et al. (2002)
	Commercialization of research idea	Perkmann et al. (2013)
	Leverage on industry requirement to train future talent	Deloitte Global and the Global Business Coalition (2018)
	Access to industry insights and operation data	Sannö et al. (2019)
	Enhance R&D facility and capability	Grimpe and Hussinger (2013)
	Improves the university's reputation	van Rijnsoever et al. (2008)
	Academic aims to publish finding at reputable journals to be the outstand their competitor.	Newberg and Dunn (2002)
Industry	Leverage on expensive research infrastructure	Ankrah and AL-Tabbaa (2015)
	Access to high-quality talents	Myoken (2013)
	Access to high-technology and knowledge	Barnes et al. (2002)
	Low-risk exploration study	Wallin et al. (2014)
	Commercialization of patent and IP	Han (2017)

With regards to the mutual benefits that the academic and industry sector can leverage upon, challenges arise especially in this data-driven era (Li et al., 2016) indicated that the information generated from big-data analytics can improve decision making. With the available technology, advanced technology can analyse Big Data to obtain insight and high-value output that can bring higher value to the energy aspect of an organization. The sustainable of an organization is becoming more dependent on the organization's ability to manage big data, knowledge and information (Ngai et al., 2017). Therefore, the collaborative relationship between the industry and academic forms an important element in exploring the potential of energy saving with big-data analysis. However, the concern of data-sharing in every organization has become a barrier to establish academic collaboration and implementation of new technology (Al Nuaimi et al., 2015). To form a healthy collaboration, both parties have to establish a common understanding of possible challenges resulting from operation structure, culture and constraint (Wallin et al., 2014). The challenges and conflicts are identified to maximize collaboration output in Table 2.12.

Table 2.12: Academic and industry collaboration challenges

Barrier and challenges	Academic	Industry
Data access and handling	Energy consumption behaviour varies with industry sectors, the researchers need access to reliable real-time industry data to produce impactful outcome (O'Donovan et al., 2015).	(Kaisler et al., 2013) highlight the concerns over ownership of data where the data privacy is concerned. Industries are not technological and physically ready for big data such as upgrading IT infrastructure, developing new cultures and new employee skillset (Manyika et al., 2011). Lack of expert in data-driven technology in the organization (Liberatore et al., 2017). Many energy-related organizations are not aware of their capability and competency in handling big-data especially in electricity, oil and gas and transportation sectors (Rusitschka and Curry, 2016).
Project schedule	Academic are exploring for long term collaboration to develop, explore and validate the energy-saving model. The inflow volume of data is too huge where the researcher needs to invest in proper data storage management especially in the non-IT sector such as energy (Strohbach et al., 2016).	Industries are looking for short term outcomes to maintain their position in the market (Muscio and Vallanti, 2014). The industry expects positive research outcomes to be produced in order to be ahead of their competitors.

Table 2.12 (Cont.): Academic and industry collaboration challenges

Barrier and challenges	Academic	Industry
Relationship	Long term collaboration (Rybnicek and Königsgruber, 2019) especially on the researching funding aspect (OECD, 2015).	(Howells et al., 2012) highlighted that some industries have problems accessing the university's knowledge and information as they do not have any contact with the university. Establishment of collaboration is highly depending on the university's expertise, capability and infrastructure of laboratories that align with the industry's direction (Karlson and Callagher, 2012).
Resource	Support from Academic is restricted by the university's schedule. For example, the usage of expensive lab infrastructure is highly depending on the availability of the equipment (Ankrah and AL-Tabbaa, 2015).	Industries are looking for high availability of support from the academic (Wu, 2017) such as human resource (Myoken, 2013) and laboratory (Boardman and Bozeman, 2015). The bargaining power, financial power and setback handling from research outcome differ with the scale of the company (Rybnicek and Königsgruber, 2019).
Expectation	Funding from the industry is an important, industry data and data validation	Develop talent, technology, and commercial able solutions. Data security is lagging falling from the current research ecosystem (Strohbach et al., 2016).

The collaborations between academics and industry are very important to create new innovative solutions and data exchange. In the data-driven era, the barrier and gap between the industry and academics are mainly challenged by the lack of talent, suitable collaboration partner and the infrastructure of data processing. In order to address the challenges in energy sectors using data-driven approaches, the contextual understanding should be established. The

proper policy and agreement shall be in place to avoid unnecessary misunderstanding where data-driven policy is concerned.

2.6 The Way Forward

The future for data-driven energy savings for Industry 4.0 is bright and promising. However, certain efforts from researchers, industrialist and policymakers will certainty accelerate field developments. From the perspective of researchers, more research effort is required in addressing the full data pipeline which includes data acquisition, data cleaning, modelling, and industrial implementation. The responsibility of researchers in this field includes:

(i) Provide more low-level research and halt discrimination on “data janitor”-type research.
(ii) Focus on realistic industrial implementations instead of pseudo-theoretical problems.
(iii) Accelerate development in enabling technologies.
(iv) Carry out more interdisciplinary collaborations and communications.

As the implementation of data-driven energy-saving systems will be physically within the industrial area, industrialists also play an important role in the future development of this field. The responsibility of industrialist includes:

(i) Respect research advances from the academic world and participates in collaborations.
(ii) Provide honest feedback and problems to researchers.
(iii) Allocate funds for R&D projects and technological transitions.
(iv) Enhance information infrastructure and data collection methods within the facility.

Project economics is one of the bigger constraints for both academics and industrialists. Contrarily, the task for policymakers to encourage industrialists and researchers to further develop novel industrial systems that can contribute to global energy savings and energy efficiency elevation. The responsibility of policymakers is less in quantity, but significant in quality, which includes:

(i) Provide effective funding schemes for academic-industrial energy-saving projects.
(ii) Encourage energy audits and regulate policies to favour advanced energy-saving systems.

With concise cost and energy consideration, the development of digital twin-based infrastructure will pose to be a beneficial step in the fields of energy development for mankind. It will be thrilling to see what future technological development will be unveiled to us soon.

2.7 Conclusion

This chapter discusses that the future of digital twin-based infrastructures for data-driven energy savings remains optimistic. As SCADA system remains the *de facto* standards in typical industrial facilities, there are many industries that have incomplete data acquisition systems due to the costs of implementation. A potential solution for low-cost data acquisition with high coverage is by using IIoT sensors. However, the connection reliability of such devices needs to be improved by using 5G connections as redundancy. For this matter, more technological development is required to lower the costs of devices and shorten the roll-out timing for industrial implementations. This chapter also points out the importance of modularizing and standardizing data infrastructure during implementation. Moreover, there are many limitations in these directions, such as ensuring the reliability of sensor devices, balancing the accuracy in simulation and optimization of digital twins, bounding the complexity of the computation, and putting all the data infrastructure together within feasible investment costs. Hence, the timeliness of research in this field is critical towards its significance and relevance. In terms of digital twin modelling, data sensor integration, data security, computational and data storage services, there are already many commercial services that can support the implementation. Nevertheless, this chapter recommends that further developments in the fields of Artificial Intelligence (AI), Blockchain 3.0, 5G-enabled IIoT, Digital Twins are essential to accelerate research advances in the practical implementation of this field. In utilizing such technologies for energy saving, there is international interest in the forms of government initiative and policies (in regions such as Europe, United States, China, India, Malaysia, etc.) that can support smart energy-saving projects. Future developments in the field require close communication and collaboration between academic researchers, industrialists, and policymakers. To secure an energy-sustainable future, each party should provide responsible collaboration and contribute to their speciality. A strong symbiosis between multi-parties in a multi-disciplinary setting will contribute greatly to the success of digital twin-based infrastructures for data-driven energy savings. To conclude, the novelty of this chapter is that the current context of industrial

energy-savings was extended towards a more digitalized paradigm for smart industrial energy-saving in Industry 4.0.

Nomenclature

Abbreviations	**Definition**
AI	Artificial Intelligence
AMO	Advanced Manufacturing Office
APC	Advanced Process Control
BI	Business Intelligence
CAD	Computer-aided Design
CFD	Computational Fluid Dynamics
CHP	Combined Heat and Power Plant
CNC	Computer Numerical Control
CPPS	Cyber-Physical Production System
CPS	Cyber-Physical System
CPU	Central Processing Unit
CVD	Chemical Vapour Deposition
DaaS	Data as a Service
dApps	Decentralized Applications
EMEA	Europe, the Middle East and Africa
ERP	Enterprise Resource Planning
EU	European Union
FEA	Finite Element Analysis
GB/s	Gigabytes per seconds
GPU	Graphics Processing Unit
HMI	Human Machine Interface
HVAC	Heating, Ventilation and Air Conditioning
IaaS	Infrastructure as a Service
IIoT	Industrial Internet of Things
Industry 4.0	Forth Industrial Revolution
IoT	Internet of Things
IP	Intellectual Property

IP69k	International Electrotechnical Commission's protection rating for Dust Tight and Able to Sustain High Pressure Cleaning or Steam Jet
ISO	International Standard Organization
IT	Information Technology
KPI	Key Performance Indicator
LCA	Life Cycle Analysis
MES	Manufacturing Execution System
MYR	Malaysian Ringgit
NDRC	National Development and Reform Commission
NEA	National Energy Administration
OII	Open Innovation Intermediaries
PaaS	Platform as a Service
PASPO	Principal Component-aided Statistical Process Optimization
PAT	Perform Achieve and Trade scheme
PLC	Programmable Logic Control
PwC	PricewaterhouseCoopers
QR	Quick Response
R&D	Research and Development
RB-FEA	Reduced Basis Finite Element Analysis
RFID	Radio-frequency Identification
RTU	Remote Terminal Unit
SaaS	Software as a Service
SCADA	Supervisory Control and Data Acquisition System
Sim-to-Real	A field of artificial intelligence dealing with transferring simulation to the real world.
SPP	Standard Payback Period
SVM	Support Vector Machine
TEE	Trusted Execution Environment
TOE	Tonnes of Oil Equivalent
TPU	Tensor Processing Unit
UI	User Interface

UPS	Uninterruptable Power Supply
USD	United States Dollar (Used synonymously with "$")
UX	User Experience

References

Abd Rahman, N.A., Kamaruzzaman, S.N., Akashah, F.W., 2019. Scenario and Strategy towards Energy Efficiency in Malaysia: A Review. MATEC Web Conf. https://doi.org/10.1051/matecconf/201926602012

Abdelaziz, E.A., Saidur, R., Mekhilef, S., 2011. A review on energy saving strategies in industrial sector. Renew. Sustain. Energy Rev. 15, 150–168. https://doi.org/10.1016/j.rser.2010.09.003

Abele, E., Panten, N., Menz, B., 2015. Data collection for energy monitoring purposes and energy control of production machines, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2015.01.035

Acciona, 2019. Business as unusual [WWW Document]. Acciona. URL https://www.acciona.com/

Adenuga, O.T., Mpofu, K., Boitumelo, R.I., 2019. Energy efficiency analysis modelling system for manufacturing in the context of industry 4.0, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2019.01.002

Ahmed, E., Yaqoob, I., Hashem, I.A.T., Khan, I., Ahmed, A.I.A., Imran, M., Vasilakos, A. V., 2017. The role of big data analytics in Internet of Things. Comput. Networks 129, 459–471. https://doi.org/10.1016/j.comnet.2017.06.013

Akselos, 2019. Akselos Technology [WWW Document]. Technology. URL https://akselos.com/technology/

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., Al-Jaroodi, J., 2015. Applications of big data to smart cities. J. Internet Serv. Appl. https://doi.org/10.1186/s13174-015-0041-5

Alam, K.M., El Saddik, A., 2017. C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems. IEEE Access. https://doi.org/10.1109/ACCESS.2017.2657006

Alammary, A., Alhazmi, S., Almasri, M., Gillani, S., 2019. Blockchain-based applications in education: A systematic review. Appl. Sci. https://doi.org/10.3390/app9122400

Alonso, I.G., Fernández, M.R., Peralta, J.J., García, A.C., 2013. A Holistic Approach to Energy Efficiency Systems through Consumption Management and Big Data Analytics. Int. J. Adv. Softw.

Amazon Web Services, 2019a. Amazon EC2 [WWW Document]. EC2. URL https://aws.amazon.com/ec2/

Amazon Web Services, 2019b. Amazon S3 pricing [WWW Document]. Amaz. S3. URL https://aws.amazon.com/s3/pricing/

Andoni, M., Robu, V., Flynn, D., Abram, S., Geach, D., Jenkins, D., McCallum, P., Peacock, A., 2019. Blockchain technology in the energy sector: A systematic review of challenges and opportunities. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2018.10.014

Angelis, J., Ribeiro da Silva, E., 2019. Blockchain adoption: A value driver perspective. Bus. Horiz. https://doi.org/10.1016/j.bushor.2018.12.001

Ankrah, S., AL-Tabbaa, O., 2015. Universities-industry collaboration: A systematic review. Scand. J. Manag. https://doi.org/10.1016/j.scaman.2015.02.003

Ansari, F., Seidenberg, U., 2016. A portfolio for optimal collaboration of human and cyber physical production systems in problem-solving. Proc. 13th Int. Conf. Cogn. Explor. Learn. Digit. Age, CELDA 2016 311–314.

Antos, J.R., 1983. 4 Analysis of Labor Cost: Data Concepts and Sources, in: The Measurement of Labor Cost. pp. 153–182.

Arduino, 2016. Arduino - Home. Hardware.

Atkeson, A., Kehoe, P.J., 2001. The Transition to a New Economy After the Second Industrial Revolution. Natl. Bur. Econ. Res. Work. Pap. Ser. No. 8676. https://doi.org/10.3386/w8676

Backlund, S., Thollander, P., Palm, J., Ottosson, M., 2012. Extending the energy efficiency gap. Energy Policy 51, 392–396. https://doi.org/10.1016/j.enpol.2012.08.042

Barnes, T., Pashby, I., Gibbons, A., 2002. Effective university - Industry interaction: A multi-case evaluation of collaborative R&D projects. Eur. Manag. J. https://doi.org/10.1016/S0263-2373(02)00044-0

Basl, J., 2018. Companies on the Way to Industry 4.0 and their Readiness. J. Syst. Integr. 9, 3–6. https://doi.org/10.20470/jsi.v9i3.351

Beal, V., 2019. Mobile Cloud Storage [WWW Document]. Webopedia. URL https://www.webopedia.com/TERM/M/mobile_cloud_storage.html

Bekkers, R., Bodas Freitas, I.M., 2008. Analysing knowledge transfer channels between universities and industry: To what degree do sectors also matter? Res. Policy. https://doi.org/10.1016/j.respol.2008.07.007

Bello, O., Yang, D., Lazarus, S., Wang, X.S., Denney, T., 2017. Next generation downhole big data platform for dynamic data-driven well and reservoir management, in: Society of Petroleum Engineers - SPE Reservoir Characterisation and Simulation Conference and Exhibition, RCSC 2017. https://doi.org/10.2118/186033-ms

Bhatnagar, R., 2018. Machine Learning and Big Data Processing: A Technological Perspective and Review, in: Advances in Intelligent Systems and Computing. Springer Verlag, pp. 468–478. https://doi.org/10.1007/978-3-319-74690-6_46

BittWatt Pte. Ltd, 2019. The platform for entrepreneurs in mind & spirit. [WWW Document]. BittWatt Platf.

Block.one, 2019. EOSIO [WWW Document]. EOSIO Technol. Blockchain. URL https://eos.io/

Blocklabs, 2017. Four New Energy Innovation Projects [WWW Document]. Blockchain Energy. URL https://www.blocklab.nl/four-energy-innovation-projects-launched/

Boardman, C., Bozeman, B., 2015. Academic faculty as intellectual property in university-industry research alliances. Econ. Innov. New Technol. https://doi.org/10.1080/10438599.2014.988499

Boelaert, J., Ollion, E., 2018. The Great Regression. Machine Learning, Econometrics, and the Future of Quantitative Social Sciences.

Bolt, W., 2017. Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. J. Econ. Lit.

Bornschlegl, M., Drechsel, M., Kreitlein, S., Bregulla, M., Franke, J., 2013. A new approach to increasing energy efficiency by utilizing cyber-physical energy systems. Proc. 11th Int. Work. Intell. Solut. Embed. Syst. WISES 2013 25–32.

Borozan, D., 2013. Exploring the relationship between energy consumption and GDP: Evidence from Croatia. Energy Policy. https://doi.org/10.1016/j.enpol.2013.03.061

Brandon, D., Ph, D., Naucler, C., Evans, M., Bernard, S.A., Manning, M., 2016. Industrial Blockchain Platforms : An Exercise in Use Case Development in the Energy Industry. Int. J. Acad. Bus. World. https://doi.org/10.1017/CBO9781107415324.004

Brock, S., 2018. Can Grid+ Disintermediate Utilities? [WWW Document]. Stina.Io. URL https://www.stina.io/blog/2018/01/can-grid-actually-disrupt-energy.html

Brundage, M., Chang, Q., Wang, S., Feng, S., Xiao, G., Arinez, J., 2013. Energy savings opportunities and energy efficiency performance indicators for a serial production line, in: IFIP Advances in Information and Communication Technology. https://doi.org/10.1007/978-3-642-41266-0_37

Bureau of Energy Efficiency, 2018. Enhancing Energy Efficiency Through Industry Partnership. Minist. Power, Goverment India.

Bureau of Energy Efficiency, 2008. The Action Plan for Energy Efficiency. Minist. Power, Goverment India 20–21. https://doi.org/10.1007/978-1-4020-6581-1_16

Buterin, V., 2014. A next-generation smart contract and decentralized application platform, Etherum.

C2ES, 2019. Global Emissions. [WWW Document]. Cent. Clim. Energy Solut. CAIT. URL https://www.c2es.org/content/international-emissions/

Cagala, T., 2017. Improving Data Quality and Closing Data Gaps with Machine Learning.

Cai, L., Zhu, Y., 2015. The challenges of data quality and data quality assessment in the big data era. Data Sci. J. 14, 1–10. https://doi.org/10.5334/dsj-2015-002

CarbonX Personal Carbon Trading Inc., 2019. Blockchain for Good [WWW Document]. CarbonX. URL https://www.carbonx.ca/

Cardano, 2019. Cardano [WWW Document]. Decent. public blockchain. URL https://www.cardano.org/en/home/

Caylar, P.-L., Oliver, N., Kedar, N., 2016. Digital in industry: From buzzword to value creation. McKinsey Digit. 1–9.

Chen, H.H., Chen, S., Lan, Y., 2016. Attaining a sustainable competitive advantage in the smart grid industry of China using suitable open innovation intermediaries. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2016.03.008

Chen, S., 2011. Application of inherent safety explosion-proof technology in oil storage & transportation device. Procedia Eng. 15, 4814–4818. https://doi.org/10.1016/j.proeng.2011.08.899

Cheng, J., Chen, W., Tao, F., Lin, C.L., 2018. Industrial IoT in 5G environment towards smart manufacturing. J. Ind. Inf. Integr. 10, 10–19. https://doi.org/10.1016/j.jii.2018.04.001

Chu, X., Ilyas, I.F., Krishnan, S., Wang, J., 2016. Data cleaning: Overview and emerging challenges. Proc. ACM SIGMOD Int. Conf. Manag. Data 26-June-20, 2201–2206. https://doi.org/10.1145/2882903.2912574

Chua, W.S., Chan, J.C.L., Tan, C.P., Chong, E.K.P., Saha, S., 2020. Robust fault reconstruction for a class of nonlinear systems. Automatica 113, 108718. https://doi.org/10.1016/j.automatica.2019.108718

CONTACT software, 2019. Elements for IoT [WWW Document]. Prod. Inf. URL https://www.contact-software.com/en/media/product-information/contact-elements-for-iot/

Cordeau, S., Barrington, S., 2010. Instrumentation strategies for energy conservation in broiler barns with ventilation air solar pre-heaters. Energy Build. https://doi.org/10.1016/j.enbuild.2010.02.023

Cosgrove, J., Littlewood, J., Wilgeroth, P., 2017. Development of a holistic method to analyse the consumption of energy and technical services in manufacturing facilities, in: Smart Innovation, Systems and Technologies. https://doi.org/10.1007/978-3-319-52076-6_9

Creydt, M., Fischer, M., 2019. Blockchain and more - Algorithm driven food traceability. Food Control. https://doi.org/10.1016/j.foodcont.2019.05.019

Dai, H.N., Wang, H., Xu, G., Wan, J., Imran, M., 2019. Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. Enterp. Inf. Syst. https://doi.org/10.1080/17517575.2019.1633689

Danzi, P., Hambridge, S., Stefanovic, C., Popovski, P., 2018. Blockchain-Based and Multi-Layered Electricity Imbalance Settlement Architecture, in: 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018. https://doi.org/10.1109/SmartGridComm.2018.8587577

Dar, A.A., 2018. Cloud Computing-Positive Impacts and Challenges in Business Perspective. J. Comput. Sci. Syst. Biol. 12, 15–18. https://doi.org/10.4172/jcsb.1000294

Dawkins, V., 2019. How the energy industry is embracing cloud computing: Three key areas of success [WWW Document]. CloudTech. URL https://www.cloudcomputing-news.net/news/2019/aug/07/how-energy-industry-embracing-cloud-computing-three-key-areas-success/

Delmas, M.A., Fischlein, M., Asensio, O.I., 2013. Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. Energy Policy 61, 729–739. https://doi.org/10.1016/j.enpol.2013.05.109

Deloitte Global and the Global Business Coalition, 2018. Preparing tomorrow's workforce for the Fourth Industrial Revolution | Deloitte | About 1–58.

Deng, C., Guo, R., Liu, C., Zhong, R.Y., Xu, X., 2018. Data cleansing for energy-saving: a case of Cyber-Physical Machine Tools health monitoring system. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2017.1394596

Di Orio, G., Cândido, G., Barata, J., Bittencourt, J.L., Bonefeld, R., 2013. Energy efficiency in machine tools - A self-learning approach. Proc. - 2013 IEEE Int. Conf. Syst. Man, Cybern. SMC 2013 4878–4883. https://doi.org/10.1109/SMC.2013.830

Duan, Y., Li, W., Fu, X., Yang, L., 2016. Reliable Data Transmission Method for Hybrid Industrial Network Based on Mobile Object BT - Internet and Distributed Computing Systems,

in: Li, W., Ali, S., Lodewijks, G., Fortino, G., Di Fatta, G., Yin, Z., Pathan, M., Guerrieri, A., Wang, Q. (Eds.), . Springer International Publishing, Cham, pp. 466–476.

Durrani, M.A., Ahmad, I., Kano, M., Hasebe, S., 2018. An artificial intelligence method for energy efficient operation of crude distillation units under uncertain feed composition. Energies 11. https://doi.org/10.3390/en11112993

EASME, 2015. Energy Efficiency trends and policies in Industry. Odyssee-Mure.

Eco Power Supplies Ltd, 2019. Industrial UPS Systems [WWW Document]. UPS Syst. URL https://www.ecopowersupplies.com/industrial-ups

Elegant Energy, n.d. What are Bitcoins? [WWW Document]. Peer-to-peer Technol. URL https://www.elegant.be/be/nl/bitcoins/

Elshawi, R., Sakr, S., Talia, D., Trunfio, P., 2018. Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. Big Data Res. https://doi.org/10.1016/j.bdr.2018.04.004

Enegie, 2019. Utility connections and metering services [WWW Document]. Energy Serv. URL https://www.engie.co.uk/energy-services/utility-connections-and-metering-services/

Enercity AG, n.d. Enercity- your public utilities for Hannover and the region [WWW Document]. 2019. URL https://www.enercity.de/privatkunden/index.html

Errico, M., Tola, G., Mascia, M., 2009. Energy saving in a crude distillation unit by a preflash implementation. Appl. Therm. Eng. https://doi.org/10.1016/j.applthermaleng.2008.07.011

European Comission, 2019. The revised energy efficiency directive. Clean Energy All Eur. Packag.

European Commission, 2019. United in delivering the Energy Union and Climate Action - Setting the foundations for a successful clean energy transition. Commun. from Comm. to Eur. Parliam. Counc. Eur. Econ. Soc. Comm. Comm. Reg.

European Commission, 2017. EU 2020 target for energy efficiency [WWW Document]. Targets Dir. Rules. URL https://ec.europa.eu/energy/en/topics/energy-efficiency/targets-directive-and-rules/eu-targets-energy-efficiency#content-heading-0

European Energy Manager, n.d. EREM Project News [WWW Document]. URL https://www.energymanager.eu/en/euremnext-project/information/

European Union, 2012. Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC. Off. J. Eur. Union 55.

Eva Energy Romania, 2019. Eva Energy [WWW Document]. Electr. Power Supplier, Nat. Gas Serv. URL https://www.eva-energy.ro/

Farhan, M.N., Habib, M.A., Ali, M.A., 2018. A study and Performance Comparison of MapReduce and Apache Spark on Twitter Data on Hadoop Cluster. Int. J. Inf. Technol. Comput. Sci. https://doi.org/10.5815/ijitcs.2018.07.07

Federal Ministry for Economics Affair and Energy, 2019. Map of Industrie 4.0 Use Cases [WWW Document]. Platf. Ind. 4.0. URL https://www.plattform-i40.de/PI40/Navigation/EN/InPractice/UseCases/use-cases.html

Gallagher, C. V., Leahy, K., O'Donovan, P., Bruton, K., O'Sullivan, D.T.J., 2019. IntelliMaV: A cloud computing measurement and verification 2.0 application for automated, near real-time energy savings quantification and performance deviation detection. Energy Build. https://doi.org/10.1016/j.enbuild.2018.12.034

Gao, J., 2013. Machine Learning Applications for Data Center Optimization. Google white Pap. 1–13.

Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C., Santamaria, V., 2018. To Blockchain or Not to Blockchain: That Is the Question. IT Prof. https://doi.org/10.1109/MITP.2018.021921652

General Electric, 2019. Predix Platform [WWW Document]. IIoT Platf. URL https://www.ge.com/digital/iiot-platform

General Electric, 2016. Edge Computing: Driving New Outcomes from Intelligent Industrial Machines. White Pap. 10.

George, G., Zahra, S.A., Wood, D.R., 2002. The effects of business-university alliances on innovative output and financial performance: A study of publicly traded biotechnology companies. J. Bus. Ventur. https://doi.org/10.1016/S0883-9026(01)00069-6

Giacone, E., Mancò, S., Gabriele, P., 2008. Energy management techniques for small- and medium-sized companies (ESDA2006-95808). J. Energy Resour. Technol. Trans. ASME 130, 0120021–0120027. https://doi.org/10.1115/1.2835614

Gill, K.S., 2002. Knowledge networking in cross-cultural settings. AI Soc. https://doi.org/10.1007/s001460200021

Glaessgen, E.H., Stargel, D.S., 2012. The digital twin paradigm for future NASA and U.S. Air force vehicles, in: Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. https://doi.org/10.2514/6.2012-1818

Google Cloud, 2019. Serverless Computing [WWW Document]. App Engine. URL https://cloud.google.com/appengine/

GridPlus Energy, 2019. Shop Texas Electricity Plans for your home [WWW Document]. First App Energy. URL https://gridplus.io/

Grieves, M., 2014. Digital Twin: Manufacturing Excellence Through Virtual Factory Replication. Nc-Race 18. https://doi.org/10.5281/zenodo.1493930

Grieves, M., Vickers, J., 2016. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, in: Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. https://doi.org/10.1007/978-3-319-38756-7_4

Grimpe, C., Hussinger, K., 2013. Formal and Informal Knowledge and Technology Transfer from Academia to Industry: Complementarity Effects and Innovation Performance. Ind. Innov. https://doi.org/10.1080/13662716.2013.856620

Grueneich, D.M., 2015. The Next Level of Energy Efficiency: The Five Challenges Ahead. Electr. J. https://doi.org/10.1016/j.tej.2015.07.001

Han, J., 2017. Technology commercialization through sustainable knowledge sharing from university-industry collaborations, with a focus on patent propensity. Sustain. 9. https://doi.org/10.3390/su9101808

Hesselbach, J., 2011. A Practical Guide to Energy Efficiency in Production Processes 56.

Higginson, M., Nadeau, M.-C., Rajgopal, K., 2019. Blockchain's Occam problem.

Hou, H., 2017. The application of blockchain technology in E-government in China, in: 2017 26th International Conference on Computer Communications and Networks, ICCCN 2017. https://doi.org/10.1109/ICCCN.2017.8038519

Howells, J., Ramlogan, R., Cheng, S.L., 2012. Innovation and university collaboration: Paradox and complexity within the knowledge economy. Cambridge J. Econ. https://doi.org/10.1093/cje/bes013

Hu, J., 2017. IOTA Tangle: Introductory overview of white paper for Beginners. Hackernoon.

Huang, Y.T., Cheng, F.T., Chen, Y.T., 2006. Importance of data quality in virtual metrology, in: IECON Proceedings (Industrial Electronics Conference). https://doi.org/10.1109/IECON.2006.347318

IBM, 2019. IBM Cloud Object Storage [WWW Document]. Cloud Object Storage. URL https://www.ibm.com/cloud/object-storage/pricing

ICF Consulting Ltd, 2015. Study on Energy Efficiency and Energy Saving Potential in Industry and on.

ICON Foundation, 2019. Connecting Crypto To Real World [WWW Document]. ICON Netw. URL https://icon.foundation/?lang=en

International Energy Charter, 2018. China Energy efficiency report. Energy Charter Secretariat.

J.C.R., L., 2001. Memorandum For Members and Affiliates of the Intergalactic Computer Network [WWW Document]. Kurzweil Accel. Intell. URL

https://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network

Jain, P., Poon, J., Singh, J.P., Spanos, C., Sanders, S.R., Panda, S.K., 2020. A digital twin approach for fault diagnosis in distributed photovoltaic systems. IEEE Trans. Power Electron. https://doi.org/10.1109/TPEL.2019.2911594

Jänicke, M., Jacob, K., 2009. A third Industrial Revolution? Solutions to the crisis of resource-intensive growth. Forschungsstelle fur Umweltpolitik 35.

Januteniene, J., Lenkauskas, T., Didžiokas, R., Obsta, R., Jakovlev, S., 2012. Energy saving in industrial processes using modern data acquisition systems, in: 2012 2nd International Conference on Digital Information Processing and Communications, ICDIPC 2012. https://doi.org/10.1109/ICDIPC.2012.6257287

Ji, W., Yin, S., Wang, L., 2019. A big data analytics based machining optimisation approach. J. Intell. Manuf. https://doi.org/10.1007/s10845-018-1440-9

Johansson, M.T., Thollander, P., 2018. A review of barriers to and driving forces for improved energy efficiency in Swedish industry– Recommendations for successful in-house energy management. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2017.09.052

Jouppi, N.P., Borchers, A., Boyle, R., Cantin, P., Chao, C., Clark, C., Coriell, J., Daley, M., Dau, M., Dean, J., Gelb, B., Young, C., Ghaemmaghami, T.V., Gottipati, R., Gulland, W., Hagmann, R., Ho, C.R., Hogberg, D., Hu, J., Hundt, R., Hurt, D., Ibarz, J., Patil, N., Jaffey, A., Jaworski, A., Kaplan, A., Khaitan, H., Killebrew, D., Koch, A., Kumar, N., Lacy, S., Laudon, J., Law, J., Patterson, D., Le, D., Leary, C., Liu, Z., Lucke, K., Lundin, A., MacKean, G., Maggiore, A., Mahony, M., Miller, K., Nagarajan, R., Agrawal, G., Narayanaswami, R., Ni, R., Nix, K., Norrie, T., Omernick, M., Penukonda, N., Phelps, A., Ross, J., Ross, M., Salek, A., Bajwa, R., Samadiani, E., Severn, C., Sizikov, G., Snelham, M., Souter, J., Steinberg, D., Swing, A., Tan, M., Thorson, G., Tian, B., Bates, S., Toma, H., Tuttle, E., Vasudevan, V., Walter, R., Wang, W., Wilcox, E., Yoon, D.H., Bhatia, S., Boden, N., 2017. In-Datacenter Performance Analysis of a Tensor Processing Unit. ACM SIGARCH Comput. Archit. News. https://doi.org/10.1145/3140659.3080246

Kahan, A., n.d. EIA projects nearly 50% increase in world energy usage by 2050, led by growth in Asia.

Kaisler, S., Armour, F., Espinosa, J.A., Money, W., 2013. Big data: Issues and challenges moving forward, in: Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.1109/HICSS.2013.645

Kang, D., Robles, R.J., 2009. Compartmentalization of Protocols in SCADA Communication. Int. J. Adv. Sci. Technol. 8, 27–36.

Karlson, B., Callagher, L., 2012. Which university to partner with: An investigation into partner selection motives among small innovative firms, in: International Journal of Innovation Management. https://doi.org/10.1142/S1363919612400026

Katchasuwanmanee, K., Bateman, R., Cheng, K., 2016. Development of the Energy-smart Production Management system (e-ProMan): A Big Data driven approach, analysis and optimisation. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. 230, 972–978. https://doi.org/10.1177/0954405415586711

Khan, M.A., Salah, K., 2018. IoT security: Review, blockchain solutions, and open challenges. Futur. Gener. Comput. Syst. https://doi.org/10.1016/j.future.2017.11.022

Kim, Suwon, Kim, Seongcheol, 2016. A multi-criteria approach toward discovering killer IoT application in Korea. Technol. Forecast. Soc. Change 102, 143–155. https://doi.org/10.1016/j.techfore.2015.05.007

Klein, A., 2017. The Economics of Hybrid Cloud Storage [WWW Document]. Backblaze. URL https://www.backblaze.com/blog/hybrid-cloud-storage-economics/

Koch, V., Kuge, S., Geissbauer, R., Schrauf, S., 2014. Industry 4.0 - Opportunities and challenges of the industrial internet. Strateg. Former. Booz Company, PwC 13, 1–51. https://doi.org/10.1016/j.futures.2014.12.002

Koutsiamanis, R.A., Papadopoulos, G.Z., Fafoutis, X., Fiore, J.M. Del, Thubert, P., Montavont, N., 2018. From Best Effort to Deterministic Packet Delivery for Wireless Industrial IoT Networks. IEEE Trans. Ind. Informatics 14, 4468–4480. https://doi.org/10.1109/TII.2018.2856884

Kraft, E.M., 2016. The US air force digital thread/digital Twin – life cycle integration and use of computational and experimental knowledge. 54th AIAA Aerosp. Sci. Meet. 1–22. https://doi.org/10.2514/6.2016-0897

Krome, C., Sander, V., 2018. Time series analysis with apache spark and its applications to energy informatics. Energy Informatics. https://doi.org/10.1186/s42162-018-0043-1

Kshetri, N., 2017. Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecomm. Policy. https://doi.org/10.1016/j.telpol.2017.09.003

Kumar, S., Goudar, R.H., 2012. Cloud Computing – Research Issues, Challenges, Architecture, Platforms and Applications: A Survey. Int. J. Futur. Comput. Commun. https://doi.org/10.7763/ijfcc.2012.v1.95

Kusiak, A., 2017. Smart manufacturing must embrace big data. Nature. https://doi.org/10.1038/544023a

Lade, P., Ghosh, R., Srinivasan, S., 2017. Manufacturing analytics and industrial Internet of Things. IEEE Intell. Syst. 32, 74–79. https://doi.org/10.1109/MIS.2017.49

Lardinois, F., 2014. Google Drive Gets A Big Price Drop, 100GB Now Costs $1.99 A Month [WWW Document]. Techcrunch. URL https://techcrunch.com/2014/03/13/google-drive-gets-a-big-price-drop-100gb-now-costs-1-99-a-month/

Le, C.V., Pang, C.K., 2013. An energy data-driven decision support system for high-performance manufacturing industries. Int. J. Autom. Logist. 1, 61. https://doi.org/10.1504/ijal.2013.057453

Lee, J., Kao, H.A., Yang, S., 2014. Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP 16, 3–8. https://doi.org/10.1016/j.procir.2014.02.001

Lenz, J., Wuest, T., Westkämper, E., 2018. Holistic approach to machine tool data analytics. J. Manuf. Syst. https://doi.org/10.1016/j.jmsy.2018.03.003

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P., Ponnambalam, S.G., 2019. Adaptive analytical approach to lean and green operations. J. Clean. Prod. 235, 190–209. https://doi.org/10.1016/j.jclepro.2019.06.143

Lesjak, C., Hein, D., Winter, J., 2015. Hardware-security technologies for industrial IoT: TrustZone and security controller. IECON 2015 - 41st Annu. Conf. IEEE Ind. Electron. Soc. 2589–2595. https://doi.org/10.1109/IECON.2015.7392493

Li, K., 2019. The Blockchain Scalability Problem & the Race for Visa-Like Transaction Speed [WWW Document]. Hackernoon. URL https://hackernoon.com/the-blockchain-scalability-problem-the-race-for-visa-like-transaction-speed-5cce48f9d44

Li, M., Porter, A.L., 2019. Can nanogenerators contribute to the global greening data centres? Nano Energy. https://doi.org/10.1016/j.nanoen.2019.03.046

Li, X., Bowers, C.P., Schnier, T., 2010. Classification of energy consumption in buildings with outlier detection. IEEE Trans. Ind. Electron. https://doi.org/10.1109/TIE.2009.2027926

Li, Y., Thomas, M.A., Osei-Bryson, K.M., 2016. A snail shell process model for knowledge discovery via data analytics. Decis. Support Syst. https://doi.org/10.1016/j.dss.2016.07.003

Li, Z., Mahbobur Rahman, S.M., Vega, R., Dong, B., 2016. A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. Energies. https://doi.org/10.3390/en9010055

Liberatore, M.J., Pollack-Johnson, B., Heller Clain, S., 2017. Analytics capabilities and the decision to invest in analytics. J. Comput. Inf. Syst. https://doi.org/10.1080/08874417.2016.1232995

Lifset, R., 2017. Panic at the Pump: The Energy Crisis and the Transformation of American Politics in the 1970s. J. Am. Hist. https://doi.org/10.1093/jahist/jax143

Liu, Q., Zhang, H., Leng, J., Chen, X., 2019. Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2018.1471243

Long, S., Zhao, Y., Chen, W., 2014. A three-phase energy-saving strategy for cloud storage systems. J. Syst. Softw. https://doi.org/10.1016/j.jss.2013.08.018

Lu, B., Habetler, T.G., Harley, R.G., Gutiérrez, J.A., 2005. Applying wireless sensor networks in industrial plant energy management systems - Part I: A closed-loop scheme. Proc. IEEE Sensors 2005, 145–150. https://doi.org/10.1109/ICSENS.2005.1597657

Lu, H., Guo, L., Azimi, M., Huang, K., 2019. Oil and Gas 4.0 era: A systematic review and outlook. Comput. Ind. https://doi.org/10.1016/j.compind.2019.06.007

Luo, W., Hu, T., Zhang, C., Wei, Y., 2019. Digital twin for CNC machine tool: modeling and using strategy. J. Ambient Intell. Humaniz. Comput. https://doi.org/10.1007/s12652-018-0946-5

M-PAYG, 2017. Democratizing access to energy [WWW Document]. M-PAYG Platf. URL http://www.mpayg.com/

M, D., Biradar, N.B., 2018. IOTA-Next Generation Block chain. Int. J. Eng. Comput. Sci. https://doi.org/10.18535/ijecs/v7i4.05

Ma, Y.-J., Zhai, M.-Y., 2019. A Dual-Step Integrated Machine Learning Model for 24h-Ahead Wind Energy Generation Prediction Based on Actual Measurement Data and Environmental Factors. Appl. Sci. 9, 2125. https://doi.org/10.3390/app9102125

Manyika, J., Michael, C., Brad, B., Jacques, B., Richard, D., Charles, R., Byers, A.H., 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Glob. Inst. 156. https://doi.org/10.1080/01443610903114527

Marinakis, V., Doukas, H., Tsapelas, J., Mouzakitis, S., Sicilia, Á., Madrazo, L., Sgouridis, S., 2018. From big data to smart energy services: An application for intelligent energy management. Futur. Gener. Comput. Syst. https://doi.org/10.1016/j.future.2018.04.062

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018. Key pillars of successful energy saving projects in small and medium industrial enterprises. Energy 158, 293–304. https://doi.org/10.1016/j.energy.2018.06.018

Mattern, F., Floerkemeier, C., 2010. From the internet of computers to the internet of things. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 6462 LNCS, 242–259. https://doi.org/10.1007/978-3-642-17226-7_15

Mayilvaganan, M., Sabitha, M., 2013. A cloud-based architecture for Big-Data analytics in smart grid: A proposal, in: 2013 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2013. https://doi.org/10.1109/ICCIC.2013.6724168

Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., Weinhardt, C., 2018. Designing microgrid energy markets: A case study: The Brooklyn Microgrid. Appl. Energy. https://doi.org/10.1016/j.apenergy.2017.06.054

Meyer-Krahmer, F., Schmoch, U., 1998. Science-based technologies: University-industry interactions in four fields. Res. Policy. https://doi.org/10.1016/S0048-7333(98)00094-8

Migueláñez, E., Lane, D., 2010. Predictive diagnosis for offshore wind turbines using holistic condition monitoring, in: MTS/IEEE Seattle, OCEANS 2010. https://doi.org/10.1109/OCEANS.2010.5664584

Mihaylov, M., Jurado, S., Avellana, N., Van Moffaert, K., De Abril, I.M., Nowé, A., 2014. NRGcoin: Virtual currency for trading of renewable energy in smart grids, in: International Conference on the European Energy Market, EEM. https://doi.org/10.1109/EEM.2014.6861213

Mikšovský, P., Matoušek, K., Kouba, Z., 2002. Data pre-processing support for data mining. Proc. IEEE Int. Conf. Syst. Man Cybern. 5, 51–54. https://doi.org/10.1109/icsmc.2002.1176327

Miller, A.M.D., Alvarez, R., Hartman, N., 2018. Towards an extended model-based definition for the digital twin. Comput. Aided. Des. Appl. https://doi.org/10.1080/16864360.2018.1462569

Miller, B., Rowe, D.C., 2012. A survey of SCADA and critical infrastructure incidents. RIIT'12 - Proc. ACM Res. Inf. Technol. 51–56. https://doi.org/10.1145/2380790.2380805

MingJi, J., Ping, Z., 2014. Research on the Patent Innovation Performance of University-Industry Collaboration Based on Complex Network Analysis. J. Business-to-bus. Mark. https://doi.org/10.1080/1051712X.2014.903454

Ministry of Energy, G.T. and W., 2015. National Energy Efficiency Action Plan 1–3.

Ministry of Energy Green Technology and Water, 2017. Green Technology Master Plan 2017 - 2030, Ministry of Energy, Green Technology and Water Malaysia (KeTTHA).

Ministry of Power, 2001. The Energy Conservation Act. Gaz. India 60, 22.

Mishra, S., Karamchandani, N., Tabuada, P., Diggavi, S., 2014. Secure state estimation and control using multiple (insecure) observers, in: Proceedings of the IEEE Conference on Decision and Control. https://doi.org/10.1109/CDC.2014.7039631

Mittal, S., Khan, M.A., Romero, D., Wuest, T., 2018. A critical review of smart manufacturing & Industry 4.0 maturity models: Implications for small and medium-sized enterprises (SMEs). J. Manuf. Syst. https://doi.org/10.1016/j.jmsy.2018.10.005

Motlaghi, S., Jalali, F., Ahmadabadi, M.N., 2008. An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework. Expert Syst. Appl. 35, 1540–1545. https://doi.org/10.1016/j.eswa.2007.08.105

Mourtzis, D., Vlachou, E., Milas, N., 2016. Industrial Big Data as a Result of IoT Adoption in Manufacturing, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2016.07.038

Mueck, M., Strinati, E.C., Kim, I.G., Clemente, A., Doré, J.B., De Domenico, A., Kim, T., Choi, T., Chung, H.K., Destino, G., Pärssinen, A., Pouttu, A., Latva-Aho, M., Chuberre, N., Gineste, M., Vautherin, B., Monnerat, M., Frascolla, V., Fresia, M., Keusgen, W., Haustein, T., Korvala, A., Pettissalo, M., Liinamaa, O., 2016. 5G CHAMPION - Rolling out 5G in 2018. 2016 IEEE Globecom Work. GC Wkshps 2016 - Proc. https://doi.org/10.1109/GLOCOMW.2016.7848798

Mueller, P., 2006. Exploring the knowledge filter: How entrepreneurship and university-industry relationships drive economic growth. Res. Policy. https://doi.org/10.1016/j.respol.2006.09.023

Müller, M., Kuhlenkötter, B., Nassmacher, R., 2014. Robots in food industry challenges and chances. Proc. Jt. Conf. ISR 2014 - 45th Int. Symp. Robot. Robot. 2014 - 8th Ger. Conf. Robot. ISR/ROBOTIK 2014 232–238.

Muscio, A., Vallanti, G., 2014. Perceived Obstacles to University–Industry Collaboration: Results from a Qualitative Survey of Italian Academic Departments. Ind. Innov. https://doi.org/10.1080/13662716.2014.969935

Mylrea, M., 2019. Distributed Autonomous Energy Organizations: Next-Generation Blockchain Applications for Energy Infrastructure, in: Artificial Intelligence for the Internet of Everything. https://doi.org/10.1016/b978-0-12-817636-8.00012-0

Myoken, Y., 2013. The role of geographical proximity in university and industry collaboration: case study of Japanese companies in the UK. Int. J. Technol. Transf. Commer. https://doi.org/10.1504/ijttc.2013.064170

Naboni, R., Paoletti, I., 2015. The third industrial revolution. SpringerBriefs Appl. Sci. Technol. 7–27. https://doi.org/10.1007/978-3-319-04423-1_2

Nadel, S., Elliott, N., Langer, T., 2015. Energy Efficiency in the United States: 35 Years and Counting. Am. Counc. an Energy-Efficient Econ.

Nagata, K., 2016. Utility venture to promote cheaper electricity payments via bitcoin [WWW Document]. The Japan Times. URL https://www.japantimes.co.jp/news/2016/09/26/business/utility-venture-promote-cheaper-electricity-payments-via-bitcoin/#.Xe4e8OhKhPZ

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system, Bigcoin.

Newberg, J.A., Dunn, R.L., 2002. Keeping Secrets in The Campus Lab:Law, Values and Rules of Engagement for Industry-University R&D Partnerships. Am. Bus. Law J. https://doi.org/10.1111/j.1744-1714.2002.tb00298.x

Newmarch, J., Newmarch, J., 2017. Raspberry Pi, in: Linux Sound Programming. https://doi.org/10.1007/978-1-4842-2496-0_31

NexGen Energy Ltd., 2019. NexGen: Delivering a Generational Project for Canada and the Global Environment [WWW Document]. Corp. Present. URL https://www.nexgenenergy.ca/

Ngai, E.W.T., Gunasekaran, A., Wamba, S.F., Akter, S., Dubey, R., 2017. Big data analytics in electronic markets. Electron. Mark. 27, 243–245. https://doi.org/10.1007/s12525-017-0261-6

Nunes, J., Silva, P.D., Andrade, L.P., Domingues, C., Gaspar, P.D., 2015. Opportunities for the energy efficiency improvement in the dairy food sector - The case study of Portuguese traditional cheese industries, in: Refrigeration Science and Technology. https://doi.org/10.18462/iir.icr.2015.0177

Nunes, J., Silva, P.D., Domingues, L., Andrade, L.P., Gaspar, P.D., 2014. Energy evaluation of refrigeration systems in Portuguese fruit and vegetable industry, in: Refrigeration Science and Technology.

NVIDIA, 2019. Jetson Nano Developer Kit [WWW Document]. Auton. Mach. URL https://developer.nvidia.com/embedded/jetson-nano-developer-kit

O'Donovan, P., Leahy, K., Bruton, K., O'Sullivan, D.T.J., 2015. Big data in manufacturing: a systematic mapping study. J. Big Data. https://doi.org/10.1186/s40537-015-0028-x

OECD, 2015. OECD Science, Technology and Industry Scoreboard. Innov. growth Soc. https://doi.org/10.1787/sti_scoreboard-2015-en

Office, A.M., 2012. Advanced Manufacturing Office Industrial Technical.

Omar, A. Al, Bhuiyan, M.Z.A., Basu, A., Kiyomoto, S., Rahman, M.S., 2019. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Futur. Gener. Comput. Syst. https://doi.org/10.1016/j.future.2018.12.044

Omnisite, 2009. OmniSite vs. SCADA [WWW Document]. Omnisite docs. URL https://www.omnisite.com/docs/omnisite-vs-scada-TCO-sheet.xls

Oracle, 2019. IoT Production Monitoring [WWW Document]. Internet of Things. URL https://www.oracle.com/internet-of-things/iot-production-monitoring-cloud.html

Oses, N., Legarretaetxebarria, A., Quartulli, M., García, I., Serrano, M., 2016. Uncertainty reduction in measuring and verification of energy savings by statistical learning in manufacturing environments. Int. J. Interact. Des. Manuf. 10, 291–299. https://doi.org/10.1007/s12008-016-0302-y

Peng, X. Bin, Andrychowicz, M., Zaremba, W., Abbeel, P., 2018. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization, in: Proceedings - IEEE International Conference on Robotics and Automation. https://doi.org/10.1109/ICRA.2018.8460528

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: A review of the literature on university-industry relations. Res. Policy. https://doi.org/10.1016/j.respol.2012.09.007

Peter Harrell, 2019. 5G: National Security Concerns, Intellectual Property Issues, and the Impact on Competition and Innovation. Cent. a New Am. Secur. 1–8.

Petrecca, G., 1993. Industrial Energy Management: Principles and Applications. Springer Science & Business Media.

Pinheiro, R., Langa, P. V., Pausits, A., 2015. One and two equals three? The third mission of higher education institutions. Eur. J. High. Educ. https://doi.org/10.1080/21568235.2015.1044552

Pinto, S., Gomes, T., Pereira, J., Cabral, J., Tavares, A., 2017. IIoTEED: An Enhanced, Trusted Execution Environment for Industrial IoT Edge Devices. IEEE Internet Comput. 21, 40–47. https://doi.org/10.1109/MIC.2017.17

Popov, S., 2018. IOTA whitepaper v1.4.3. New Yorker.

Power Ledger Pty Ltd, 2019. Energy, reimagined [WWW Document]. Power Ledger. URL https://www.powerledger.io/

Preuveneers, D., Ilie-Zudor, E., 2017. The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in Industry 4.0. J. Ambient Intell. Smart Environ. 9, 287–298. https://doi.org/10.3233/AIS-170432

Prime Minister's Office, 2010. Economic Transformation Programme Report, Economic Planning Unit Malaysia.

Pylon Network, 2019. Pylon Network. The Energy-Wise Blockchain Platform [WWW Document]. Pylon Netw. Netw. URL https://pylon-network.org/

Qi, Q., Tao, F., Zuo, Y., Zhao, D., 2018. Digital Twin Service towards Smart Manufacturing, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2018.03.103

Qiu, J., Wu, Q., Ding, G., Xu, Y., Feng, S., 2016. A survey of machine learning for big data processing. EURASIP J. Adv. Signal Process. 2016, 67. https://doi.org/10.1186/s13634-016-0355-x

Radoglou Grammatikis, P.I., Sarigiannidis, P.G., Moscholios, I.D., 2019. Securing the Internet of Things: Challenges, threats and solutions. Internet of Things. https://doi.org/10.1016/j.iot.2018.11.003

Rajnai, Z., Kocsis, I., 2017. Labor market risks of industry 4.0, digitization, robots and AI. SISY 2017 - IEEE 15th Int. Symp. Intell. Syst. Informatics, Proc. 343–346. https://doi.org/10.1109/SISY.2017.8080580

Rao, S.K., Prasad, R., 2018. Impact of 5G Technologies on Industry 4.0. Wirel. Pers. Commun. 100, 145–159. https://doi.org/10.1007/s11277-018-5615-7

Reitsma, S., 2019. Cost comparison of deep learning hardware: Google TPUv2 vs Nvidia Tesla V100 [WWW Document]. Medium. URL https://medium.com/bigdatarepublic/cost-comparison-of-deep-learning-hardware-google-tpuv2-vs-nvidia-tesla-v100-3c63fe56c20f

Ringel, M., Knodt, M., 2018. The governance of the European Energy Union: Efficiency, effectiveness and acceptance of the Winter Package 2016. Energy Policy. https://doi.org/10.1016/j.enpol.2017.09.047

Ronay, A.K., Bhinge, R., 2015. Data analytics and uncertainty quantification for energy prediction in manufacturing, in: Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015. https://doi.org/10.1109/BigData.2015.7364081

Rosenberg, S.A., Winkler, H., 2011. Policy review and analysis: Energy efficiency strategy for the Republic of South Africa. J. Energy South. Africa. https://doi.org/10.17159/2413-3051/2011/v22i4a3230

Rusitschka, S., Curry, E., 2016. Big Data in the Energy and Transport Sectors. New Horizons a Data-Driven Econ. A Roadmap Usage Exploit. Big Data Eur. 1–303. https://doi.org/10.1007/978-3-319-21569-3_13

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., Harnisch, M., 2015. Pillars of Industry 4.0. Bost. Consult. Gr. 1–20. https://doi.org/10.1007/s12599-014-0334-4

Rusu, A.A., Vecerik, M., Rothörl, T., Heess, N., Pascanu, R., Hadsell, R., 2016. Sim-to-Real Robot Learning from Pixels with Progressive Nets 1–9.

Rybnicek, R., Königsgruber, R., 2019. What makes industry–university collaboration succeed? A systematic review of the literature. J. Bus. Econ. https://doi.org/10.1007/s11573-018-0916-6

Sadeghi, A.R., Wachsmann, C., Waidner, M., 2015. Security and privacy challenges in industrial Internet of Things. Proc. - Des. Autom. Conf. 2015-July, 1–6. https://doi.org/10.1145/2744769.2747942

Sajid, A., Abbas, H., Saleem, K., 2016. Cloud-Assisted IoT-Based SCADA Systems Security: A Review of the State of the Art and Future Challenges. IEEE Access 4, 1375–1384. https://doi.org/10.1109/ACCESS.2016.2549047

Salah, K., Rehman, M.H.U., Nizamuddin, N., Al-Fuqaha, A., 2019. Blockchain for AI: Review and open research challenges. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2890507

Sannö, A., Öberg, A.E., Flores-Garcia, E., Jackson, M., 2019. Increasing the Impact of Industry–Academia Collaboration through Co-Production. Technol. Innov. Manag. Rev. 9, 37–47. https://doi.org/10.22215/timreview/1232

Sato, K., Young, C., Patterson, D., 2017. An in-depth look at Google's first Tensor Processing Unit [WWW Document]. Google Cloud Big Data Mach. Learn. Blog.

ScaleOut Software, 2019. ScaleOut Digital Twin Builder [WWW Document]. Bringing Digit. Twins to Streaming Anal. URL https://www.scaleoutsoftware.com/products/digital-twin-builder/

Schueffel, P., 2018. Alternative Distributed Ledger Technologies Blockchain vs. Tangle vs. Hashgraph - A High-Level Overview and Comparison -. SSRN Electron. J. https://doi.org/10.2139/ssrn.3144241

SEE Action, 2013. Energy Audits and Retro-commissioning: State and Local Policy Design Guide and Sample Policy Language. Exist. Commer. Build. Work. Gr.

Seebo, 2019. Seebo Industry 4.0 Platform and Predictive Analytics [WWW Document]. Solut. Powered by Process. Artif. Intell. URL https://www.seebo.com/

Sequeira, H., Carreira, P., Goldschmidt, T., Vorst, P., 2014. Energy cloud: Real-time cloud-native energy management system to monitor and analyze energy consumption in multiple industrial sites. Proc. - 2014 IEEE/ACM 7th Int. Conf. Util. Cloud Comput. UCC 2014 529–534. https://doi.org/10.1109/UCC.2014.79

Shan, W.H., 2015. Report on the Compatibility of Chinese Laws and Regulations with the Energy Charter Treaty. Energy Chart. Secr.

Sharpe, K., 2018. New Project to Tackle Crypto Energy Crisis by Generating Electricity Through Waste [WWW Document]. Cointelegraph. URL https://cointelegraph.com/news/new-project-to-tackle-crypto-energy-crisis-by-generating-electricity-through-waste

Shrouf, F., Gong, B., Ordieres-Meré, J., 2017. Multi-level awareness of energy used in production processes. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2016.11.019

Shrouf, F., Ordieres, J., Miragliotta, G., 2014. Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm. IEEE Int. Conf. Ind. Eng. Eng. Manag. 2015-Janua, 697–701. https://doi.org/10.1109/IEEM.2014.7058728

Singapore Power Ltd., 2019. Powering the Nation [WWW Document]. SP Gr. URL https://www.spgroup.com.sg/home

Singh, S., 2012. Cost breakdown of Public Cloud Computing and Private Cloud Computing and Security Issues. Int. J. Comput. Sci. Inf. Technol. https://doi.org/10.5121/ijcsit.2012.4202

Sittón-Candanedo, I., Alonso, R.S., García, Ó., Muñoz, L., Rodríguez-González, S., 2019. Edge computing, iot and social computing in smart energy scenarios. Sensors (Switzerland) 19, 1–20. https://doi.org/10.3390/s19153353

Solarcoin, 2019. Produce One Megawatt Hour. Get One Free Solar Coin [WWW Document]. Incent. a solar-powered planet. URL https://solarcoin.org/

Song, M., Wang, S., 2018. Measuring environment-biased technological progress considering energy saving and emission reduction. Process Saf. Environ. Prot. 116, 745–753. https://doi.org/10.1016/j.psep.2017.08.042

Sorrell, W.H., Young, S.R., Griffin, W.E., 2017. Guidance for Renewable Energy Marketing Claims. Off. Atty. Gen. State Vermont. 1–6.

Spectral Energy, 2019. Jouliette at De Ceuvel [WWW Document]. Jouliette Energy. URL https://spectral.energy/news/jouliette-at-deceuvel/

Statt, N., 2015. Apple's cheaper iCloud storage plans are now available [WWW Document]. The Verge. URL https://www.theverge.com/2015/9/16/9341083/apple-icloud-storage-plans-cheaper

Stojkovic, B., Vujosevic, I., 2002. A compact SCADA system for a smaller size electric power system control-an fast, object-oriented and cost-effective approach. Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf. 1, 695–700.

Strohbach, M., Daubert, J., Ravkin, H., Lischka, M., 2016. Big Data Storage. New Horizons a Data-Driven Econ. A Roadmap Usage Exploit. Big Data Eur. 1–303. https://doi.org/10.1007/978-3-319-21569-3_7

Suarez, A., Kirkwood, G., 2019. Cloud Wars! Public vs Private Cloud Economics [WWW Document]. OpenStack Found. URL https://www.youtube.com/watch?v=oYK9y8rEhKI

Sun, H., Guo, Q., Zhang, B., Wu, W., Wang, B., Shen, X., Wang, J., 2018. Integrated Energy Management System: Concept, Design, and Demonstration in China. IEEE Electrif. Mag. https://doi.org/10.1109/MELE.2018.2816842

Sung, T.K., 2018. Industry 4.0: A Korea perspective. Technol. Forecast. Soc. Change 132, 40–45. https://doi.org/10.1016/j.techfore.2017.11.005

Supriadi, B., Haryadi, S., 2017. An academic study of roadmap of 5G implementation in Indonesia. Proceeding 2016 10th Int. Conf. Telecommun. Syst. Serv. Appl. TSSA 2016 Spec. Issue Radar Technol. 1–5. https://doi.org/10.1109/TSSA.2016.7871067

Swords, B., Coyle, E., Norton, B., 2008. An enterprise energy-information system. Appl. Energy 85, 61–69. https://doi.org/10.1016/j.apenergy.2007.06.009

Sykes, N., 2019. The Cloud's Impact on the Energy Industry [WWW Document]. EnergyCentral. URL https://www.energycentral.com/c/iu/clouds-impact-energy-industry

Tao, F., Qi, Q., 2019. Make more digital twins. Nature 573, 490–491. https://doi.org/10.1038/d41586-019-02849-1

Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S.C.Y., Nee, A.Y.C., 2019. Digital twin-driven product design framework. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2018.1443229

TechNative, 2016. Study finds 94% of businesses in EMEA use cloud-based IT services [WWW Document]. Enterprise. URL https://www.technative.io/study-finds-94-of-businesses-in-emea-use-cloud-based-it-services/

Teng, Sin Yong, How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2019.03.272

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. 225. https://doi.org/10.1016/j.jclepro.2019.03.272

The Apache Software Foundation, 2019a. Apache Hadoop [WWW Document]. Hadoop Proj. URL https://hadoop.apache.org/

The Apache Software Foundation, 2019b. Apache Spark: Lightning-fast unified analytics engine [WWW Document]. Apache Spark. URL https://spark.apache.org/

The Share&Charge Foundation, 2019. Seamless, Smart and Green Charging [WWW Document]. Emiss. Mobil. as a Serv. URL https://shareandcharge.com/

Tian, J., Shi, H., Li, X., Chen, L., 2012. Measures and potentials of energy-saving in a Chinese fine chemical industrial park. Energy. https://doi.org/10.1016/j.energy.2012.08.003

Touš, M., Pavlas, M., Putna, O., Stehlík, P., Crha, L., 2015. Combined heat and power production planning in a waste-to-energy plant on a short-term basis. Energy 90, 137–147. https://doi.org/10.1016/j.energy.2015.05.077

Tripathi, G., Ahad, M.A., Paiva, S., 2019. S2HS- A blockchain based approach for smart healthcare system. Healthcare. https://doi.org/10.1016/j.hjdsi.2019.100391

Tu, C., He, X., Shuai, Z., Jiang, F., 2017. Big data issues in smart grid – A review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2017.05.134

Tuegel, E.J., Ingraffea, A.R., Eason, T.G., Spottswood, S.M., 2011. Reengineering aircraft structural life prediction using a digital twin. Int. J. Aerosp. Eng. https://doi.org/10.1155/2011/154798

Tuo, J., Liu, F., Liu, P., Zhang, H., Cai, W., 2018. Energy efficiency evaluation for machining systems through virtual part. Energy. https://doi.org/10.1016/j.energy.2018.06.096

Uhlemann, T.H.J., Lehmann, C., Steinhilper, R., 2017. The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2016.11.152

US Department of Energy, 2016. Federal Energy Management Program. Small Bus. Progr.

US Department of Energy, 2011. State Energy Program and Energy Efficiency and Conservation Block Grant Program. Energy Effic. Renew. Energy 24.

van Rijnsoever, F.J., Hessels, L.K., Vandeberg, R.L.J., 2008. A resource-based view on the interactions of university researchers. Res. Policy. https://doi.org/10.1016/j.respol.2008.04.020

Vidal, J., 2017. 'Tsunami of data' could consume one fifth of global electricity by 2025 [WWW Document]. Clim. Home NEws. URL https://www.climatechangenews.com/2017/12/11/tsunami-data-consume-one-fifth-global-electricity-2025/

Volyand, A.F., Woodman, P., Hook, C., Reece, M., Fagan, J.M., n.d. The Importance of Energy Conservation : Raising Awareness in today ' s youth . Rutgers students attend local middle school " Rabbi Pesach Ramon Yeshiva " to inform students on how to conserve energy .

Vukolić, M., 2016. The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-39028-4_9

Wallin, J., Isaksson, O., Larsson, A., Elfström, B.O., 2014. Bridging the gap between university and industry: Three mechanisms for innovation efficiency. Int. J. Innov. Technol. Manag. https://doi.org/10.1142/S0219877014400057

Wang, Yu Emma, Wei, G.-Y., Brooks, D., 2019. Benchmarking TPU, GPU, and CPU Platforms with Deep Learning Models. ASPLOS-sbm.

Wang, J., Ye, L., Gao, R.X., Li, C., Zhang, L., 2019. Digital Twin for rotating machinery fault diagnosis in smart manufacturing. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2018.1552032

Wang, J.F., Xue, J., Feng, Y., Li, S.Q., Fu, Y., Chang, Q., 2016. Active energy saving strategy for sensible manufacturing systems operation based on real time production status. IEEE Int. Conf. Ind. Eng. Eng. Manag. 2016-Decem, 1001–1005. https://doi.org/10.1109/IEEM.2016.7798028

Wang, Q., Zhu, X., Ni, Y., Gu, L., Zhu, H., 2019. Blockchain for the IoT and industrial IoT: A review. Internet of Things. https://doi.org/10.1016/j.iot.2019.100081

Wang, X.V., Wang, L., 2019. Digital twin-based WEEE recycling, recovery and remanufacturing in the background of Industry 4.0. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2018.1497819

Wei, M., Hong, S.H., Alam, M., 2016. An IoT-based energy-management platform for industrial facilities. Appl. Energy. https://doi.org/10.1016/j.apenergy.2015.11.107

Wei, Y., Blake, M.B., 2010. Service-oriented computing and cloud computing: Challenges and opportunities. IEEE Internet Comput. https://doi.org/10.1109/MIC.2010.147

WePower, 2019. Next generation green energy procurement and trading platform [WWW Document]. WePower Platf. URL https://wepower.network/

Weyer, S., Schmitt, M., Ohmer, M., Gorecky, D., 2015. Towards industry 4.0 - Standardization as the crucial challenge for highly modular, multi-vendor production systems. IFAC-PapersOnLine 28, 579–584. https://doi.org/10.1016/j.ifacol.2015.06.143

Wisniak, J., 2018. James Watt. The steam engine. Educ. Química 18, 323. https://doi.org/10.22201/fq.18708404e.2007.4.65879

Wolff, G.B., 2018. How should the EU position itself in a global trade war? Intereconomics 53, 50–51. https://doi.org/10.1007/s10272-018-0719-6

World Economic Forum, 2018. Insight Report: Readiness for the Future of Production Report 2018, World Economic Forum.

Worrell, E., 2010. Managing Your Energy An ENERGY STAR ® Guide for Identifying Energy Savings in Manufacturing Plants. Environ. Prot.

Wriggers, P., 2008. Nonlinear finite element methods, Nonlinear Finite Element Methods. https://doi.org/10.1007/978-3-540-71001-1

Wu, H., 2017. The Essentiality of Sustainability and Variety for Industry Collaborations with University Partners. Int. J. Adv. Corp. Learn. https://doi.org/10.3991/ijac.v10i2.7272

Wu, J., Tran, N.K., 2018. Application of blockchain technology in sustainable energy systems: An overview. Sustain. https://doi.org/10.3390/su10093067

Xiaojun, C., Xianpeng, L., Peng, X., 2015. IOT-based air pollution monitoring and forecasting system. 2015 Int. Conf. Comput. Comput. Sci. ICCCS 2015 257–260. https://doi.org/10.1109/ICCACS.2015.7361361

Xu, Wei, Liu, Q., Xu, Wenjun, Zhou, Z., Pham, D.T., Lou, P., Ai, Q., Zhang, X., Hu, J., 2017. Energy Condition Perception and Big Data Analysis for Industrial Cloud Robotics, in: Procedia CIRP. https://doi.org/10.1016/j.procir.2016.11.164

Xue, C.T.S., Xin, F.T.W., 2016. Benefits and Challenges of the Adoption of Cloud Computing in Business. Int. J. Cloud Comput. Serv. Archit. https://doi.org/10.5121/ijccsa.2016.6601

Yan, J., Meng, Y., Lu, L., Li, L., 2017. Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance. IEEE Access 5, 23484–23491. https://doi.org/10.1109/ACCESS.2017.2765544

Yang, S.Y., 2013. A novel cloud information agent system with Web service techniques: Example of an energy-saving multi-Agent system. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2012.09.025

Yuan, J., Chen, M., Jiang, T., Li, T., 2017. Complete tolerance relation based parallel filling for incomplete energy big data. Knowledge-Based Syst. 132, 215–225. https://doi.org/10.1016/j.knosys.2017.06.027

Zhang, C., Wu, J., Long, C., Cheng, M., 2017. Review of Existing Peer-to-Peer Energy Trading Projects, in: Energy Procedia. https://doi.org/10.1016/j.egypro.2017.03.737

Zhang, L., Li, Z., Królczyk, G., Wu, D., Tang, Q., 2019. Mathematical modeling and multi-attribute rule mining for energy efficient job-shop scheduling. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2019.118289

Zhang, Y., Ma, S., Yang, H., Lv, J., Liu, Y., 2018. A big data driven analytical framework for energy-intensive manufacturing industries. J. Clean. Prod. 197, 57–72. https://doi.org/10.1016/j.jclepro.2018.06.170

Zhao, J.L., Fan, S., Yan, J., 2016. Overview of business innovations and research opportunities in blockchain and introduction to the special issue. Financ. Innov. https://doi.org/10.1186/s40854-016-0049-2

Zhao, Y., Liu, Y., Tian, A., Yu, Y., Du, X., 2019. Blockchain based privacy-preserving software updates with proof-of-delivery for Internet of Things. J. Parallel Distrib. Comput. https://doi.org/10.1016/j.jpdc.2019.06.001

Zheng, J., Simplot-ryl, D., Mouftah, H.T., 2011. The internet of things [Guest Editorial]. IEEE Commun. Mag. 49, 30–31. https://doi.org/10.1109/MCOM.2011.6069706

Zheng, Z., Xie, S., Dai, H., Chen, X., Wang, H., 2017. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends, in: Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017. https://doi.org/10.1109/BigDataCongress.2017.85

Zhou, Kaile, Fu, C., Yang, S., 2016a. Big data driven smart energy management: From big data to big insights. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2015.11.050

Zhou, Kaile, Fu, C., Yang, S., 2016b. Big data driven smart energy management: From big data to big insights. Renew. Sustain. Energy Rev. 56, 215–225. https://doi.org/10.1016/j.rser.2015.11.050

Zhou, Keliang, Liu, T., Zhou, L., 2016. Industry 4.0: Towards future industrial opportunities and challenges. 2015 12th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2015 2147–2152. https://doi.org/10.1109/FSKD.2015.7382284

Zhou, N., 2008. Status of China's Energy Efficiency Standards and Labels for Appliances and International Collaboration. Ernest Orlando Lawrence Berkeley Natl. Lab.

Zhu, X., Wu, X., 2005. Cost-constrained data acquisition for intelligent data preparation. IEEE Trans. Knowl. Data Eng. 17, 1542–1556. https://doi.org/10.1109/TKDE.2005.176

Zou, J., Chang, Q., Arinez, J., Xiao, G., 2017. Data-driven modeling and real-time distributed control for energy efficient manufacturing systems. Energy 127, 247–257. https://doi.org/10.1016/j.energy.2017.03.123

Zuo, Y., Tao, F., Nee, A.Y.C., 2018. An Internet of things and cloud-based approach for energy consumption evaluation and analysis for a product. Int. J. Comput. Integr. Manuf. https://doi.org/10.1080/0951192X.2017.1285429

CHAPTER 3 A GUIDING FRAMEWORK FOR PROCESS IMPROVEMENT

Abstract: With growing demands for data-driven process improvement in various energy-intensive industries, the existence of a guide framework for generic process improvement is of high research interest. In this chapter, this work proposes a novel framework for the purpose of guiding industrialist to perform data-driven process improvement with different industrial requirement. The guide framework strongly considers data-driven approaches and surrogate models that are further discussed in the subsequent chapters. These approaches include novel modelling techniques from one-shot learning, neuro-evolution, dimension reduction-based optimization, bottleneck tree analysis, hierarchical temporal memory, and other machine learning techniques. The proposed framework also considers the availability of industrial data, number of units, the requirement of optimization or debottlenecking, real-time management, or decision-making computation. The guide framework functions to guide process improvement engineers in performing operational and design improvements with different objectives and purposes.

Keywords: Process Improvement, Framework, Process System Engineering, Data-driven Approaches, Industrial Management

3.1 Introduction

With the global initiative of Industrial 4.0, various firms and companies are repositioning themselves into utilizing data-driven analytics for process improvement. The requirement of a guide framework for different situation of process improvement is becoming desirable. Guiding framework for process improvement are conventionally using model-based approach. For example, Halim and Srinivasan (2011) proposed a simulation-optimization framework as a decision-support system using Gensym's G2 and Aspen Hysys. An mixed-integer nonlinear optimization (MINLP)-based model was also proposed for simultaneous process synthesis of heat and power of industrial facility (Baliban et al., 2011). Moreover, environmental impacts via lifecycle assessment (LCA) were also highlighted in model-based process optimization frameworks (Pieragostini et al., 2012). Besides, there is also very significant number of work being carried out using superstructure optimization approaches for process improvement (Mencarelli et al., 2020). However, with the development of data-driven analytics in this era of

Big Data (Ge, 2017; Kang et al., 2016; Lueth et al., 2016), many weaknesses of model-based approaches are being exposed during implementation. Such short-coming of model-based approaches include the requirements to be parameterized by experts, difficult to consider all the knowledge representation, inefficiencies during testing, problems related to user acceptance, and challenges for maintenance (Bell, 1985; Chu and You, 2015).

Recently, data-driven approaches has become the spotlight of interest due to their modelling flexibility, less reliance on human bias, good prediction rate, and self-learning possibilities (Ge, 2017). Nevertheless, the reliance of data-driven approaches on abundance of accurate data cannot be avoided. This condition is not always provided in industrial plants (Máša et al., 2018), where many small-and-medium enterprises cannot afford to install SCADA systems. Many data infrastructure problems exists in the this case, leading to unstructured data, inconsistencies in data, errors and other problems (Teng et al., 2021). Thus, it is of great research interest to combine advantages from both fields of model-based approach and data-driven approaches for a better process improvement paradigm (Tidriri et al., 2016).

In this work, the tasks of process improvement are classified as single unit optimization, multiple-unit optimization, debottlenecking, production management and high-level decision-making. For optimization of single units, El-Mounayri et al. (2005) used a neural network-based approach to optimize the process parameters. Large research effort was also carried out to optimize single units of solar energy panel (Elsheikh et al., 2019). Bio-systems have also received more research attention for single-unit optimization due to large uncertainties and nonlinearity in biological interactions (Desai et al., 2008; Yang et al., 2020). For multi-unit process optimization, data-driven approaches were deployed on hard-to-model process. For example, Bayesian network was used to carry out plant-wide optimization of flotation industrial processes (Yan et al., 2020). A neuro-fuzzy approach was also demonstrated to perform multi-unit process optimization in a Zinc hydrometallurgical plant (Xie et al., 2020). Moreover, dealing with process bottlenecks have also been of great interest in recent years. For example, a heuristic-based framework was proposed to debottleneck a palm-oil refinery (Kasivisvanathan et al., 2014). A level-by-level debottlenecking approach was also applied in retrofitting a hydrogen network (Zhang et al., 2001) in a refinery operation. A prognostic algorithm was also proposed (Subramaniyan et al., 2019) to identify improvement measures on process bottlenecks. In production management, recent research focuses on monitoring, control, optimization and autonomy on the enterprise resource planning (ERP) and manufacturing

execution system (MES) (Moeuf et al., 2018). Various data-driven approaches such as using adaptive expert systems (Leong et al., 2019) and ensemble solver (Leong et al., 2020). Furthermore, recent challenges in production management are focused on real-time management. For example, a real-time management system was developed for a combined heat and power (CHP) generator coupled with renewable sources in Italy (Rossi et al., 2016). Real-time energy management was also implemented with flexible assessment in a CHP plant with energy storage to produce up to 7 % energy savings (De Rosa et al., 2018). For high-level decision-making in process plants, techno-economic analysis (TEA) plays a key role in the area (Mezher et al., 2011; Moncada et al., 2013). Moreover, TEA can demonstrate the profitability of various processing technologies at a regional level (Cristóbal et al., 2018).

In short, this work concludes the following points from literature review:

- The transition towards Industry 4.0 has spiked interest in using data analytics for process improvement.
- Such application of process improvement requires a generic framework to guide industrialist on different applications and implementations.
- Model-driven approach and data-driven approaches should be used together to leverage advantageous from both sides.
- The tasks of process improvement can be specifically classified as single unit optimization, multiple-unit optimization, debottlenecking, production management and high-level decision-making.

Thereby, this work proposes a guiding framework for process improvement following this structure based on the above research gap. Chapter 3.2 will discuss which company personnel and the guiding framework.

3.2 Conceptual Framework

A typical hierarchy of a process or manufacturing plant includes unit technician at the bottom level, process system engineers, production or plant manager and shareholder (and possibly policy makers). As discussed in Chapter 3.1, the classification of tools within this work will consider (i) single unit optimization, (ii) multiple-unit optimization, (ii) process debottlenecking, (iii) production management analytics, and (iv) high-level decision-making analysis. Figure 3.1 shows the tool classification associated with the company personnel that is suitable for implementation of such process improvement projects.

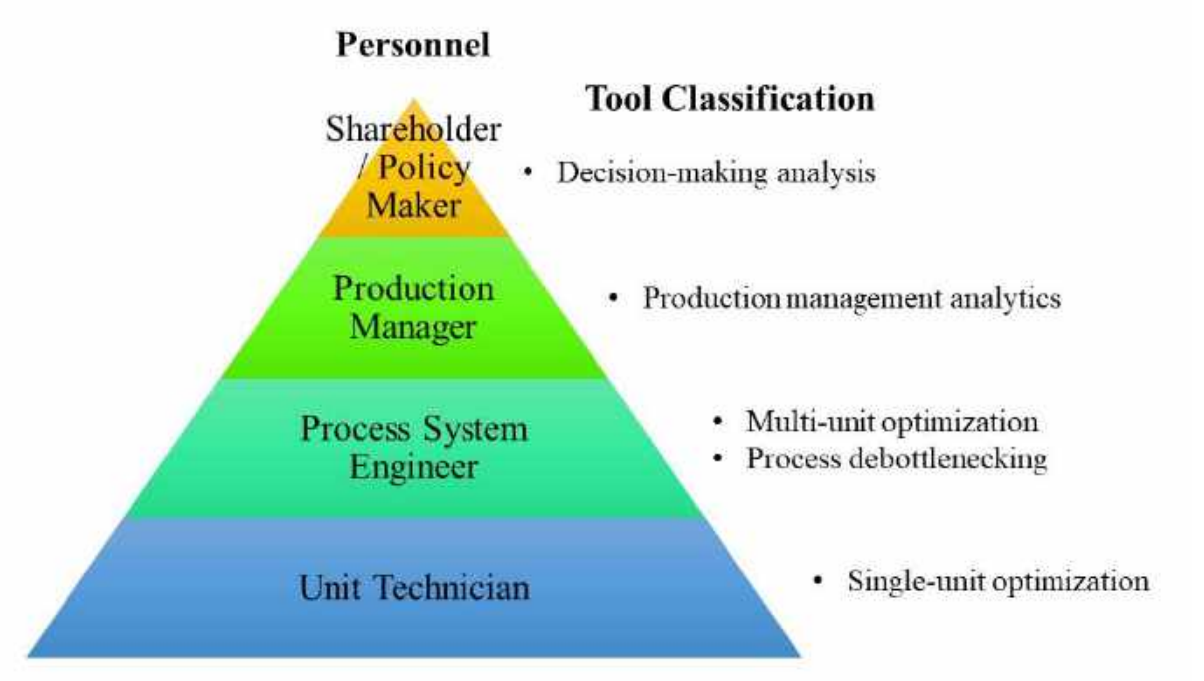


Figure 3.1: Association of company or institution personnel and process improvement tools

With understanding of the hierarchy of the company personnel and the tool classification, this work proposed a novel framework to carry out process improvement in existing plants (see Figure 3.2). The framework is flexible and serves as a guide for general process improvement. To start, the historical data availability of the target facility is evaluated. If insufficient data is available, one-shot learning is deployed to access prediction of process unit performances under data unavailability. If data availability is not an issue, data cleaning, structuring, and pre-processing would be carried out. This step varies case-by-case, but the general procedure is to remove missing or error data, change or combine the data structure to be easily processed by the algorithms, and some redundant variable removal. Next, the person-in-charge of the target facility should decide to either optimize only a single unit or multiple units. In an industrial project, this may be defined in the agreement and the project scope. If a single unit is chosen, neuro-evolution method is recommended to fully utilize computation capability and data, producing very accurate data-driven results. Otherwise, if the optimization of multiple units is required, in most cases it is not realistic to provide solutions which involves all (e.g. 1000 or more) operating variable changes in the existing facility. Therefore, this work proposed using dimension reduction to reduce the variable space and carry out optimization only on the critical variables.

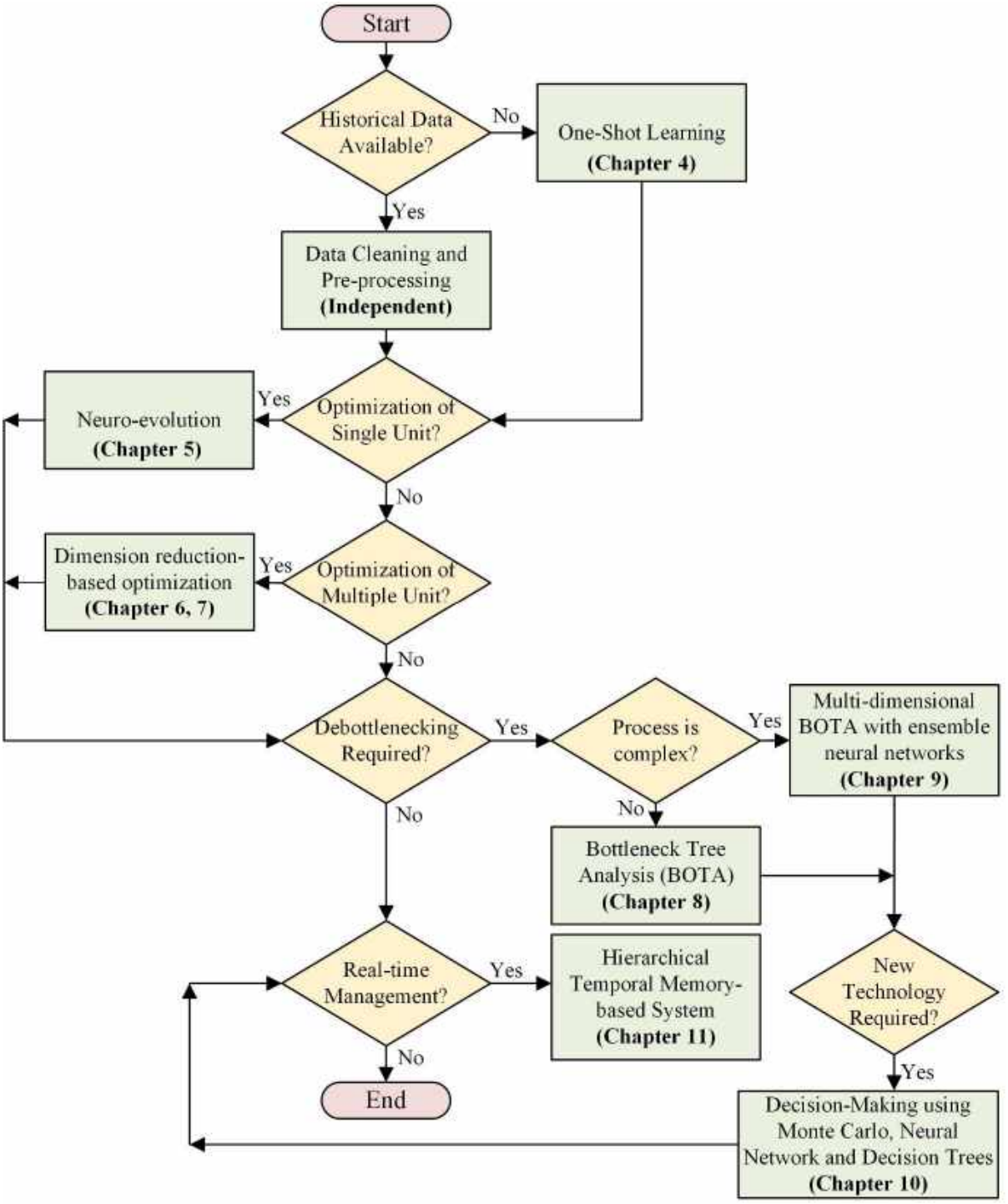


Figure 3.2: A guiding procedure for carrying out process improvement

Subsequently, if capacity debottlenecking is required, it should always be carried out after the operating variables in the process have been optimized. In this case, the process should be checked to either be complex or not. If the process contains distinct area of integrity of more than 2, then it is considered complex. For a non-complex process, the conventional bottleneck tree analysis (BOTA) is sufficient for debottlenecking study. Otherwise, a multi-dimensional

BOTA with ensemble neural networks will be used. Further on, if new technology is required, decision-making should be carried out using Monte Carlo, neural network, and decision trees to concisely evaluate the techno-economics of the process. Lastly, if the process involves real-time management, a hierarchical temporal memory-based system is proposed to be implemented.

For this work, one advantage of the data-driven and AI approach that was used is that it is process-agnostic (i.e. does not depends the nature of process of being implemented). Therefore, the focus of preparing the case study was targeted for the applicability on realistic case studies. Furthermore, in realistic industrial case studies, it is rare to have one single company willing to try out the full framework with all the methods due to limitation in operation or investment. In most cases, industrialist would find certain parts of the framework particularly suitable for their facility and recommend the use of a specific method for implementation. Therefore, many distinct company case studies were presented in this work (see Figure 3.3). This also provides a demonstration that methods in the guiding framework is process-agnostic and works on all generic processes.

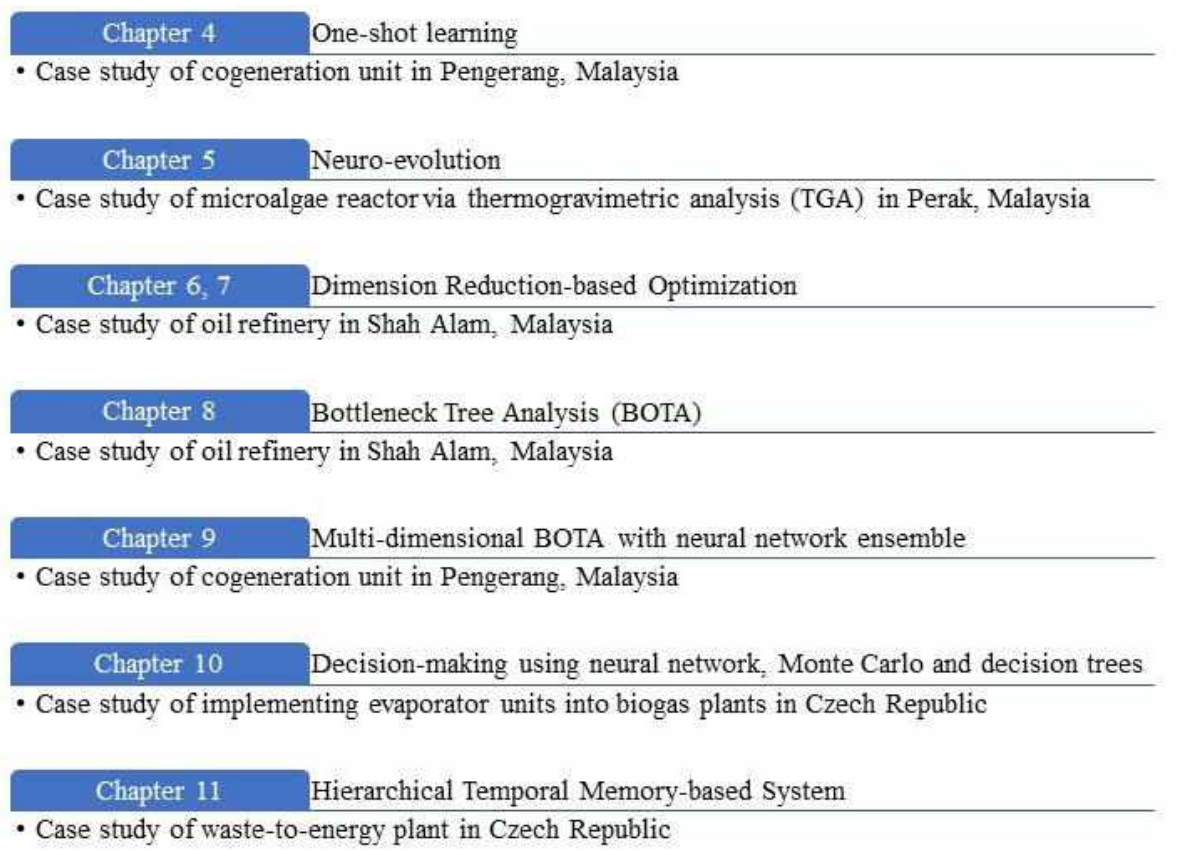


Figure 3.3: Realistic case study of each chapters of study

3.3 Conclusion

In this chapter, a guiding framework for process improvement was proposed with details in the subsequent chapters. The guiding framework mainly consist of AI and data-driven approaches for the purpose of process optimization, debottlenecking, decision-making, and real-time management. Each classification of the process improvement tools was associated to different personnel in the company or institute for more effective implementation. All parts of the framework were demonstrated in distinct case studies (i) to highlight the process-agnostic feature of the framework, (ii) to focus on the applicability of the framework on realistic industrial case studies, (iii) and to consider the feasibility of implementing process improvement in realistic case study. The case studies consist of the cogeneration industry, microalgae industry, oil refining industry, biogas industry, and waste-to-energy industry. These case studies are mainly located in Malaysia or Czech Republic. The methods that were briefly described in the proposed framework in this chapter will be discussed in detailed in the subsequent Chapter 4 to 9.

References

Baliban, R.C., Elia, J.A., Floudas, C.A., 2011. Optimization framework for the simultaneous process synthesis, heat and power integration of a thermochemical hybrid biomass, coal, and natural gas facility. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2011.01.041

Bell, M.Z., 1985. Why expert systems fail. J. Oper. Res. Soc. 36, 613–619. https://doi.org/10.1057/jors.1985.106

Chu, Y., You, F., 2015. Model-based integration of control and operations: Overview, challenges, advances, and opportunities. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2015.04.011

Cristóbal, J., Caldeira, C., Corrado, S., Sala, S., 2018. Techno-economic and profitability analysis of food waste biorefineries at European level. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2018.03.016

De Rosa, M., Carragher, M., Finn, D.P., 2018. Flexibility assessment of a combined heat-power system (CHP) with energy storage under real-time energy price market framework. Therm. Sci. Eng. Prog. https://doi.org/10.1016/j.tsep.2018.10.002

Desai, K.M., Survase, S.A., Saudagar, P.S., Lele, S.S., Singhal, R.S., 2008. Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation

media optimization: Case study of fermentative production of scleroglucan. Biochem. Eng. J. https://doi.org/10.1016/j.bej.2008.05.009

El-Mounayri, H., Kishawy, H., Briceno, J., 2005. Optimization of CNC ball end milling: A neural network-based model. J. Mater. Process. Technol. https://doi.org/10.1016/j.jmatprotec.2004.07.097

Elsheikh, A.H., Sharshir, S.W., Abd Elaziz, M., Kabeel, A.E., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: A comprehensive review. Sol. Energy. https://doi.org/10.1016/j.solener.2019.01.037

Ge, Z., 2017. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemom. Intell. Lab. Syst. https://doi.org/10.1016/j.chemolab.2017.09.021

Halim, I., Srinivasan, R., 2011. A knowledge-based simulation-optimization framework and system for sustainable process operations. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2010.08.004

Kang, H.S., Lee, J.Y., Choi, S., Kim, H., Park, J.H., Son, J.Y., Kim, B.H., Noh, S. Do, 2016. Smart manufacturing: Past research, present findings, and future directions. Int. J. Precis. Eng. Manuf. - Green Technol. 3, 111–128. https://doi.org/10.1007/s40684-016-0015-5

Kasivisvanathan, H., Tan, R.R., Ng, D.K.S., Abdul Aziz, M.K., Foo, D.C.Y., 2014. Heuristic framework for the debottlenecking of a palm oil-based integrated biorefinery. Chem. Eng. Res. Des. 92, 2071–2082. https://doi.org/10.1016/j.cherd.2014.02.024

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P., Ponnambalam, S.G., 2019. Adaptive analytical approach to lean and green operations. J. Clean. Prod. 235, 190–209. https://doi.org/10.1016/j.jclepro.2019.06.143

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Rahman, A.A., Tan, C.P., Ponnambalam, S.G., Lam, H.L., 2020. Enhancing the adaptability: Lean and green strategy towards the Industry Revolution 4.0. J. Clean. Prod. 273, 122870. https://doi.org/10.1016/j.jclepro.2020.122870

Lueth, K.L., Patsioura, C., Williams, Z.D., Kermani, Z.Z., 2016. Industrial Analytics 2016/2017: The current state of data analytics usage in industrial companies. IoT Anal. 58.

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018. Key pillars of successful energy saving projects in small and medium industrial enterprises. Energy 158, 293–304. https://doi.org/10.1016/j.energy.2018.06.018

Mencarelli, L., Chen, Q., Pagot, A., Grossmann, I.E., 2020. A review on superstructure optimization approaches in process system engineering. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2020.106808

Mezher, T., Fath, H., Abbas, Z., Khaled, A., 2011. Techno-economic assessment and environmental impacts of desalination technologies. Desalination. https://doi.org/10.1016/j.desal.2010.08.035

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., Barbaray, R., 2018. The industrial management of SMEs in the era of Industry 4.0. Int. J. Prod. Res. https://doi.org/10.1080/00207543.2017.1372647

Moncada, J., El-Halwagi, M.M., Cardona, C.A., 2013. Techno-economic analysis for a sugarcane biorefinery: Colombian case. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2012.08.137

Pieragostini, C., Mussati, M.C., Aguirre, P., 2012. On process optimization considering LCA methodology. J. Environ. Manage. https://doi.org/10.1016/j.jenvman.2011.10.014

Rossi, I., Banta, L., Cuneo, A., Ferrari, M.L., Traverso, A.N., Traverso, A., 2016. Real-time management solutions for a smart polygeneration microgrid. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2015.12.026

Subramaniyan, M., Skoogh, A., Sheikh Muhammad, A., Bokrantz, J., Turanoğlu Bekar, E., 2019. A prognostic algorithm to prescribe improvement measures on throughput bottlenecks. J. Manuf. Syst. https://doi.org/10.1016/j.jmsy.2019.07.004

Teng, S.Y., Touš, M., Leong, W.D., How, B.S., Lam, H.L., Máša, V., 2021. Recent advances on industrial data-driven energy savings: Digital twins and infrastructures. Renew. Sustain. Energy Rev. 135. https://doi.org/10.1016/j.rser.2020.110208

Tidriri, K., Chatti, N., Verron, S., Tiplica, T., 2016. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annu. Rev. Control. https://doi.org/10.1016/j.arcontrol.2016.09.008

Xie, S., Xie, Y., Ying, H., Jiang, Z., Gui, W., 2020. Neurofuzzy-Based Plant-Wide Hierarchical Coordinating Optimization and Control: An Application to Zinc Hydrometallurgy Plant. IEEE Trans. Ind. Electron. https://doi.org/10.1109/TIE.2019.2902790

Yan, H., Wang, F., He, D., Zhao, L., Wang, Q., 2020. Bayesian Network-Based Modeling and Operational Adjustment of Plantwide Flotation Industrial Process. Ind. Eng. Chem. Res. https://doi.org/10.1021/acs.iecr.9b05803

Yang, J., Huang, Y., Xu, H., Gu, D., Xu, F., Tang, J., Fang, C., Yang, Y., 2020. Optimization of fungi co-fermentation for improving anthraquinone contents and antioxidant activity using artificial neural networks. Food Chem. https://doi.org/10.1016/j.foodchem.2019.126138

Zhang, J., Zhu, X.X., Towler, G.P., 2001. A level-by-level debottlenecking approach in refinery operation. Ind. Eng. Chem. Res. 40, 1528–1540. https://doi.org/10.1021/ie990854w

CHAPTER 4 ONE-SHOT LEARNING TO MODEL PROCESS UNITS WITH LOW DATA AVAILABILITY

This chapter has been peer-reviewed and published in Chemical Engineering Transaction. Teng S.Y., Máša V., Lam H.L., Stehlík P., 2020, A One-Shot Learning Framework to Model Process Systems, Chemical Engineering Transactions, 81, 937-942.

Abstract: In the era of Big Data, the utilization of data-driven analytics for process engineering systems is rising exponentially. The abundance of data from industrial sensors and various documentation logs have served as a strong basis for such analysis. Nevertheless, there are some critical data in an industry that are simply rare and uncommon due to certain processing constraints or confidentiality. Such constraints may include economic costs for data acquisition, the complexity for data collection, the needs for qualified personnel and many other unforeseeable problems. Due to conventional data-driven approach requiring a large volume of data, such rare but critical data cannot be properly utilized. For this aspect, this work proposed a one-shot learning framework to model process systems. The novel framework utilizes prior knowledge from multi-sourced data to learn the conditional relationships of critical variables within the process. By utilizing prior generic knowledge of the system, one-shot learning can provide a better representation of the prediction space when acting as a data-driven black-box model. A combined heat and power (CHP) system is used as the case study for one-shot learning modelling which a mean squared error of 0.00616 was achieved. The efficient use of data within this framework is expected to be beneficial when modelling under high-priority and low data availability.

4.1 Introduction

In advancing sustainable and efficient processing and manufacturing, the needs of data-driven engineering analytics are essential for the transition into the Industry 4.0. The data acquisition, modelling, simulation and optimization of processing systems within small and medium (SME) industries (Máša et al., 2018) are essential for companies to move towards a digitalized future. The use of data for industrial, manufacturing and business analytics has been consolidated to provide data-enabled growth (Ritter and Pedersen, 2020). Brynjolfsson and McElheran (2016) discussed that US manufacturing is transitioning towards a data-driven paradigm for gain productivity in managerial decision-making, tracking performance and communicating of the production process. The concept of data-driven smart manufacturing (Tao et al., 2018) has

accelerated the global transition of manufacturing lifecycles towards an age of big data high, giving potential to improve manufacturing performances, process understanding and industrial management. Computational intelligence and process optimization have high potential in improving product quality and efficiency (Yin et al., 2020).

Nevertheless, the challenge with implementing data-driven analytics in the industry is on the difficulty of obtaining reliable data sources (Máša et al., 2018). The identification and collection of reliable data within manufacturing systems is a great challenge for the actual implementation of process simulation models within manufacturing industries. Moreover, from author's industrial experience, carrying out optimization on such simulation models are often tricky in which there is a great dilemma on whether to: (i) optimize the model to be more accurate or (ii) optimize the solution to give better objectives under a controlled number of samples. In some aspect, this concept is related to the well-known exploration-exploitation dilemma (Berger-Tal et al., 2014) in machine learning and optimization. In such sample-critical cases, Bayesian methods are often used to improve sampling efficiency (Baheri and Vermillion, 2017) for "expensive-to-evaluate" problems. The sampling efficiency in such methods is improved by relying on both prior and posterior, gaining statistical significance from prior data (Ghosh and Dunson, 2009). Furthermore, deep learning approaches can also be used to effectively process important industrial data for modelling and optimization (Sin Yong Teng et al., 2019f).

Despite efficient sample utilization, many manufacturing facilities face data collection difficulties related to organisation authority, machine design, information transfer methods, analytical instruments, cost of measurement etc. For example, obtaining new operational data for an oil refinery was expensive and difficult (Teng et al., 2019a). Moreover, industries such as co-generation plants (Leong et al., 2019b) have many organisational and data confidentiality challenges in providing real-time data. While other energy management systems for SME generally has poor data-acquisition systems (Máša et al., 2016). Yong et al. (2016) also demonstrated that data reconciliation is essential for successful total site integration projects. While installing new data acquisition systems such as microcontrollers (Zhang and Chen, 2008) and carrying out carefully designed experiments in laboratories (Sin Yong Teng et al., 2019d) can provide high-quality data, it is generally expensive to do as such. Furthermore, some processing equipment requires measurement devices that are expensive to operate (Hamacher et al., 2003) and manufacturing companies often turn them on only when required. Thus, this

situation gives rise to many SME companies with only one or a few critical operational data, making process modelling difficult.

One-shot learning is a field of machine learning which uses one or a few samples (sometimes known as few-shot learning) to carry out inference instead of using hundreds or thousands of samples. This concept already existed in the 1990s (Yip and Sussman, 1997), while works from Fei-Fei et al. (2006) and Miller et al. (2000) popularized the use for object detection in images. Throughout the years, the applications of one-shot learning have shown successes for face verification (Guo et al., 2011), representing human gesture (Yang Yang et al., 2013), and other mobile authentication applications. One of the most successful implementations of by using a twin neural network, which is commonly called Siamese neural network to learn the similarity of data within the prior dataset (Koch et al., 2015). The strategy of using a Siamese neural network was particularly successful even for difficult tasks such as dynamic object tracking (Guo et al., 2017) and sentence plagiarism checking (Mueller and Thyagarajan, 2016).

As such, conceptually even one or a few samples can be effectively used to model process systems by one-shot learning techniques. This work presents a novel framework to model a process system with data acquisition problems using only one or a few samples via one-shot learning techniques. The novelty of this work is that the one-shot learning technique is adapted for process manufacturing and industrial data as an alternative to the conventional field of image and sequence classification.

4.2 Method and concept

The concept of one-shot learning framework is to learn the representation of the process model from its performance database (see Figure 4.1). This performance database can be obtained from multiple manufacturers that provide a similar type of unit or even similar operating units from other facilities. The most important constraint is that the sampled unit must have the same functionality as all the units in the database, but units in the database can be of a different design model. This preserves the representation of the unit functionality. The expected result from this framework is to obtain a relatively accurate data-driven model of the desired performance of the sampled unit by lending knowledge from the aforementioned database.

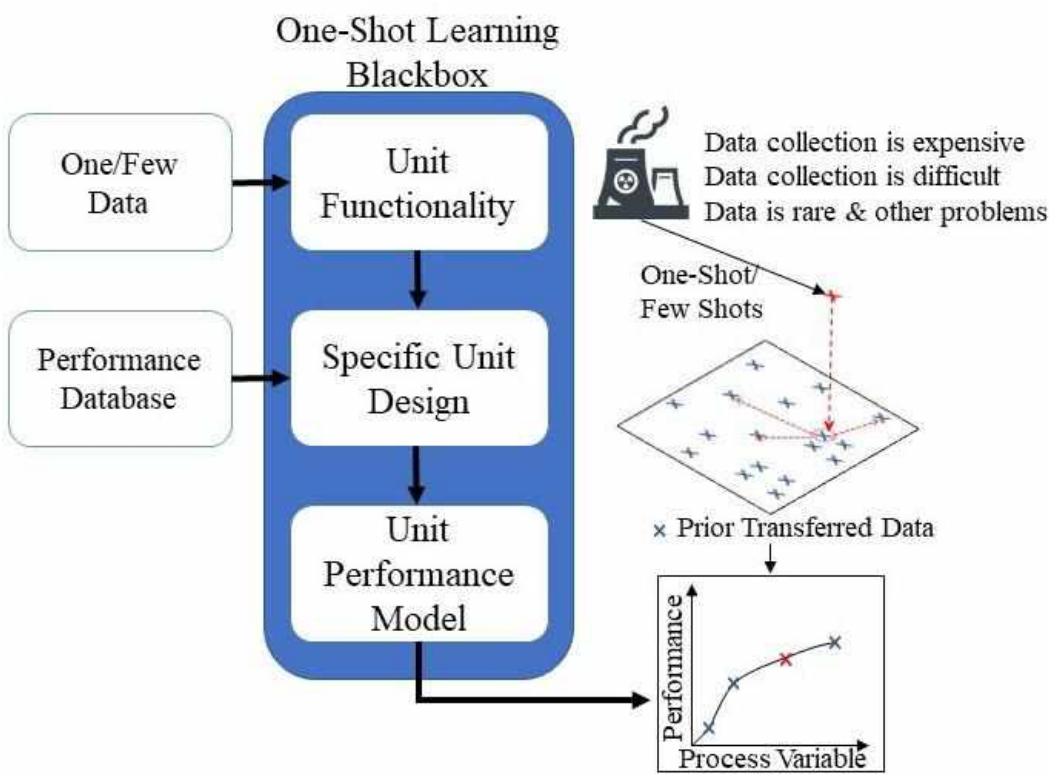


Figure 4.1: Conceptual diagram of one-shot learning for process modelling

Referring to Figure 4.2., the first step to start the one-shot learning workflow is to confirm that a data-driven modelling approach is required. Next, it is required to verify if the use of one-shot learning is suitable by evaluating data availability. If there is flexibility for data sampling or data quantity, the use of an alternative data-driven method such as principal component analysis statistical process optimization (Teng et al., 2019a), Monte-Carlo simulation (Ngan et al., 2020), or adaptive analytical approaches (Leong et al., 2019b) can be deployed. For the next step, it is required to identify the process unit functionality and obtain one or a few samples (which includes the performance variable of interest). Based on this unit functionality, prior data from other process units with similar functionality should be obtained from sources such as multiple manufacturer's databases, other facilities, commercial software, etc. As an example regarding "similar functionality", during the modelling of a 1,2-pass heat exchanger, one could use the data from a 2,2,-pass heat exchanger in similar conditions as prior data due to its similar functionality within the process. In short, this means that the data obtained from units of a different design but fulfilling the same processing purpose can be used as a knowledge basis for this one-shot learning framework. Nevertheless, the knowledge basis should have a statistically significant amount of data to give a good representation of the unit functionality.

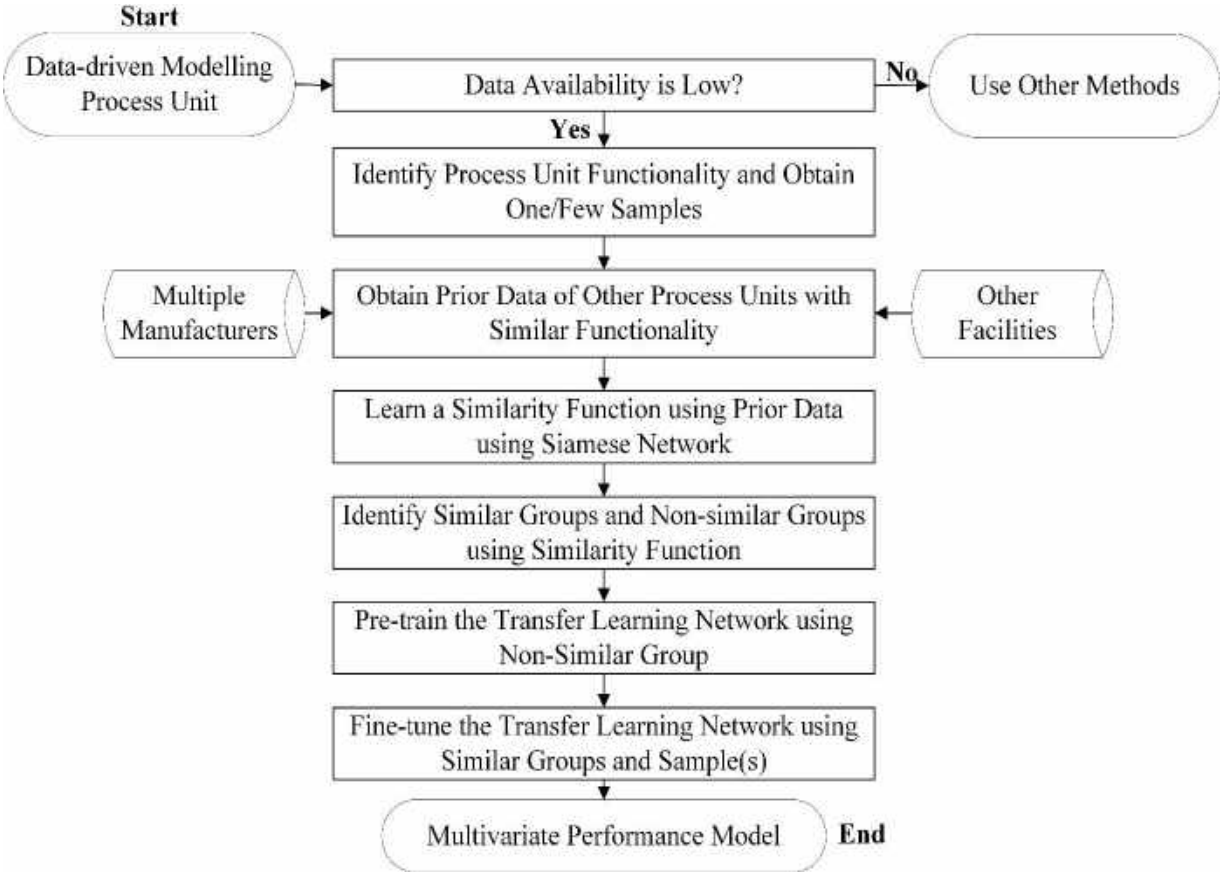


Figure 4.2: Workflow of the novel one-shot learning framework for process system modelling

Furthermore, from this prior data, a Siamese network is used to learn the similarity function between the prior grouped data. In this paper, the contrastive loss (Hadsell et al., 2006) is used as the similarity function (see Figure 4.3). The similar groups with regards to the one/few samples (group encoding is represented as *C*) are then selected using the equation:

$$C = y\left(argmin\ f_c(x, x_s)\right) \qquad \forall x \in X \tag{4.1}$$

Where x is the single data from the prior, X is the full prior dataset, x_s is the one-shot sample, and y is the group classification of the data. Other data with a different group than the similar group are all separated into a non-similar group. Next, this work proposes the use of a multi-layer perceptron (MLP) with distinct losses for transfer learning. During the pre-train, the non-similar group are split at an 80:10:10 training, validation, and testing ratio with a modified Pearson's correlation coefficient as the loss to learn the general shape of the performance space. Next, the similar group which contains the one/few samples are used to finetune the features of the pre-trained neural network by using a few extension layers. The loss function for the

fine-tuning step is set to be the mean squared error function to give an aggressive fine-tuning result, giving high accuracy. Thus, the performance characteristics of the processing unit can be predicted by the transfer learned network with the knowledge represented from the space of prior data of process unit models of similar functionality.

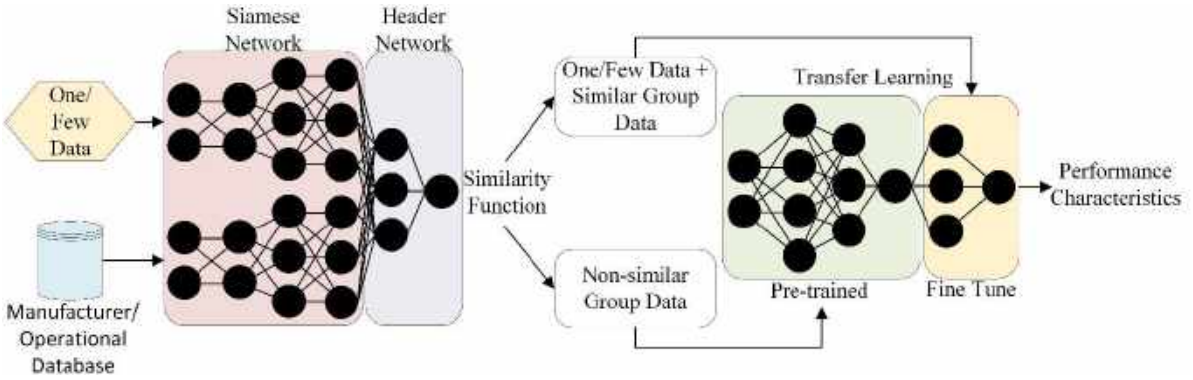


Figure 4.3: Siamese network and transfer learned MLP for one-shot learning in process system modelling

4.3 Case study problem

The modelling of a combined heat and power (CHP) unit (ECOMAX 44 NGS 1.1 HW model) is studied. CHP units commonly operate at a fix steady-state point, and changing the operational point is generally costly in terms of process economics. One sample data was collected from this CHP model. Prior data were prepared using an in-house collection of CHP performance database consisting of 613 datasets from 64 units. The sampled unit is also cross-checked with the database, and the database does not contain data from the specific CHP unit model. For learning the similarity function, all possible 15 variables were input, such as total efficiency, carbon emission, power generation etc. However, due to modelling requirement, the performance characteristics only require 3 inputs to predict the CHP's thermal efficiency and overall efficiency. These 3 variables are power utilization percentage, temperature, and fuel consumption. Finally, for validation purposes, an extra 7 data samples were obtained from the studied CHP from its operational history.

4.4 Results

Using the single data from the studied CHP unit, the Siamese network was able to learn the similarity function and allowed for a transfer learning MLP network learn the representation of unit functionality and fine-tuned on similar data groups. By comparing the predicted output using the one-shot learning framework and the 7 ground truth data (as well as the 1 data sample),

it is possible to achieve a very promising mean absolute error (MAE) of 0.02, mean squared error (MSE) of 0.000616 and R^2 of 0.9992. These results demonstrate that the framework was able to model the CHP unit only with 1 sample with acceptable error (See Figure 4.4).

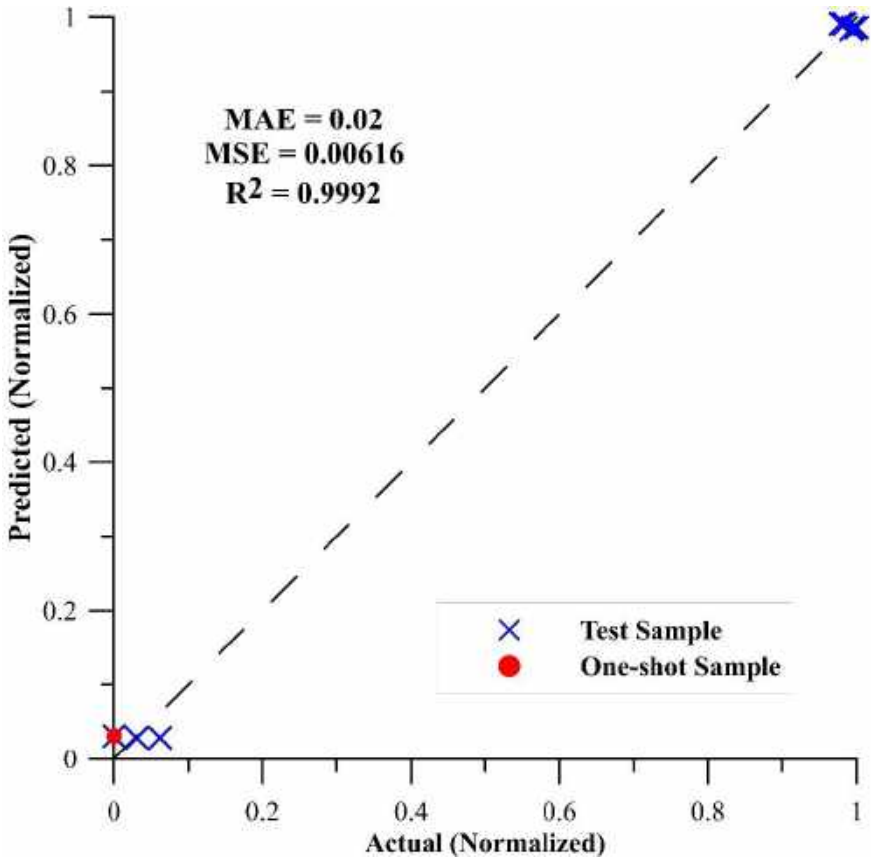


Figure 4.4: Predicted against the actual plot of test samples and one-shot sample with their overall error.

For further analysis, the surface plot of the generated performance model was plotted in Figure 4.5. It is identified that power utilization in percentage was the main factor of the efficiencies in the CHP model. Whereas fuel consumption and temperature also slightly affect CHP efficiencies. It is also observed that increase in power utilization gives a steep increase in thermal efficiency, however, creates a plateau for the overall efficiency. This implies that at over 80% power utilization, the total energy output is approximately the same, however the ratio of power and heat energy gradually increases when further increasing the power utilization. Nevertheless, temperature and fuel flowrate are providing small effects on the thermal and overall efficiencies.

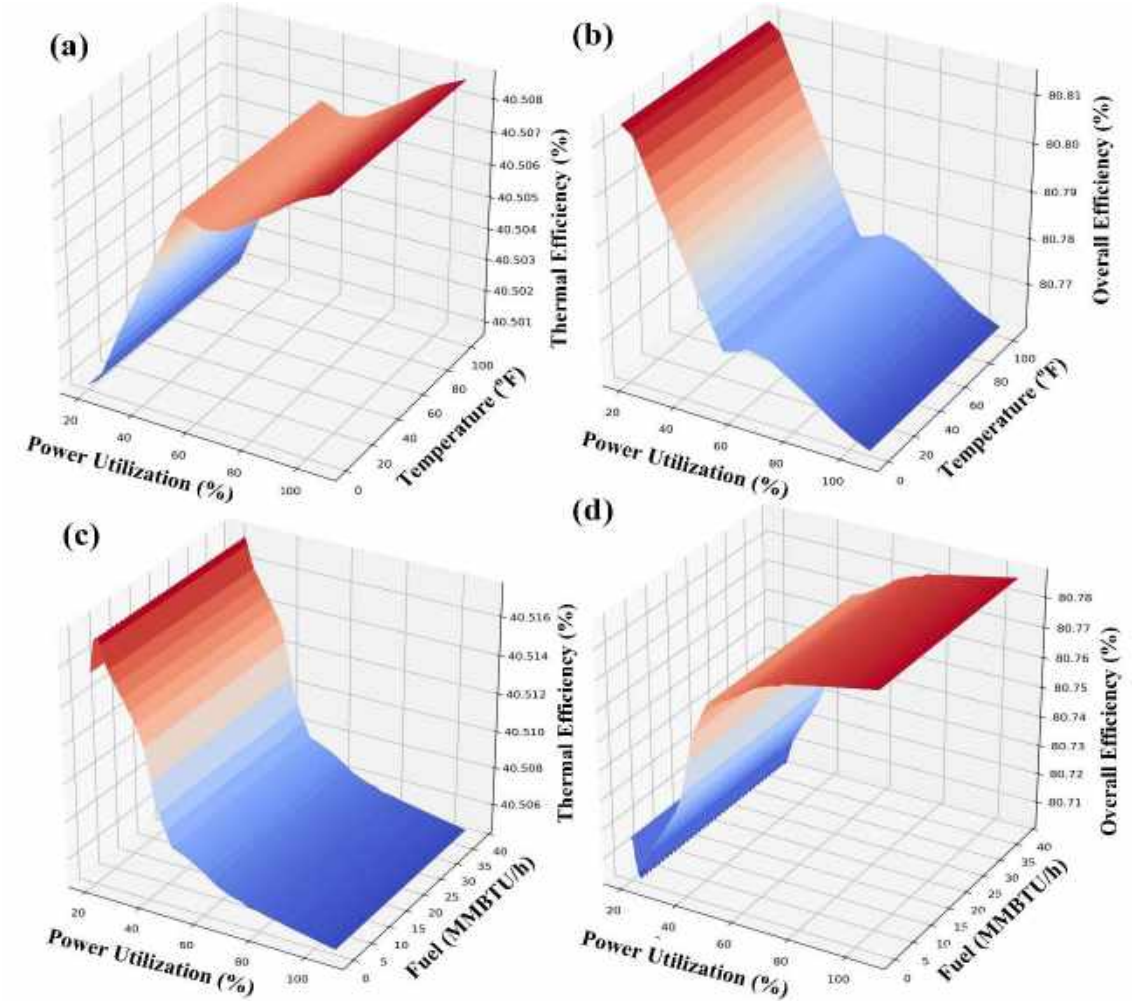


Figure 4.5: Surface plot of (a) Power utilization (PU), temperature (T) and thermal efficiency (TE) (b) PU,T, Overall Efficiency (OE) (c) PU, Fuel, TE, (d) PU, Fuel, OE

4.5 Conclusion

This chapter proposes the use of novel one-shot learning for the application of modelling process systems under low data availability problems. A novel two-step approach is proposed. First, a Siamese network is proposed to learn the similarity function within prior datasets that are from different units but share the same functionality with the studied unit. Next, these datasets are categorized as a non-similar group and similar groups to be sequentially transferred learn by a multi-layer perceptron (MLP) neural network. Using a few extra samples of the studied system that were never shown to the networks, the model was tested. A convincing mean squared error of 0.00616 was achieved only from the original one sample from the studied combined heat and power (CHP) unit. A smooth prediction space was observed from the surface plot, showing smooth continuity in the model. This demonstrates that this approach is

stable and able to model process units with high accuracy and only a single sample similarly. Additionally, the one-shot learning framework can be applied on any other units.

References

Baheri, A., Vermillion, C., 2017, Altitude optimization of Airborne Wind Energy systems: A Bayesian Optimization approach, in: 2017 American Control Conference (ACC). IEEE, Seattle, WA, 1365–1370 DOI:10.23919/ACC.2017.7963143

Berger-Tal, O., Nathan, J., Meron, E., Saltz, D., 2014, The Exploration-Exploitation Dilemma: A Multidisciplinary Framework. PLoS One 9, e95693 DOI:10.1371/journal.pone.0095693

Brynjolfsson, E., McElheran, K., 2016, Data in Action: Data-Driven Decision Making in U.S. Manufacturing. SSRN Electron. J. 1–55 DOI:10.2139/ssrn.2722502

Ghosh, J., Dunson, D.B., 2009, Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis. J. Comput. Graph. Stat. 18, 306–320 DOI:10.1198/jcgs.2009.07145

Guo, H., Schwartz, W.R., Davis, L.S., 2011, Face verification using large feature sets and one shot similarity, in: 2011 International Joint Conference on Biometrics, IJCB 2011 DOI:10.1109/IJCB.2011.6117498

Guo, Q., Feng, W., Zhou, C., Huang, R., Wan, L., Wang, S., 2017, Learning Dynamic Siamese Network for Visual Object Tracking, in: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 1781–1789 DOI:10.1109/ICCV.2017.196

Hadsell, R., Chopra, S., LeCun, Y., 2006, Dimensionality Reduction by Learning an Invariant Mapping, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2, 1735–1742 DOI:10.1109/CVPR.2006.100

Hamacher, T., Niess, J., Schulze Lammers, P., Diekmann, B., Boeker, P., 2003, Online measurement of odorous gases close to the odour threshold with a QMB sensor system with an integrated preconcentration unit. Sensors Actuators B Chem. 95, 39–45 DOI10.1016/S0925-4005(03)00400-3

Koch, G., Zemel, R., Salakhutdinov, R., 2015, Siamese Neural Networks for One Shot Image Learning, in: ICML Deep Learning Workshop. JMLR: W&CP, Lilee , France, 1–8.

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P., Ponnambalam, S.G., 2019, Adaptive analytical approach to lean and green operations. J. Clean. Prod. 235, 190–209 DOI:10.1016/j.jclepro.2019.06.143

Li Fei-Fei, Fergus, R., Perona, P., 2006, One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. 28, 594–611 DOI:10.1109/TPAMI.2006.79

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018, Key pillars of successful energy saving projects in small and medium industrial enterprises. Energy 158, 293–304 DOI:10.1016/j.energy.2018.06.018

Máša, V., Touš, M., Pavlas, M., 2016, Using a utility system grey-box model as a support tool for progressive energy management and automation of buildings. Clean Technol. Environ. Policy 18, 195–208 DOI:10.1007/s10098-015-1006-x

Miller, M.G., Matsakis, N.E., Viola, P.A., 2000, Learning from one example through shared densities on transforms, in: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662). IEEE Comput. Soc, 464–471 DOI:10.1109/CVPR.2000.855856

Mueller, J., Thyagarajan, A., 2016, Siamese recurrent architectures for learning sentence similarity, in: 30th AAAI Conference on Artificial Intelligence, AAAI 2016. AAAI Press, Phoenix, Arizona USA, 2786–2792.

Ngan, S.L., How, B.S., Teng, S.Y., Leong, W.D., Loy, A.C.M., Yatim, P., Promentilla, M.A.B., Lam, H.L., 2020, A hybrid approach to prioritize risk mitigation strategies for biomass polygeneration systems. Renew. Sustain. Energy Rev. 121, 109679 DOI:10.1016/j.rser.2019.109679

Ritter, T., Pedersen, C.L., 2020, Digitization capability and the digitalization of business models in business-to-business firms: Past, present, and future. Ind. Mark. Manag. 86, 180–190 DOI:10.1016/j.indmarman.2019.11.019

Tao, F., Qi, Q., Liu, A., Kusiak, A., 2018, Data-driven smart manufacturing. J. Manuf. Syst. 48, 157–169 DOI:10.1016/j.jmsy.2018.01.006

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019a, Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. 225, 359–375 DOI:10.1016/j.jclepro.2019.03.272

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F., Máša, V., 2019b, Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresour. Technol. 292, 121971 DOI:10.1016/j.biortech.2019.121971

Teng, S.Y., Máša, V., Stehlík, P., Lam, H.L., 2019c, Deep learning approach for industrial process improvement. Chem. Eng. Trans. 76, 487–492 DOI:10.3303/CET1976082

Yang Yang, Saleemi, I., Shah, M., 2013, Discovering Motion Primitives for Unsupervised Grouping and One-Shot Learning of Human Actions, Gestures, and Expressions. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1635–1648 DOI:10.1109/TPAMI.2012.253

Yin, X., Niu, Z., He, Z., Li, Z., Lee, D., 2020, An integrated computational intelligence technique based operating parameters optimization scheme for quality improvement oriented process-manufacturing system. Comput. Ind. Eng. 140, 106284 DOI:10.1016/j.cie.2020.106284

Yip, K., Sussman, G.J., 1997, Sparse representations for fast, one-shot learning, in: Proceedings of the National Conference on Artificial Intelligence. Providence, Rhode Island, 521–527.

Yong, J.Y., Nemet, A., Varbanov, P.S., Kravanja, Z., Klemeš, J.J., 2016, Data reconciliation for total site integration. Chem. Eng. Trans. 52, 1045–1050 DOI:10.3303/CET1652175

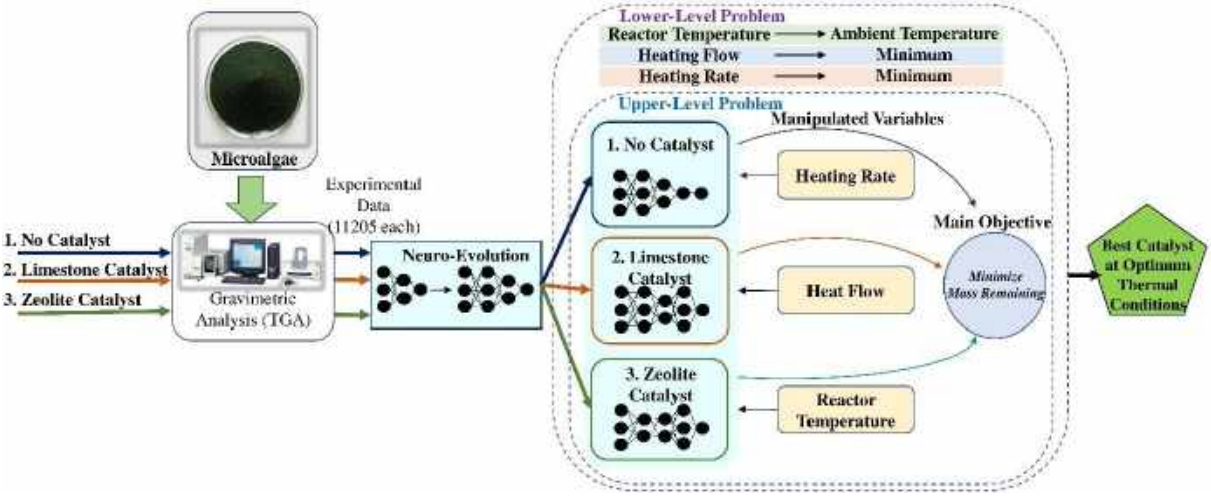
Zhang, J.Z., Chen, J.C., 2008, Tool condition monitoring in an end-milling operation based on the vibration signal collected through a microcontroller-based data acquisition system. Int. J. Adv. Manuf. Technol. 39, 118–128 DOI:10.1007/s00170-007-1186-6

CHAPTER 5 NEURO-EVOLUTION FOR PROCESS OPTIMIZATION OF A SINGLE UNIT

This chapter has been peer-reviewed and published in Bioresource Technology.

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F. and Máša, V., 2019. Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresource technology, 292, p.121971.

Graphical Abstract



Abstract

The aim of this chapter is to demonstrate that neuro-evolutionary approach can be effectively used for the determination of the optimum thermal conversion in a single unit microalgae thermal reactor. The considered temperature range is for both pyrolysis and gasification of the microalgae (*Chlorella vulgaris*). A Progressive Depth Swarm-Evolution (PDSE) neuro-evolutionary approach is proposed to model the Thermogravimetric analysis (TGA) data of Catalytic Thermal Degradation of *Chlorella Vulgaris*. Results showed that the proposed method can generate predictions which are more accurate compared to other conventional approaches (> 90 % lower in Root Mean Square Error (RMSE) and Mean Bias Error (MBE)). In addition, Simulated Annealing is proposed to determine the optimal operating conditions for microalgae conversion from multiple trained ANN. The predicted optimum conditions were reaction temperature of 900.0 ^{o}C, heating rate of 5.0 ^{o}C/min with the presence of HZSM-5 zeolite catalyst to obtain 88.3 % of Chlorella Vulgaris conversion.

Keywords
Microalgae, Thermogravimetric Analysis, Artificial Neural Network, Particle Swarm Optimization, Simulated Annealing

5.1 Introduction

Biofuel is a potential energy source that can be used as an alternative to fossil fuel (Milano et al., 2016). In the last decades, the third generation (3G) biofuel which is derived from microalgae have received great attention since first generation (1G) and second generation (2G) biofuels contain food security issue (Mohr and Raman, 2013) and sustainability concern (Sun et al., 2019) respectively. In general, the main advantages of using microalgae as biofuel production feedstock encompasses of easy cultivation; higher growth rates and productivity; attractive oil yield; and lower carbon emissions (Mata et al., 2010; Costa and Morais, 2011).
Among the various technologies established for converting microalgae to biofuels (e.g., pyrolysis, hydrothermal carbonization, and gasification (Chan et al., 2019)), catalytic pyrolysis is one of the most preferable technologies (Zainan et al., 2018). A series of catalysts have been studied thoroughly on their ability to break large hydrocarbon aromatic compounds of biomass into smaller hydrocarbon compound via decarboxylation (Mettler et al., 2014), dehydration (Barnard and Hughes, 1960) and deoxygenation (Raymundo et al., 2019) reactions. Among all the catalysts, HZSM-5 zeolite and CaO catalysts have driven more attention due to their performance in upgrading the bio-oil quality and enhance the yield of bio-oil (Zhang et al., 2019). HZSM-5 zeolite is highly selective for aromatics and effective in deoxygenation of oxygenated compounds to form olefins and phenolics (Yang et al., 2018). Meanwhile, CaO has been introduced into pyrolysis process as an alkali metal oxide catalyst. It is normally extracted from limestone and eggshell waste due to highly abundant, high CaO composition, and relatively low cost (Gan et al., 2018).

Thermogravimetric analysis (TGA) is used to investigate the thermal degradation of biomass by measuring the rate of weight loss as a function of temperature and time. Gai et al. (2013) had investigated the kinetic parameters of *Chlorella pyrenoidosa* (CP) and *Spirulina plantensis* (SP) microalgae. They reported the average activation energy of CP and SP were 77.02 and 91.56 kJ/mol. Besides, Kim et al. (2012) reported on the lumped kinetic model using *Saccharina Japonica* as feedstock. The average activation energy obtained was in the range of 102.5-269.7 kJ/mol. Most works are focusing on the kinetic parameters of the chemical

mechanism of the thermal decomposition of microalgae. There is still limited literature on the optimization study on the mass loss percentage (MLP) of thermal degradation of biomass.
Artificial neural network (ANN) have been recently applied to model the thermochemical performances of biomass pyrolysis due to its credibility in addressing complex nonlinear problems. In the early of 21st century, Abbas et al. (2003) had proposed the use of neural network model in predicting the devolatilization performances of coal and biomass. The computing efficiency of the model was proven to be more attractive compared to other existing tools Conesa et al. (2004) then developed a predictive model using a multilayer ANN, to estimate the thermal decomposition of cellulose, lignin and polyethene. This research has been extended to study the thermochemical performance of various feedstocks via pyrolysis (e.g., keratin biopolymer (Fazilat et al., 2012); sewage sludge (Naqvi et al., 2018), etc.). The popularization of ANN being used to analyse experiments datasets from TGA systems is mainly due to (i) the large size of the datasets from TGA systems (> 10,000 data per run); (ii) the complex non-linear nature of thermal conversion that requires advanced analytics for prediction and modelling; (iii) fully automatic modelling procedure that can be achieved by ANN.

Conventional practices in training neural network for the application of TGA analysis considers single-hidden layer neural networks (Abbas et al., 2003). Although researchers acknowledge that the consideration of better hyperparameters for ANN is important in TGA analysis (Mayol et al., 2018), hyperparameter optimization is rarely carried out for the applications of thermal analysis. Besides, the current norm for researchers is to use trial and error methods to find an acceptable ANN topology. This is demonstrated in a recent work of Naqvi et al. (2018), which tested neural network with lesser than 3 hidden layers. Nevertheless, Mhaskar and Poggio (2016) highlighted the importance of deeper networks from an approximation theory perspective. Another issue is on the activation functions of ANN, commonly one or only a few types (Mayol et al., 2018) of activation function are considered by trial and error for the application of TGA. From a computational intelligence perspective, a good methodology to achieve optimality instead of trial-and-error is by using Neuro-evolution (Stanley and Miikkulainen, 2002). This method is very popular in the fields of Deep Learning because natural brains themselves are the products of bio-evolutionary processes, and its implementation enables large-scale computing (Stanley et al., 2019). Based on the results obtained by the researchers from Deep Mind (Jaderberg et al., 2017) and OpenAI (Salimans et

al., 2017), it can be confirmed that Neuro-evolution is highly effective and implementable even when considering computational scalability.

To-date, none of the work has utilized this approach for the purpose thermal conversion optimization via TGA data. Thus, the objective of this work is to incorporate the use of Neuro-evolutionary approach in the applications of analysing TGA experimentation data. The increase in accuracy of the neural network prediction has allowed for a more indicative optimization study on the ANN model. However, the optimization study was not performed in the previous works (e.g., Naqvi et al. (2018) and Conesa et al. (2004)) mainly due to the complex structure of ANN, which makes conventional optimization strategies non-straightforward. To address this issue, this work proposes the use of Simulated Annealing (Metropolis et al., 1953), a metaheuristic algorithm, to search for the optimal thermal conditions for microalgae conversion from multiple trained ANN.

5.2 Materials and methods

This chapter will describe the full experimental methodology and the modelling and optimization of the pyrolysis/gasification of *C. vulgaris* microalgae.

5.2.1 Experimental description

The microalgae biomass *Chlorella vulgaris* (*C. vulgaris*) was obtained from Centre for Biofuel and Biochemical, Universiti Teknologi PETRONAS (UTP), Malaysia. The biomass was dried and sieved to a particle size less than 200 μm. Moreover, limestone and HZSM-5 zeolite catalysts were obtained from Calrock Sdn. Bhd. and Sigma-Aldrich, Malaysia, respectively. Both catalysts undergo pre-treatment such as the limestone was heated at temperature 900.0 °C for 4 hours to ensure all the $CaCO_3$ in limestone was fully converted to CaO whereas HZSM-5 zeolite catalyst was calcined at 550.0 °C for 3 hours to activate the Bronsted acid sites. After that, the *C. vulgaris* samples were analysed using thermogravimetric analyser (TGA-DSC 1, Mettler Toledo) and LECO CHNS-932 elemental analyser, respectively to determine the physical and chemical properties of the samples. From the ultimate analysis of *C. vulgaris*, the carbon, hydrogen, nitrogen, sulphur, and oxygen contents were 42.6 wt%, 8.2 wt%, 1.3 wt%, 0.8 wt% and 47.1 wt%, respectively. Whereas the moisture, volatile matter, fixed carbon, and ash content of *C. vulgaris* showed in the proximate analysis were 8.3 wt%, 59.2 wt%, 15.8 wt% and 16.7 wt%, respectively.

The pyrolysis studies were performed using five different heating rates of 10, 20, 30, 50 and 100 °C/min, respectively using the TGA equipment. Firstly, 100 ml/min of nitrogen gas (N_2) supply was introduced into the TGA for 10 min to avoid unwanted oxidation reaction of the biomass sample. After that, 10 mg of *C. vulgaris* biomass was introduced into the TGA and heated from 50.0 °C to 900.0 °C under non-isothermal conditions. Meanwhile, for catalytic pyrolysis process, the catalyst (e.g. limestone and HZSM-5 zeolite) with a ratio of 1:10 to the biomass was mixed homogenous with the *C. vulgaris* and loaded into the system. All experiments were carried out three times to ensure the reliability of the results.

5.2.2 Neural network generation

The experimentation data consists of heating rate (°C/min), heat flow (mW) and reactor temperature (°C) while providing results on the mass of microalgae remaining after the reaction. Three cases of experimentation were carried out which is the case with no catalyst, with limestone catalyst and HZSM-5 zeolite catalyst. The three separate datasets each containing 11,205 data points, giving a total of 33,615 data points. Each dataset was split with a ratio of 8:2 for training and validation. The ANN that was considered in this work is a fully connected neural network with a variable depth as shown in Fig. 5.1.

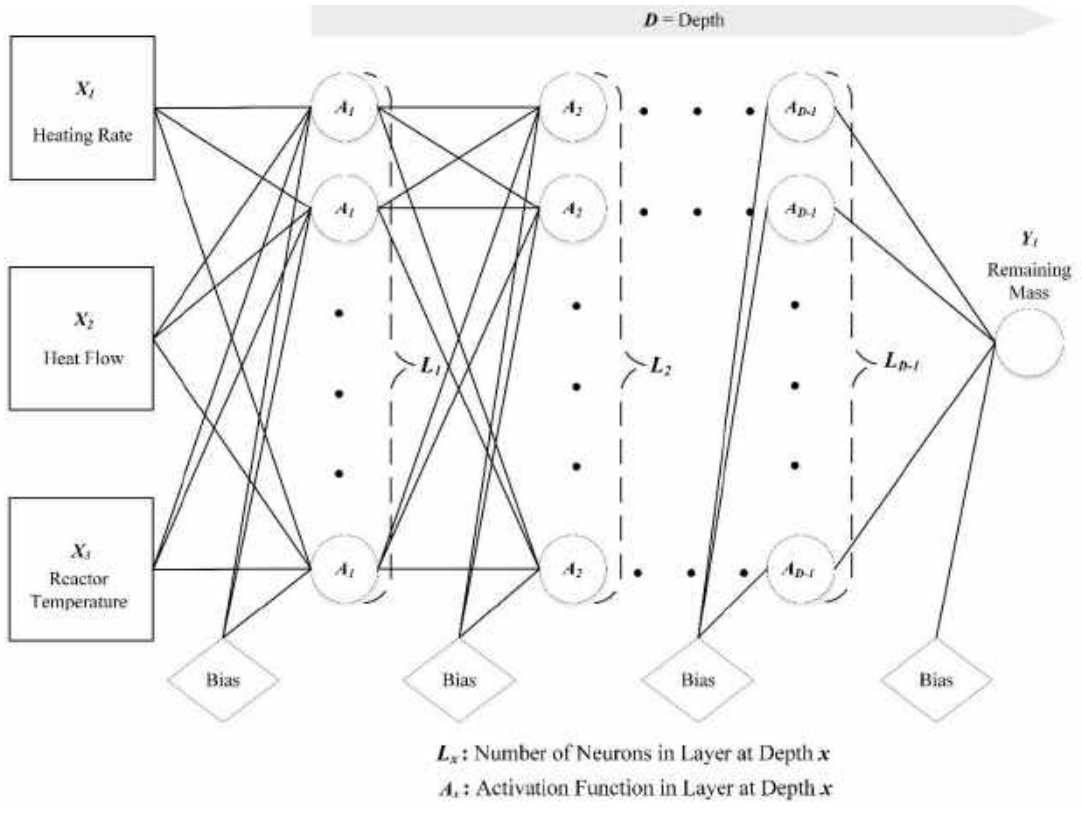


Figure 5.1: Structure of fully connected neural network with variable depth considered.

From Ding et al. (2011), neuro-evolutionary strategies can be applied on the weights of ANN, architecture of ANN, and the activation function of ANN. However, due to the formulation of ANN being mathematically differentiable, the use of stochastic gradient descent (SGD) algorithms such as Adam (Kingma and Ba, 2014) is more suitable to be used on directly training the weights of ANN (Stanley et al., 2019) as it can utilize the gradient of the ANN for more efficient weight updates. Contrarily, the search for neural architecture and activation functions can be learned using Neuro-evolution approaches on top of SGD-based weight training (Floreano et al., 2008).

Considering the computational costs of Neuro-evolution algorithms, this work proposes an algorithm called Progressive Depth Swarm-Evolution (PDSE) that is based on the efficient and robust modified Particle Swarm Optimization (PSO) algorithm (Shi and Eberhart, 1998). PDSE uses a variable length embedding for the population by progressively searching from shallow neural networks to deeper neural networks. Furthermore, the computational costs of this neuro-evolution algorithm can be effectively bounded by allocating a maximum search cost for each number of depths of ANN. The details of the algorithm are presented as pseudocode in Fig. 5.2

Progressive Depth Swarm-Evolution (PDSE) for Evolving Neural Networks

```
For i in C                          • C is datasets, which is type of catalyst in this case
  For j in D,   D = [D_L, D_h]      • D is depth of neural network, which is bounded by a range
    M ← Maximum Search Cost;  DepthBest ← 1
    Architecture ← List of size D;   Activation ← List of size D
    P ← Concatenate [D, Architecture, Activation]      • Create encoded population, P
    v ← Randomize with size P          • Initialize velocity, v which corresponds to P
    While Networks Evaluated < M
      F ← Evaluate Network(P)    • Train population of ANN according to population and get fitness, F
      v ← w v + φ_p rand_p (LocalBest-P) + φ_g rand_g (GlobalBest-P)   • Update velocity by PSO
      P ← P + v                  • Update position by PSO
      If GlobalBest < DepthBest:
        DepthBest ← GlobalBest           • Update best neural topology for the depth
```

Figure 5.2: Pseudo-code for Novel Progressive Depth Swarm-Evolution (PDSE)

In this work, a particle-to-generation ratio of 2:5 and maximum search cost of 1000 were used. The PSO algorithm is extended from the EvoOpt library (Teng, 2019). For this application, the PDSE algorithm was used to search for ANN of depth 2 to 10 with fitness set to its validation loss. For the training of the weights in ANN, the popular stochastic gradient descent-based algorithm, Adam was used with batch size of 128 (Kingma and Ba, 2014). To prevent

overfitting of the ANN, early stopping technique was implemented with a setting of minimum delta of 0.0001 and patience of 3.

The metric that is being used for training (loss) is the Mean Squared Error (MAE) as it provides a globally differentiable loss function for smooth training. For the purpose of benchmarking, 3 additional metrics were used, which are Root Mean Squared Error (RMSE), Mean Bias Error (MBE) and the coefficient of determination (R^2):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\widehat{y_t}-y_t)^2}{N}} \tag{5.1}$$

$$MSE = \frac{\sum_{t=1}^{N}(\widehat{y_t}-y_t)^2}{N} \tag{5.2}$$

$$MBE = \frac{\sum_{t=1}^{N}(\widehat{y_t}-y_t)}{N} \tag{5.3}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(\widehat{y_t}-y_t)^2}{\sum_{t=1}^{N}(\overline{y_t}-y_t)^2} \tag{5.4}$$

where y_t is actual data, $\widehat{y_t}$ is predicted data, $\overline{y_t}$ is the mean of actual data, while N is the total number of actual data.

5.2.3 Bi-layer optimization

ANN serves as a good predictive regression tool. However, an optimization tool has to be deployed on the black-box ANN in order to find the predicted optimal thermal conditions and the most suitable catalyst of thermal degradation of microalgae. This work formulates this optimization problem into a bi-level optimization problem, where the main problem is to minimize the mass of the remaining microalgae after the reaction, Y_1 enabling efficient conversion. The outer problem is to minimize the conditions of the reaction to standard conditions and achieve better energy efficiency. The formulation of this bi-level optimization problem is shown in Eqs. (5.5) and (5.6):

$$\min Y_1$$
$$\text{Subject to } X_{i,L} \leq X_i \leq X_{i,H} \quad \forall i \in 1,2,3 \tag{5.5}$$
$$\min \ |X_1| + |X_2| + |X_3 - T_{ambient}|$$
$$\text{Subject to } X_{i,L} \leq X_i \leq X_{i,H} \cap Y_1 = Y_1^* \ \forall i \in 1,2,3 \tag{5.6}$$

where X_1, X_2, X_3 are the rate of change in temperature, heat flow and the reactor temperature, respectively. Y_1* is the optimal mass of remaining microalgae from the main problem. For this

work, the constraints used are 5.0 to 100.0 °C/min for X_1, 0.0 mW to 300.0 mW for X_2, and a varying range of 50.0 °C interval for X_3. The ambient temperature is taken as 25.0 °C.

In order to solve the optimization problem, this work uses Simulated Annealing, which is a stable and well-known metaheuristics algorithm. The choice of this metaheuristic algorithm is due to Simulated Annealing being able to find global optimum points by escaping local minima and guarantees statistically robust results in short computation time (Goffe et al., 1994). The advantages of Simulated Annealing are particularly useful for this application, where the trained ANN is found to have multiple minimum points and a complex surface. The Simulated Annealing algorithm has its inspiration from metallurgical annealing, which involves heating and cooling of material to reduce their defects. This algorithm can be presented in five simple steps:

Step 1: Generate initial solution randomly.

Step 2: Find a neighbor solution for each of the initial solutions.

Step 3: Stochastically choose a solution between the initial and neighbour solution based on the probability, $p = \exp(-(Fitness_{neighbour}-Fitness_{initial})/Temperature)$.

Step 4: Decrease the temperature.

Step 5: Repeat from step 2 if stop condition is not met.

For the application of optimizing ANN trained on TGA experimentation data, both the inner and outer problem uses a population size of 30 and a maximum generation of 50. This optimization procedure ensures that the optimal thermal conditions of the microalgae can be found together with the recommended catalyst.

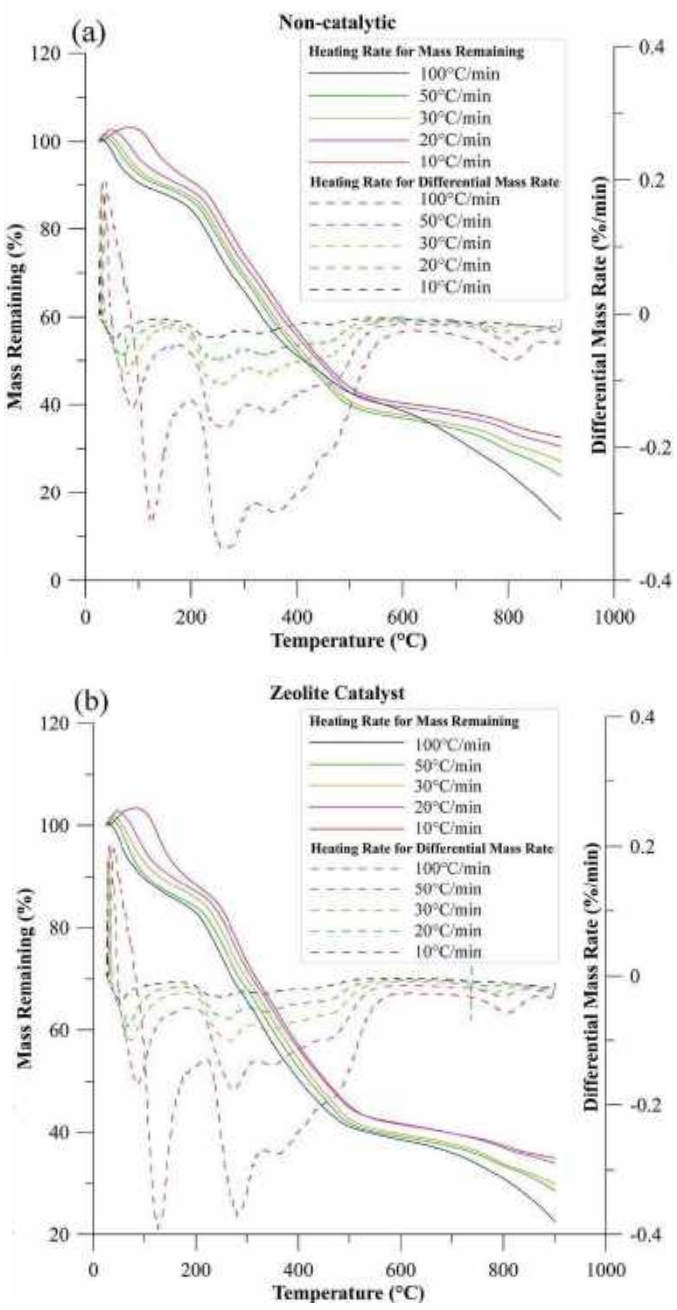
5.3 Results and discussion

This chapter presents the thermogravimetry results, results of ANN, and optimization result of the microalgae thermal conversion.

5.3.1 TG-DTG analysis of *C. vulgaris*

The thermal degradation profiles of thermogravimetry (TG) and derivative thermogravimetry (DTG) curves of *C. vulgaris* with and without the presence of HZSM-5 zeolite and limestone catalyst are shown in Fig. 5.3. The TGA curves were used to analyse the degradation profile of the samples at different temperature stages. Meanwhile, the DTG curves were used to determine the temperature in which the samples undergo the maximum degradation and it was divided into three main volatilization stages namely stage I, stage II, and stage III. The first degradation stage (Stage I) of *C. vulgaris* occurred from 50.0 ºC to 190.0 ºC was due to the

intrinsic breakdown of lipids and proteins as well as the removal of moisture in the microalgae cells. Then, the second degradation stage (Stage II) started at 190.0 ºC to 600.0 ºC (Bach et al., 2017). A significant mass loss of *C. vulgaris* was observed in this stage as it has the highest maximum degradation as shown in the DTG curve. This phenomenon was due to the degradation of the organic compounds in the cell such as protein, carbohydrate, and lipids. Lastly, the third degradation stage (Stage III) observed at higher temperature than 600.0 ºC. This stage is the final breakdown of lipids which associated with the breakdown of the fatty acid chains (FFA) in the cell (Figueira et al., 2015). The leftover solid residues are the undecomposed ash.



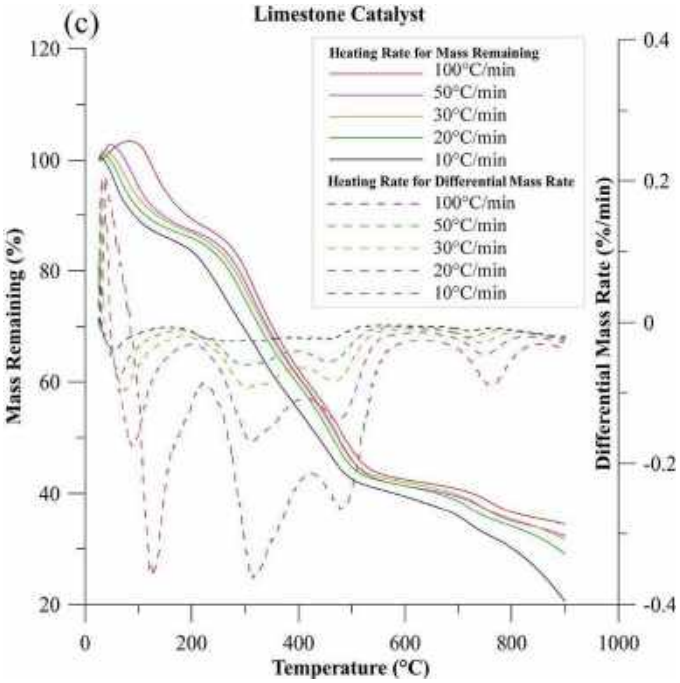


Figure 5.3: TGA and DTG curves of (a) non-catalytic degradation of *C. vulgaris*, (b) catalytic degradation of *C. vulgaris* using limestone catalyst, (c) catalytic degradation of *C. vulgaris* using HZSM-5 zeolite.

Table 5.1 shows the mass loss percentage for both catalytic and non-catalytic thermal degradation at different heating rate ranging from 10.0 ºC/min to 100.0 ºC/min. Highest average mass loss percentage was obtained in catalytic thermal degradation of *C. vulgaris* using limestone, followed by catalytic thermal degradation of *C. vulgaris* using HZSM-5 zeolite and non-catalytic thermal degradation of *C. vulgaris*. This phenomenon can be explained through the base catalytic effect of CaO metal oxide element in limestone. Previous study has reported that the addition of base catalyst can improve the quality and yield of the biomass pyrolysis oil. CaO can be used as an absorber or catalyst to enhance the efficiency of the process as well as capture the CO_2 produced and convert it to $CaCO_3$. This will further reduce the tar formation and enhance the quality of biofuel production by converting long hydrocarbon chains to short hydrocarbon chains fuel (Chen et al., 2019).

Table 5.1: Mass loss during non-catalytic and catalytic thermal decomposition of *C. vulgaris* using HZSM-5 zeolite and limestone catalyst.

	Heating Rate (β, °C/min)	$T_{initial}$ (°C)	T_{final} (°C)	T_{max} (°C)	Mass loss (%)
C. vulgaris (without catalyst)	10	140.34	553.97	247.80	86.3
	20	154.59	565.64	254.70	76.11
	30	160.42	578.58	255.70	72.87
	50	173.94	596.09	268.77	69.41
	100	189.60	619.41	272.54	67.42
Average	-	163.78	582.74	259.90	74.42
C. vulgaris – HZSM-5 zeolite	10	157.94	533.10	251.93	86.68
	20	173.15	562.11	264.55	80.47
	30	183.07	596.11	266.39	79.17
	50	191.96	605.09	274.52	75.17
	100	210.59	620.55	281.02	74.17
Average	-	183.34	583.39	267.68	79.13
C. vulgaris – Limestone	10	154.05	554.05	264.49	88.58
	20	173.94	562.16	300.66	79.97
	30	178.53	576.21	307.24	77.47
	50	199.23	593.39	322.59	76.87
	100	215.94	609.33	324.10	74.67
Average	-	184.34	579.03	303.82	79.51

5.3.2 Training and Validation of ANN

The actual optimum conditions of the thermal decomposition process cannot be determined simply based on Fig. 5.3. Therefore, this work employs the TG data to a neural network and determine the optimal condition by optimizes the outputs of the neural network instead. Using the PDSE algorithm, a total of 30,000 ANN with different activation function and topologies were evaluated. A total of 11 activation functions are considered, which includes softmax, exponential linear unit (elu), scaled exponential linear unit (selu), softplus, softsign, rectified linear unit (ReLu), hyperbolic tangent (tanh), sigmoid, hard sigmoid, exponential (Exp) and linear functions. One of the advantages of the PDSE algorithm is that evolution can easily be parallelized or distributed across multiple computation machines with ease. In this work, the training was distributed over 3 machines with Intel(R) Core (TM) i5-4460, i5-8250 and i7-6700HQ with NVIDIA GeForce GTX 950M. The training time took approximately 12 hours in total. Referring to Fig. 5.4(a) to 5.4(c) the PDSE algorithm steadily reduced the loss of the ANN with best neural architecture on extra evaluated ANNs. The algorithm has shown success in achieving optimality within a reasonable context for every single case of catalyst and depth of ANN. This demonstrates the robustness of the algorithm in obtaining an optimal neural architecture.

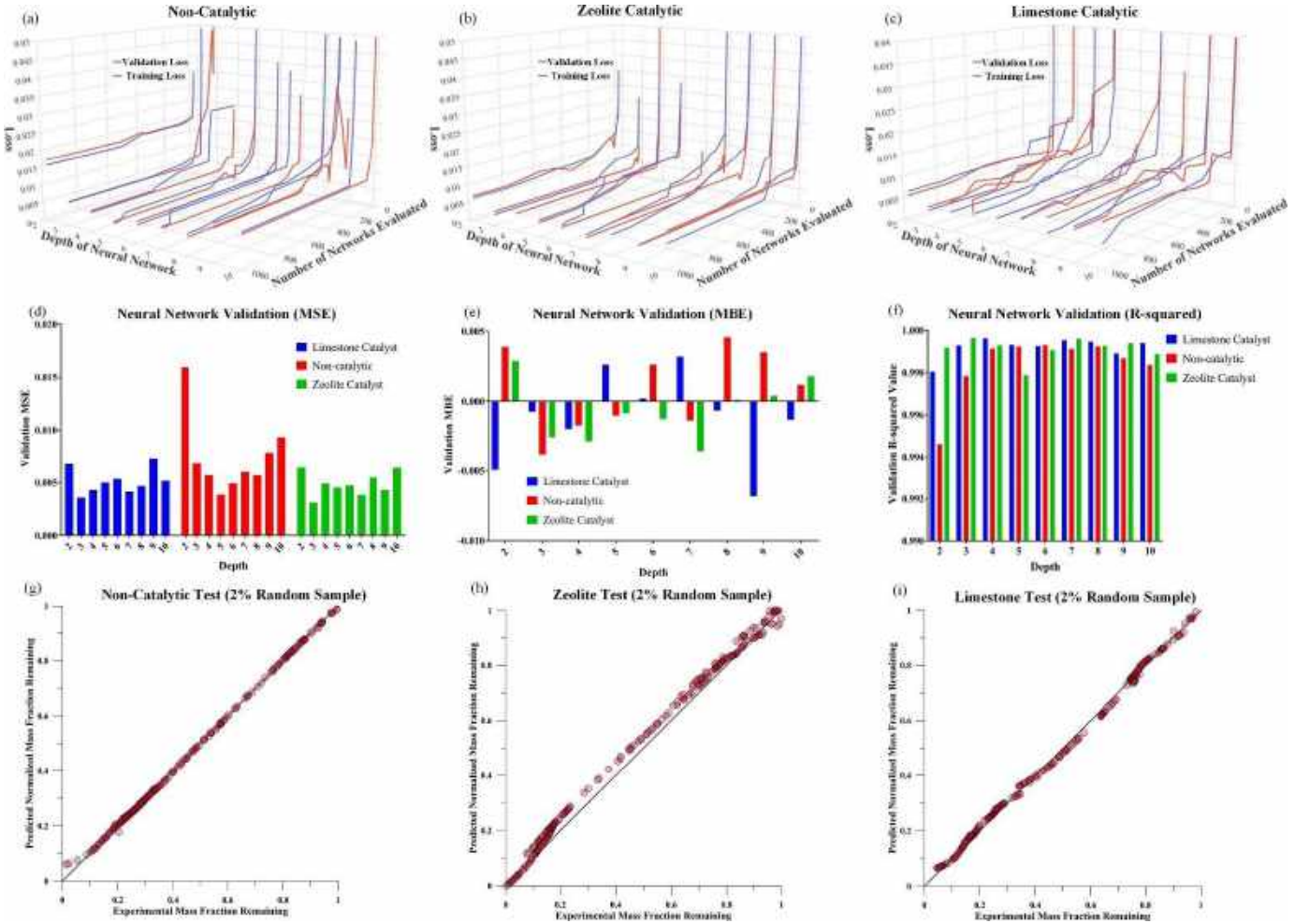


Figure 5.4: Training and Validation of ANN (a) non-catalytic degradation of C. vulgaris, (b) catalytic degradation of C. vulgaris using HZSM-5 zeolite catalyst, (c) catalytic degradation of C. vulgaris using limestone catalyst dataset; Accuracy validation of ANN by (d) Mean Squared Error, (e) Mean Bias Error, (f) R-squared value; Predictions against (g) experimental data for non-catalytic testing, (h) experimental data for zeolite testing, (i) experimental for limestone testing.

From Fig. 5.4(d), it can be observed that there is an overall trend of improvement of loss (MSE) until a specific depth, and then the loss increases giving a "U" shaped trend. This phenomenon is due to the limitations of computational power (costs) in a larger search space as constrained by the maximum search costs in the PDSE algorithm. It is worth mentioning that the maximum search cost is set to reflect an acceptable computation time for the algorithm, and hence it is representative of the machine's computation power. With a more powerful computational machine, "maximum search cost" within the PDSE algorithm can be improved. Hence, the algorithm can more effectively exploit better architectures of neural network with the same depth. The authors also highlight that the problem of evolving neural networks faces the exploitation-exploration dilemma (Tan et al., 2009) where computation power and time is fixed, and search quality (exploitation) is controlled by "maximum search costs" while search area (exploration) is controlled by the range of network's depth searched. Nevertheless, referring to Fig. 5.4(e) and 5.4(f), the PDSE algorithm has found very convincing neural architecture for all cases and depths with MBE overall between -0.0050 to 0.0050 and R-squared overall above 0.9950. Note that the predictions made by the model against the experimental data were shown in Fig. 5.4(g) to 5.4(h).

From Fig. 5.5, the prediction space of each evolutionary neural network can be observed to be relatively smooth with no sudden disconnections which infer that the networks are well trained with low chance of overfitting. For the experiment of limestone catalyst, PDSE produced a neural network with a depth of 3, having a topology of [173, 81, 1] and activation function of [ReLU, ReLU, Exp]. Non-catalytic experiment required a neural network with deeper depth of 5, topology of [512, 282, 68, 52, 1] and activation function of [tanh, elu, softmax, Relu, Exp]. A very interesting phenomenon in this topology is the evolution of a softmax function which re-normalized the weights in the middle of the neural network, leading to better performances. In the case of HZSM-5 zeolite catalytic experiment, the neural architecture is found to be [348, 214, 1] and activation of [Exp, Linear, Hard Sigmoid].

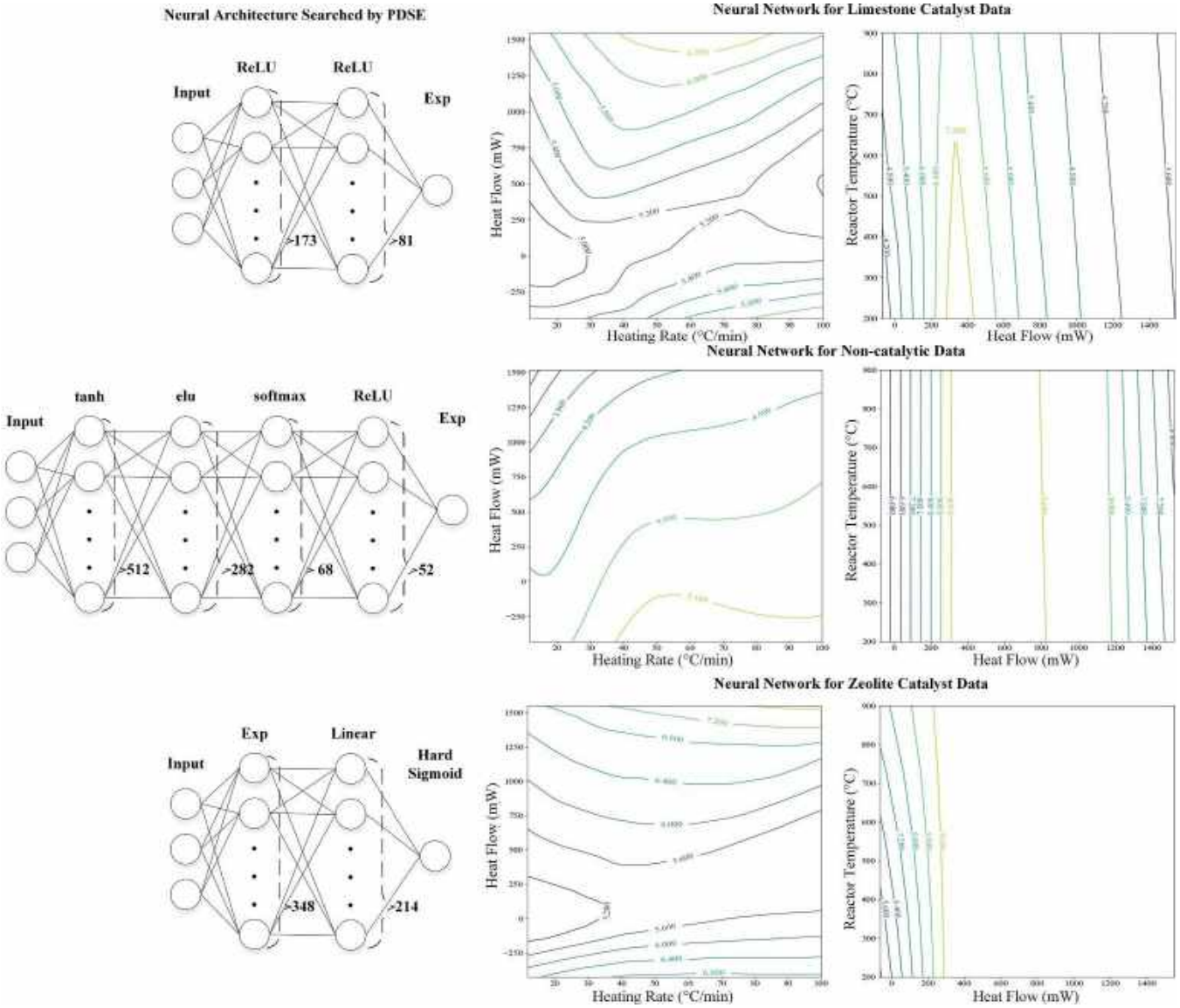


Figure 5.5: Neural architecture by PDSE and prediction space (contour represents mass remaining).

The best obtained validation metrics of the trained evolutionary neural networks are tabularized in Table 5.2. Validation RMSE of all experiments are simultaneously below 0.0075 (i.e., 0.0071 for the case using limestone as catalyst; 0.0075 for the case without using catalyst; 0.0050 for the case using HZSM-5 zeolite catalyst), demonstrating the robustness in the evolutionary neural network model. Furthermore, validation MBE were all smaller than a magnitude of 0.0026 (i.e., -0.0007 for the case using limestone as catalyst; -0.0010 for the case without using catalyst; -0.0026 for the case using HZSM-5 zeolite catalyst), this shows that the model is not biased towards a certain effect of association. It is also worth pointing out that the R^2 metric does not reflect the real performance of models when the values of two models are close (e.g. 0.9991 and 0.9996). A popular example to demonstrate this is when the predicted value is a dataset of [3,4,5] and actual values are [1,2,3], the R^2 value is 1.0 but the average error is 2. Alexander et al. (2015) also observed this phenomenon and encouraged the use of error metrics over R^2. Nevertheless, all models obtained validation R^2 value above 0.9990 (i.e., 0.9993 for the case using limestone as catalyst; 0.9992 for the case without using catalyst; 0.9996 for the case using HZSM-5 zeolite catalyst).

To underline the performance of evolutionary neural networks within the TGA modelling ecosystem, this work states the performances of the models within some related works. Firstly, a recent work from Xie et al. (2018) tested Radial Basis Function Network (RBFN) and Bayesian Regularized Network (BRN) modelling on the thermal analysis of textile dyeing sludge and pomelo peel. In the work, RBFN attained 0.8506 RMSE and 0.9989 R^2 while BRN attained 0.3277 RMSE and 0.9989 R^2. Buyukada (2016) also used Multi-Layered Perceptrons (MLP) to model TGA analysis of peanut hull and coal blend, giving a result of 1.5678 RMSE, 0.0501 MBE and 0.9999 R^2. For the feedstock of hazelnut husks and lignite coal, Yıldız et al. (2016) also obtained a result of 0.6240 RMSE, 0.4840 MBE and 0.9994 R^2 using MLP model. It is also worth pointing out that previous works related to the modelling of C. Vulgaris attained 0.9000 to 0.9800 regression R^2 values using Kissinger-Akahira-Sunose (KAS) model and 0.9180 to 0.9490 regression R^2 values using Flynn-Wall-Ozawa (FWO) model (Figueira et al., 2015). Considering a mixed feedstock of C. Vulgaris, wood and polypropylene, Azizi et al. (2017) obtained 0.6899 to 0.9906 regression R^2 values with KAS model and 0.7336 to 0.9910 regression R^2 values with FWO model. All previous work stated had provided admirable contribution to the fields of thermogravimetric analysis and can be considered as the current state-of-art. Contrarily, it is clear that the proposed modelling technique from this work

achieved results with two order of magnitude better in RMSE and at least one order of magnitude better in MBE than the current state-of-art results.

5.3.3 Optimized results using Simulated Annealing method

The bi-level optimization by Simulated Annealing to find an optimum thermal condition for the conversion of *C. vulgaris* took 27 minutes on a machine with Intel(R) Core(TM) i5-8250. In general, the Simulated Annealing algorithm performs well with increasing fitness on new generations within the algorithm. Fig. 5.6(a) is taken as an example (temperature range from 700.0 °C to 750.0 °C), the averaged fitness within the population was also steadily increasing, showing that searching agents within the population are gradually improving their search solution on average. The optimum conditions were found in generation 41. However, the algorithm was allowed to run until generation 100 for confirmation of optimality. At higher number of generations, the mass remaining fraction was already quite stable, attaining good values below 0.2. For other temperature ranges, very similar phenomena were observed, and authors think it is trivial to present similar figures. Overall, the Simulated Annealing algorithm shows good improvement of mass remaining (in mass fraction), while consistently have its population of searchers move to a better solution on average. This shows that an optimal point is statistically guaranteed to be found.

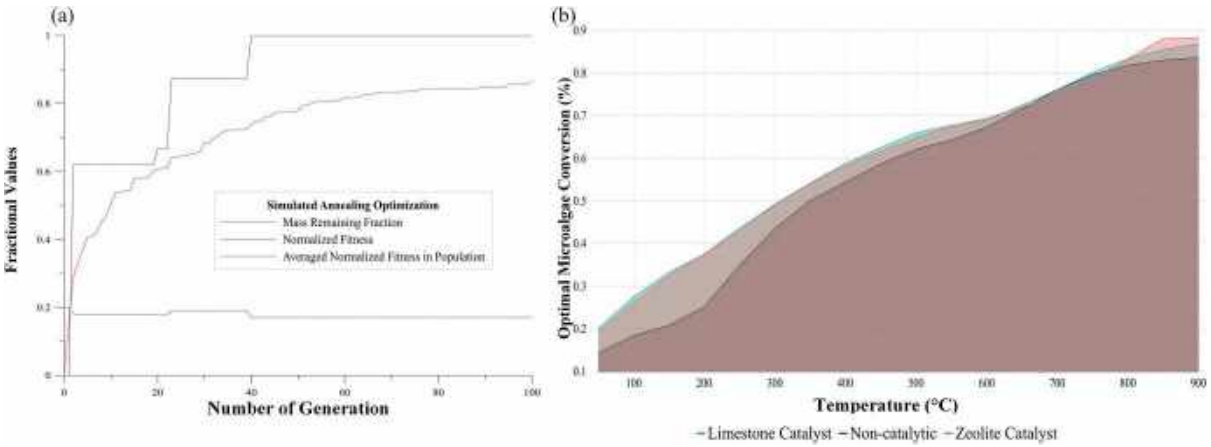


Figure 5.6: Optimization by simulated annealing (a) example of convergence curve at 700°C to 750 °C, (b) optimal conversion against reactor temperature.

Referring to the Pareto curve shown in Fig. 5.6(b), each point of the curve is the local optimal of conversion under a specified temperature, while the one with the highest conversion is denoted as global optimal under a specified temperature. At temperatures below 550.0 °C,

limestone catalyst showed a significant improvement on the *C. vulgaris* conversion. This is due to the alkali CaO element in limestone catalyst which reduced the activation energy of the pyrolysis reaction and resulting in a higher conversion of *C. vulgaris*. For this temperature range, the limestone catalyst results in higher conversion than HZSM-5 zeolite catalyst converting almost all the carbohydrates, protein, and lipid to bio-oil. Meanwhile at temperatures between 550.0 °C to 800.0 °C, both limestone and HZSM-5 zeolite catalysts shown almost similar catalytic effect in converting the *C. vulgaris* as shown in Table 5.3. Previous study has reported that Brønsted acid sites predominate in HZSM-5 zeolites will be activated at temperature above 500.0 °C and this could further enhance the secondary tar cracking such as dehydrogenation, decarboxylation, and aromatization to convert the light vapours and tar to form water, carbon dioxide, carbon monoxide, alkanes, alkene, methane and hydrogen (Vitolo et al., 2001). Thus, higher gaseous yield will be obtained in this stage. At temperature above 800.0 °C, the limestone catalyst started to lose its stability and catalytic activity due to the degradation of CaO to CO_2. Thus, HZSM-5 zeolite is a more suitable catalyst in thermal degradation process at temperature above 800.0 °C.

Table 5.2: Optimum conditions to enhance *C. vulgaris* conversion at different temperature range.

Temperature Range (°C)	Optimal Heating Rate (°C/min)	Optimal Heating Flow (mW)	Optimal Temperature (°C)	Best Catalyst	Microalgae Conversion (%)
30 - 50	5.0	300.0	50	Limestone	20.0
50 - 100	5.0	300.0	100	Limestone	27.4
100 - 150	5.0	300.0	150	Limestone	33.1
150 - 200	5.0	300.0	200	Limestone	37.5
200 - 250	5.0	300.0	250	Limestone	43.6
250 - 300	5.0	278.4	300	Limestone	49.2
300 - 350	5.0	76.1	350	Limestone	54.2
350 - 400	5.0	59.1	400	Limestone	58.7
400 - 450	15.8	0.0	450	Limestone	62.5
450 - 500	20.2	0.0	500	Limestone	64.9

Table 5.2 (Cont.): Optimum conditions to enhance *C. vulgaris* conversion at different temperature range.

500 - 550	19.9	0.0	550	Limestone	67.7
550 - 600	5.0	174.8	600	HZSM-5 Zeolite	69.4
600 - 650	5.0	0.0	650	HZSM-5 Zeolite	72.0
650 - 700	5.0	0.0	700	Limestone	76.1
	5.0	300.0	700	Non-Catalytic*	76.1
	5.0	0.0	700	HZSM-5 Zeolite**	75.0
700 - 750	5.0	0.0	750	Limestone	80.1
	5.0	300.0	750	Non-Catalytic**	79.6
	5.0	0.0	750	HZSM-5 Zeolite**	78.6
750 - 800	5.0	0.0	800	Limestone	83.3
	5.0	300.0	800	Non-Catalytic**	81.8
	5.0	0.0	800	HZSM-5 Zeolite**	83.3
800 - 850	5.0	0.0	850	HZSM-5 Zeolite	88.0
850 - 900	5.0	0.0	900	HZSM-5 Zeolite	88.3

*Non-catalytic can reach maximum conversion, but more heating flow is required. This is not the bi-level optimum, however, authors decide to present it.

**Near to optimum points that can be considered.

5.4 Conclusion

Thermal conversion of *C. vulgaris* was successfully modelled using neuro-evolutionary approach. Based on the benchmarking results, the simulation performance of the proposed approach significantly outperforms other conventional methods with lower RMSE and MBE values (> 90 % lesser error) compared to other methods, while maintaining a high R^2 value (> 0.9990). A bi-level optimization was performed using Simulated Annealing algorithm to determine the optimal operating conditions. The highest conversion of *C. vulgaris* (88.3 %) to bio-oil and gaseous product was achieved at a temperature of 900.0 ^{o}C, heating rate of 5.0 ^{o}C/min, and with the presence of HZSM-5 zeolite catalyst.

References

Abbas, T., Awais, M.M., Lockwood, F.C., 2003. An artificial intelligence treatment of devolatilization for pulverized coal and biomass in co-fired flames. Combust Flame 132(3), 305-318. doi.org/10.1016/S0010-2180(02)00482-0

Alexander, D.L., Tropsha, A. and Winkler, D.A., 2015. Beware of R 2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J Chem Inf Model 55(7), 1316-1322. doi.org/10.1021/acs.jcim.5b00206

Azizi, K., Moraveji, M.K., Najafabadi, H.A., 2017. Characteristics and kinetics study of simultaneous pyrolysis of microalgae Chlorella vulgaris, wood and polypropylene through TGA. Bioresource Technol 243, 481-491. doi.org/10.1016/j.biortech.2017.06.155

Bach, Q.V., Chen, W.H., 2017. Pyrolysis characteristics and kinetics of microalgae via thermogravimetric analysis (TGA): A state-of-the-art review. Bioresource Technol 246, 88-100. doi.org/10.1016/j.biortech.2017.06.087

Barnard, J.A., Hughes, H.W.D., 1960. The pyrolysis of ethanol. T Faraday Soc 56, 55-63.

Buyukada, M., 2016. Co-combustion of peanut hull and coal blends: Artificial neural networks modeling, particle swarm optimization and Monte Carlo simulation. Bioresource Technol 216, 280-286. doi.org/10.1016/j.biortech.2016.05.091

Carlson, T.R., Jae, J., Lin, Y.-C., Tompsett, G.A., Huber, G.W., 2010. Catalytic fast pyrolysis of glucose with HZSM-5: The combined homogeneous and heterogeneous reactions. J Catal 270(1), 110-124. doi.org/10.1016/j.jcat.2009.12.013

Chan, Y.H., Cheah, K.W., How, B.S., Loy, A.C.M., Shahbaz, M., Singh, H.K.G., Yusuf, N.R., Ahmad Shuhaili, A.F., Yusup, S., Ghani, W.A.W.A.K., Rambli, J., Kansha, Y., Lam, H.L., Hong, B.H., Ngan, S.L., 2019. An overview of biomass thermochemical conversion technologies in Malaysia. Sci Total Environ 680, 105-123. doi.org/10.1016/j.scitotenv.2019.04.211

Chen, X., Li, S., Liu, Z., Chen, Y., Yang, H., Wang, X., Che, Q., Chen, W., Chen, H., 2019. Pyrolysis characteristics of lignocellulosic biomass components in the presence of CaO. Bioresource Technol 287, 121493. doi.org/10.1016/j.biortech.2019.121493

Conesa, J.A., Caballero, J.A., Reyes-Labarta, J.A., 2004. Artificial neural network for modelling thermal decompositions. J Anal Appl Pyrol 71(1), 343-352. doi.org/10.1016/S0165-2370(03)00093-7

Costa, J.A.V., Morais, M.G., 2011. The role of biochemical engineering in the production of biofuels from microalgae, Bioresource Technol 102(1), 2-9. doi.org/10.1016/j.biortech.2010.06.014

Ding, S., Li, H., Su, C., Yu, J., Jin, F., 2011. Evolutionary artificial neural networks: a review. Artif Intell Rev 39(3), 251-260. doi.org/10.1007/s10462-011-9270-6

Fazilat, H., Akhlaghi, S., Shiri, M.E., Sharif, A., 2012. Predicting thermal degradation kinetics of nylon6/feather keratin blends using artificial intelligence techniques. Polymer 53(11), 2255-2264. doi.org/10.1016/j.polymer.2012.03.053

Figueira, C.E., Moreira Jr, P.F., Giudici, R., 2015. Thermogravimetric analysis of the gasification of microalgae Chlorella vulgaris. Bioresource Technol 198, 717-724. doi.org/10.1016/j.biortech.2015.09.059

Fister Jr, I., Yang, X.S., Fister, I., Brest, J. and Fister, D., 2013. A brief review of nature-inspired algorithms for optimization. arXiv preprint arXiv:1307.4186.

Floreano, D., Dürr, P., Mattiussi, C., 2008. Neuroevolution: from architectures to learning. Evol Intell 1(1), 47-62. doi.org/10.1007/s12065-007-0002-4

Gai, C., Zhang, Y., Chen, W.-T., Zhang, P., Dong, Y., 2013. Thermogravimetric and kinetic analysis of thermal decomposition characteristics of low-lipid microalgae. Bioresource Technol 150, 139-148. doi.org/10.1016/j.biortech.2013.09.137

Gan, D.K.W., Loy, A.C.M., Chin, B.L.F., Yusup, S., Unrean, P., Rianawati, E., Acda, M.N., 2018. Kinetics and thermodynamic analysis in one-pot pyrolysis of rice hull using renewable calcium oxide based catalysts. Bioresource Technol 265, 180-190. doi.org/10.1016/j.biortech.2018.06.003

Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. J Econometrics 60(1-2), 65-99. doi.org/10.1016/0304-4076(94)90038-8

Jaderberg, M., Dalibard, V., Osindero, S., Czarnecki, W.M., Donahue, J., Razavi, A., Vinyals, O., Green, T., Dunning, I., Simonyan, K. and Fernando, C., 2017. Population based training of neural networks. arXiv preprint arXiv:1711.09846.

Kim, S.-S., Ly, H.V., Choi, G.-H., Kim, J., Woo, H.C., 2012. Pyrolysis characteristics and kinetics of the alga Saccharina japonica. Bioresource Technol 123, 445-451. doi.org/10.1016/j.biortech.2012.07.097

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Mata, T.M., Martins, A.A., Caetano, N.S., 2010. Microalgae for biodiesel production and other applications: A review. Renew Sust Energ Rev 14(1): 217-232. doi.org/10.1016/j.rser.2009.07.020

Mayol, A., Maningo, J., Chua-Unsu, A., Felix, C., Rico, P., Chua, G., Manalili, E., Fernandez, D., Cuello, J., Bandala, A., Ubando, A., Madrazo, C., Dadios, E. and Culaba, A., 2018.

Application of Artificial Neural Networks in prediction of pyrolysis behavior for algal mat (LABLAB) biomass. In 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, Baguio City, Philippines. doi.org/ 10.1109/HNICEM.2018.8666376

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., 1953. Equation of state calculations by fast computing machines. J Chem Phys 21(6), 1087-1092. doi.org/10.1063/1.1699114

Mettler, M.S., Paulsen, A.D., Vlachos, D.G., Dauenhauer, P.J., 2014. Tuning cellulose pyrolysis chemistry: selective decarbonylation via catalyst-impregnated pyrolysis. Catal Sci Technol 4, 3822-3825. doi.org/10.1039/C4CY00676C

Mhaskar, H.N., Poggio, T., 2016. Deep vs. shallow networks: An approximation theory perspective. Analy Appl 14(6), 829-848. doi.org/10.1142/S0219530516400042

Milano, J., Ong, H.C., Masjuki, H.H., Chong, W.T., Lam, M.K., Loh, P.K., Vellayan, V., 2016. Microalgae biofuels as an alternative to fossil fuel for power generation. Renew Sust Energ Rev 58, 180-197. doi.org/10.1016/j.rser.2015.12.150

Mohr, A., Raman, S., 2013. Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. Energ Policy 63, 114-122. doi./10.1016/j.enpol.2013.08.033

Naqvi, S.R., Tariq, R., Hameed, Z., Ali, I., Taqvi, S.A., Naqvi, M., Niazi, M.B.K., Noor, T., Farooq, W., 2018. Pyrolysis of high-ash sewage sludge: Thermo-kinetic study using TGA and artificial neural networks. Fuel 233, 529-538. doi.org/10.1016/j.fuel.2018.06.089

Raymundo, L.M., Mullen, C.A., Strahan, G.D., Boateng, A.A., Trierweiler, J.O., 2019. Deoxygenation of Biomass Pyrolysis Vapors via in Situ and ex Situ Thermal and Biochar Promoted Upgrading. Energ Fuels 33(3), 2197-2207. doi.org/10.1021/acs.energyfuels.8b03281

Salimans, T., Ho, J., Chen, X., Sidor, S. and Sutskever, I., 2017. Evolution strategies as a scalable alternative to reinforcement learning. arXiv preprint arXiv:1703.03864.

Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. In 1998 IEEE international conference on evolutionary computation proceedings. In 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360). IEEE, Anchorage, USA. doi.org/10.1109/ICEC.1998.699146

Stanley, K.O., Clune, J., Lehman, J., Miikkulainen, R., 2019. Designing neural networks through neuroevolution. Nat Mach Intell 1(1), 24-35. doi.org/10.1038/s42256-018-0006-z

Stanley, K.O., Miikkulainen, R., 2002. Evolving neural networks through augmenting topologies. Evol Comput 10(2), 99-127. doi.org/10.1162/106365602320169811

Sun, J., Xiong, X., Wang, M., Du, H., Li, J., Zhou, D., Zuo, J., 2019. Microalgae biodiesel production in China: A preliminary economic analysis. Renew Sust Energ Rev 104, 296-306. doi.org/10.1016/j.rser.2019.01.021

Tan, K.C., Chiam, S.C., Mamun, A.A., Goh, C.K., 2009. Balancing exploration and exploitation with adaptive variation for evolutionary multi-objective optimization. Eur J Oper Res 197(2), 701-713. doi.org/10.1016/j.ejor.2008.07.025

Teng, S.Y., 2019. tsyet12/EvoOpt: EvoOpt v0.12 pre-release. doi.org/10.5281/zenodo.3241951

Vitolo, S., Bresci, B., Seggiani, M., Gallo, M.G., 2001. Catalytic upgrading of pyrolytic oils over HZSM-5 zeolite: behaviour of the catalyst when used in repeated upgrading–regenerating cycles. Fuel 80(1), 17-26. doi.org/10.1016/S0016-2361(00)00063-6

Xie, C., Liu, J., Zhang, X., Xie, W., Sun, J., Chang, K., Kuo, J., Xie, W., Liu, C., Sun, S., Buyukada, M., 2018. Co-combustion thermal conversion characteristics of textile dyeing sludge and pomelo peel using TGA and artificial neural networks. Appl Energ 212, 786-795. doi.org/10.1016/j.apenergy.2017.12.084

Yang, H., Norinaga, K., Li, J., Zhu, W., Wang, H., 2018. Effects of HZSM-5 on volatile products obtained from the fast pyrolysis of lignin and model compounds. Fuel Process Technol 181, 207-214. doi.org/10.1016/j.fuproc.2018.09.022

Yıldız, Z., Uzun, H., Ceylan, S., Topcu, Y., 2016. Application of artificial neural networks to co-combustion of hazelnut husk–lignite coal blends. Bioresource Technol 200, 42-47. doi.org/10.1016/j.biortech.2015.09.114

Zainan, N.H., Srivatsa, S.C., Li, F., Bhattacharya, S., 2018. Quality of bio-oil from catalytic pyrolysis of microalgae Chlorella vulgaris. Fuel 223, 12-19. doi.org/10.1016/j.fuel.2018.02.166

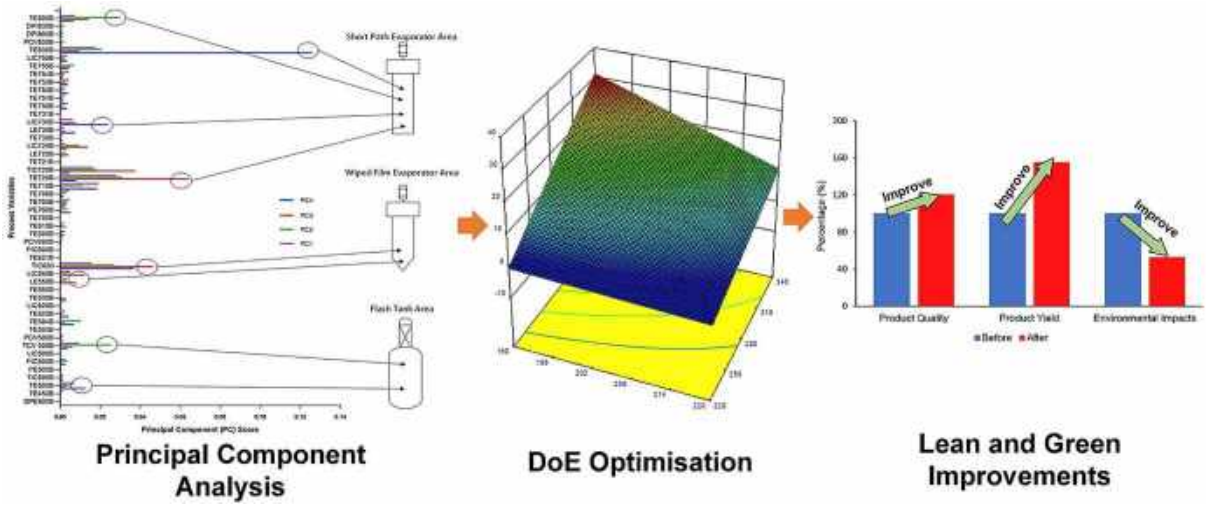
Zhang, X., Li, C., Tian, A., Guo, Q., Huang, K., 2019. Influence of CaO and HZSM-5 on oxygen migration in Chlorella vulgaris polysaccharide pyrolysis. Carbon Resources Conversion, 2(2), 111-116. doi.org/10.1016/j.crcon.2019.05.00

CHAPTER 6 DATA-DRIVEN MULTI-OBJECTIVE OPTIMIZATION FOR A PROCESS WITH MULTIPLE UNITS

This chapter has been peer-reviewed and published in Journal of Cleaner Production.

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Cheah, A.C.S., Motavasel, Z. and Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. Journal of Cleaner Production, 225, 359-375.

Graphical Abstract



Abstract

Integrated refineries and industrial processing plant in the real-world always face management and design difficulties to keep the processing operation lean and green. These challenges highlight the essentiality to improving product quality and yield without compromising environmental aspects. For various process system engineering application, traditional optimisation methodologies (i.e., pure mix-integer non-linear programming) can yield very precise global optimum solutions. However, for plant-wide optimisation, the generated solutions by such methods highly rely on the accuracy of the constructed model and often require an enumerate amount of process changes to be implemented in the real world. This work solves this issue by using a special formulation of correlation-based principal component analysis (PCA) and Design of Experiment (DoE) methodologies to serve as statistical process optimisation for industrial refineries. The contribution of this work is that it provides an efficient framework for plant-wide optimisation based on plant operational data while not compromising on environmental impacts. Fundamentally, PCA is used to prioritise statistically significant process variables based on their respective contribution scores. The variables with

high contribution score are then optimised by the experiment-based optimisation methodology. By doing so, the number of experiments run for process optimisation and process changes can be reduced by efficient prioritisation. Process cycle assessment ensures that no negative environmental impact is caused by the optimisation result. As a proof of concept, this framework is implemented in a real oil re-refining plant. The overall product yield was improved by 55.25% while overall product quality improved by 20.6%. Global Warming Potential (GWP) and Acidification Potential (AP) improved by 90.89% and 3.42% respectively.

Keyword: Principal Component Analysis, Design of Experiment, Plant-wide Optimisation, Statistical Process Optimization, PASPO, Big Data Analytics

6.1 Introduction

Common challenges occurring in the oil refinery industry are illiterate behaviour towards fading process performance and indiscriminate nature from the management level. As a problem statement, the abundance of operating parameters has been a formidable hurdle to process engineers in the determination of optimal operating conditions. For example, in the work of Joly et al. (2002), a case study of oil refinery optimisation requires 912 discrete variables and 5599 equations. In addition, many industry players tend to increase production capacity due to surges in market demand without careful and in-depth analysis of the equipment's capability and efficiency (Halvorsen et al., 2012). Inevitably, operating conditions deviates from optimal, process efficiency declines, and environmental performances are subpar coupling with deteriorating product quality.

The statistical approach is one of the simple yet effective ways to address all the mentioned issues. Notably, *Design of Experiments* (DoE) was first invented by Fisher (1935). The underlying principle is simply such that a sequence of test whereby intentional adjustment is made towards process variables and responses from the process are measured (Fisher, 1935). It can measure all correlations between process variables and responses by varying them simultaneously instead of individually. The process variables and responses are fitted in a mathematical model that used to effectively accelerate the optimisation (Toyota et al., 2017). DoE can also generate a predictive model for a great number of variables with a minimum experimental run (Gunst and Mason, 2009). Nevertheless, it is not practical for a complex problem which contains an extensive number of variables. In such problems, the required resources demand and computational time will increase exponentially with the increasing

number of variables within the problem boundary (Telford, 2007). This weakness has become apparent when DoE is being applied in problems related to Big Data (Drovandi et al., 2017). With growing needs for Big Data analytics in industries including chemicals, energy, semiconductors, pharmaceuticals, and food (Chiang et al., 2017), DoE becomes impracticable.

Fortunately, it is possible to prioritise the number of variables within a problem boundary through a multivariate statistical approach, called *Principal Component Analysis* (PCA). In brief, Hostelling (1933) formalised the novel instantiation of PCA. PCA is a well-known multivariate analysis method conducted to identify the principal components between variables that are interrelated and thereby reducing complicated data sets to smaller dimensions (Shlens, 2014). To-date, PCA has been broadly implemented in various research fields. Particularly for 3D imaging technology and machine learning algorithms which needs to deal with an enormous amount of data. For instance, Aida et al. (2017) have successfully improved the quality of images obtained from micro X-ray fluorescence analysis by performing PCA to improve the standard deviation of intensities peak. Ning and You (2018) reported that integrating PCA with kernel smoothing methods in the application of machine learning and data analysis for decision making has effectively reduced computational load and efficiency. Some other reported PCA applications are tabulated in Table 6.1. Despite the natures of these works being different, the motives for using PCA are certainly identical (i.e., to reduce the dimensionality of the research problem without losing too much information).

Table 6.1: A list of related PCA applications

PCA Application	**Authors**
Total energy efficiency assessment and optimization in manufacturing sectors	Azadeh et al. (2007)
Effective assessment of water quality network	Ou et al. (2012)
Multi-mode plant-wide process monitoring scheme for complex chemical industries	Jiang and Yan (2014)
Chiller sensor fault detection	Hu et al. (2016)
Dynamic response of commodity markets	Nobi et al. (2017)
Plant-wide process monitoring with minimal redundancy maximal relevance	Xu et al. (2017)
Fault detectability analysis in nuclear power plant	Li et al. (2018)
Biomass supply chain optimisation	How and Lam (2018a)
Biomass supply chain debottlenecking	How and Lam (2018b)

Plant-wide optimisation is often carried out using mathematical programming. Notable works related were reviewed by Klemes and Kravanja (2013), where pinch analysis and mathematical programming were used for process integration and optimisation. Furthermore, Klemes et al. (1997) also proposed a Total Site targeting and design methodology to optimise fuel, power, and carbon dioxide in processing plants. For these applications, Cucek et al. (2012) have highlighted the importance of footprint analysis tools for monitoring impacts on sustainability. The idea of integrating PCA and DoE which was initially proposed by Bratchell (1989) has been carried out in a few research works. For instance, Zhang et al. (2008) focused on the utility of PCA and DoE on dynamic systems. The work aimed to optimise the control performance of a yeast fermentation reactor model by using this hybrid framework. The combination of PCA and DoE has also been used to optimise chemical reactions. Murray and Forfar (2017) have used this hybrid method for solvent and ligand selection of three different chemical reactions. In their work, the DoE factors required for optimisation was reduced from 35 to 19 factors, which also reduced the number of experiments from 51.2 million to 6400. Furthermore, PCA and DoE have been utilised for the optimisation of a turning unit in the

works of Madhavi et al. (2017). In which, the operating conditions of a single turning unit are optimised to give optimal product hardness and surface roughness.

All the reviewed works are admirable, however, none of them has tested the capability and applicability of the proposed method with the validation in a real chemical plant for plant-wide optimisation. This work implements a novel usage of PCA and DOE that is formulated specifically for plant-wide optimization, called the Principal Component Analysis-aided Statistical Process Optimisation (PASPO). The novel PASPO framework reduces the dimensionality of the plant-wide optimization by screening computed correlation-based principal components while decoupling and recombining the principal components into process variables for critical variable selection. This work also demonstrates the effectiveness of the PASPO framework by using an actual industrial processing plant as a case study. The PASPO framework is aimed to reduce analysis time and cost, minimise process changes required for a relatively good benefit while making plant-wide optimisation more data-oriented (instead of model-oriented). Furthermore, environmental impact analysis is carried out to study the environmental performance of the process. To achieve this, a performance indicator known as *Process Cycle Assessment* (a simplification of *Life Cycle Analysis* (LCA) to target process systems) is developed to allow instant and effortless assessment of environmental performance.

6.2 Methodology

The proposed PASPO framework started off with performing correlational PCA on collected data to obtain the principal components which would add up and capture more than 90% of the variance. These principal components are then decoupled into individual process variables and recombined as contribution scores. Using different coverage of contribution scores, DoE is executed on the statistically significant variables to generate a regression model in Design Expert software. Due to the complexity of the data, the regression model is very high in mathematical dimensions. Hence the best method to visualise such models is by plotting multiple three-dimensional response surface diagram of significant relations. The regression model is then numerically optimised by maximizing product yields and quality responses. It is utilised to establish a combination of optimal operating conditions whilst ensuring that product quality meets the requirement and environmental performance is improved (see Figure 6.1). The optimal operating conditions can then be plotted in a solution diagram to show its desirability and its relativeness with the high and low limits. The newly established optimum

operating conditions are tested out by Process Cycle Assessment to further analyse the environmental performances.

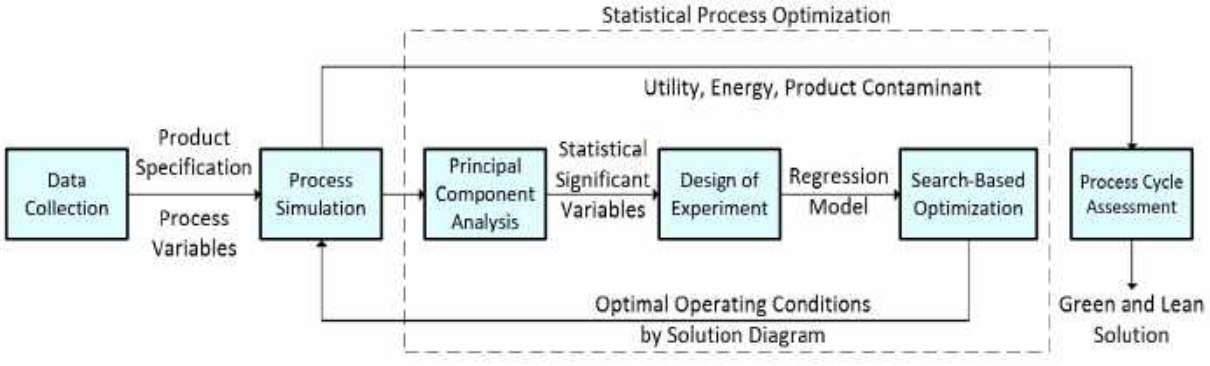


Figure 6.1: Overall proposed methodology for PCA-aided statistical process optimisation (PASPO)

The overall strategies are as below:

1. Analyse the process and plan data collection strategy
2. Data is collected for all unit operations such as operating conditions, equipment capacity and safety limits
3. The process is modelled using appropriate process simulation tool (Aspen HYSYS)
4. PCA is performed to reduce the dimensions of data
5. DoE is performed to generate a regression model
6. The regression model is used to plot a surface response curve
7. Surface response plots are utilised to visualise the regression model and study the relations between process variables.
8. Numerical optimisation is used to find an optimal operating condition that would maximise yields and quality.
9. Plot solution graph to show desirability of solution and relativeness to process limits.
10. Optimised operating conditions is inputted into the simulated process model in step 3
11. Product quality before and after statistical optimisation are compared
12. Process Cycle Assessment is performed to evaluate the environmental performance of the process (before and after statistical process optimisation)

The detailed methodology for each strategy is demonstrated based on the Pentas Flora case study (introduced in Chapter 6.2) in the subchapters below.

6.3 Case Study

The work focuses on addressing the optimisation problem in a waste oil re-refinery plant. In general, a waste oil re-refinery plant is a refinery plant which aims to recover the quality of the used oil. In this work, a real industrial case study from Pentas Flora Sdn. Bhd. is applied to show the effectiveness of the proposed method. Firstly, the waste oil is collected from Peninsular, Malaysia and sent into a series of pre-treatment facilities, i.e., flash tank dehydration unit remove moisture and light hydrocarbons from waste oil. The remaining oil is then fed into a vacuum wiped film evaporator (WFE) which has high efficiency and minimal production degradation for further processing. From there, oil is separated into light lube oil (WFE product) and heavy oil. Following is vacuum short path evaporator (SPE) that further separates the heavy oil into asphalts and medium lube oil (SPE product) which is the main product of this waste oil refinery process. Figure 6.2 demonstrates the overall process flow in Pentas Flora Sdn. Bhd. To achieve higher lean and green attainments, a lean and green optimisation framework is therefore presented.

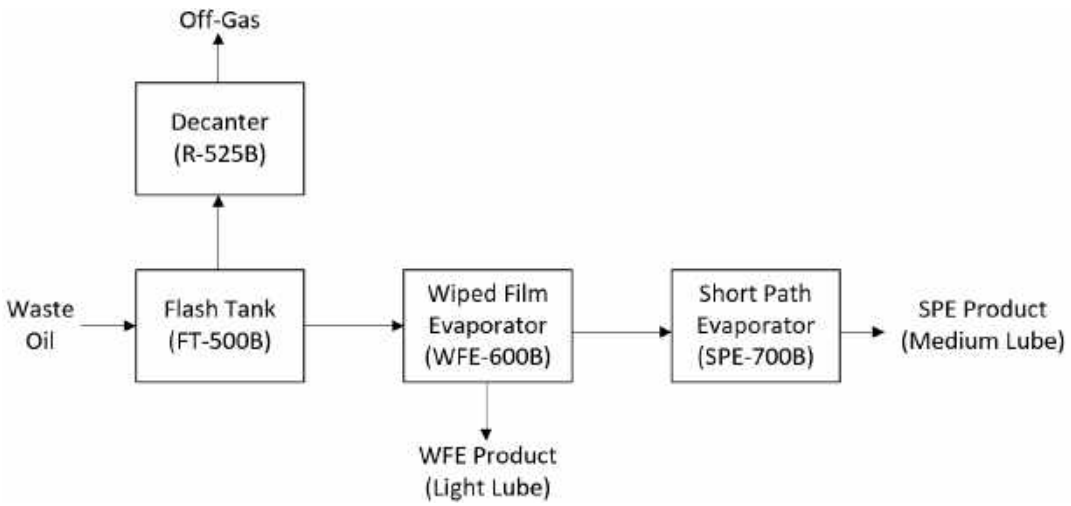


Figure 6.2: Simplified block flow diagram of oil re-refining case study

6.3.1 Data Collection

Data of the process is collected systematically to ensure all the potential correlations between the operating conditions and the responses are captured by the model. Therefore, this chapter presents the data collection step of this work. There are two chapters of data collection, i.e., operating conditions and product specifications.

The steady-state operating conditions of all equipment in the process (e.g. temperature, pressure and flowrate) were recorded for 120 minutes with 5 minutes' interval. Step-changes are introduced towards the process to determine the significances of all operating parameters by calculating their covariance within the PCA study (see Table 6.2). Likewise, the operating conditions of all equipment are extracted after the introduction of step-changes for a time span of 2 hours with 5 minutes' interval.

Table 6.2: Step-changes introduced into the process for data collection

Step Change Variable	**Location**	**Deviation**	**Study Duration (minutes)**	**Data Interval (minutes)**
No change	Full process in steady state	-	120	5
Temperature	Bottom temperature (TIC500B) of Flash tank	Increase by 5 °C	120	5
Flowrate	Waste oil feed flowrate (FIC450B)	Increase by 5 m^3/h	120	5

Oil samples were collected at waste oil feed, post flash tank product, WFE product, SPE product and asphalt during steady-state operation and two hours after step-changes were introduced. Followed by a series of lab tests in Pentas Flora Sdn. Bhd. which are Malaysian Standard Accredited to determine the properties of oil with appropriate standards and equipment as shown in Table 6.3.

Table 6.3: Lab test, codes, and equipment

Properties	ASTM Code	Lab Equipment
Sulphur Content	ASTM D4294	X-ray fluorescence spectroscopy
Specific Gravity	ASTM D1298	Hydrometer
Moisture Content	ASTM D6304	Coulometric Karl Fisher
Flash point	ASTM D93	Pernsky Marten Close Cup Flash Point
Viscosity	ASTM D4684	Mini-rotary Viscometer (MRV)
Boiling Point Range	ASTM D2887	Gas Chromatography-Mass Spectroscopy (GCMS)
Molecular Weight	ASTM D1481	Pycnometer

6.3.2 Process Simulation

Process simulation is carried out using HYSYS V8.8. The fluid package chosen is Sour Peng-Robinson (Sour PR) as it provides a good estimation for hydrocarbons and process consisting of hydrogen sulphide (H_2S) contaminant (Aspen Process Engineering Webinar, 2006). The waste oil feed is then modelled using an oil blend function, while its accuracy is further enhanced by incorporating bulk properties of waste oil and boiling point range (BPR) identified from lab testing as shown in Table 6.4.

Table 6.4: Bulk Properties and boiling point range of waste feed oil

Oil Sample	Feed oil
Density (kg/m^3)	877.10
Moisture content (% w/w)	14.27
Sulphur content (ppm)	4719.70
Boiling point range	
Initial Boiling Point (IBP)	133°C
10%Vol.	141°C
20%Vol.	182°C
30%Vol.	236°C
40%Vol.	286°C
50%Vol.	326°C
60%Vol.	353°C
70%Vol.	373°C
80%Vol.	399°C
90%Vol.	450°C
Final Boiling Point (FBP)	546°C

Subsequently, a simulation process flow sheet is generated as shown in Figure 6.3. The robustness of the model is tested by comparing the simulated and actual quality of WFE and SPE product (see Table 6.5). Note that the viscosity of asphalt is not considered in the robustness test due to technical constraint (i.e., its viscosity exceeds the measuring limit of the viscometer in the lab). Evidently, the results show the deviations between simulated and actual data are less than 10%. In other words, this indicates that the developed model is capable to provide accurate results as compared to the real scenario.

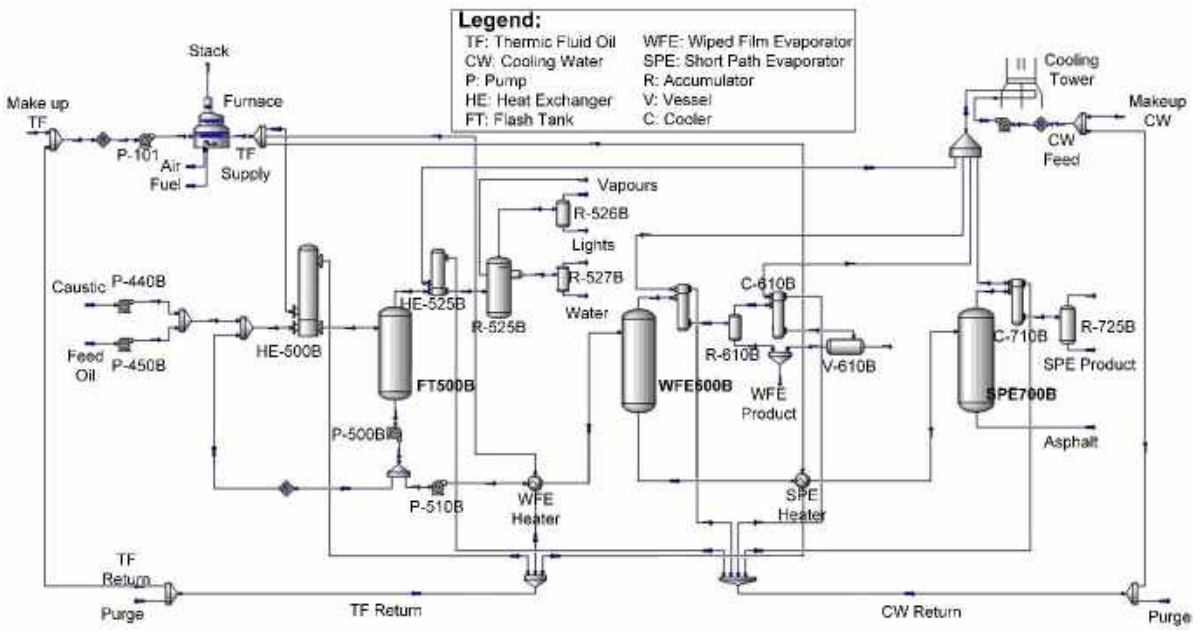


Figure 6.3: Simulation process flow diagram (main processing units in bold)

Table 6.5: Comparison between simulated and actual product specification

Oil Product	Kinematic Viscosity (cSt)			Density (kg/m^3)		
	Simulated	**Actual**	**Deviation (%)**	**Simulated**	**Actual**	**Deviation (%)**
WFE Product	16.82	15.37	9.43	891.5	848.7	5.04
SPE Product	33.49	31.81	5.28	892.1	854.8	4.36
Asphalt		-	-	938.7	935.6	0.33

By doing a paired t-test on the simulated and actual data, the two-tailed p-value is 0.1384 (t=1.8473). This shows that by conventional criteria, the difference between the simulated and actual data is not statistically significant (Detailed paired t-test analysis in Appendix Table A6.16).

6.4 Statistical Process Optimization

The aim of the statistical process optimization is to find the optimum operating conditions for the entire process. However, most of the variables are insignificant towards product quality. Thus, it is unnecessary to include these parameters in the optimisation model. To address this

issue, the PASPO framework which integrates PCA and DoE methodologies is proposed. The conceptual idea of this hybrid framework is illustrated in Figure 6.4. Initially, PCA is conducted to prioritise the variables based on their respective contribution scores. Only those (variables) with high contribution score are considered in DoE methodology. With the aid of the PCA methodology, the required number of experimental runs is expected to be reduced. This further lead to lower use of experimental resources (e.g. raw material) and time spent (e.g. working hours). Hence, the overall cost (material cost, energy cost, operator wages, etc.) is gradually reduced.

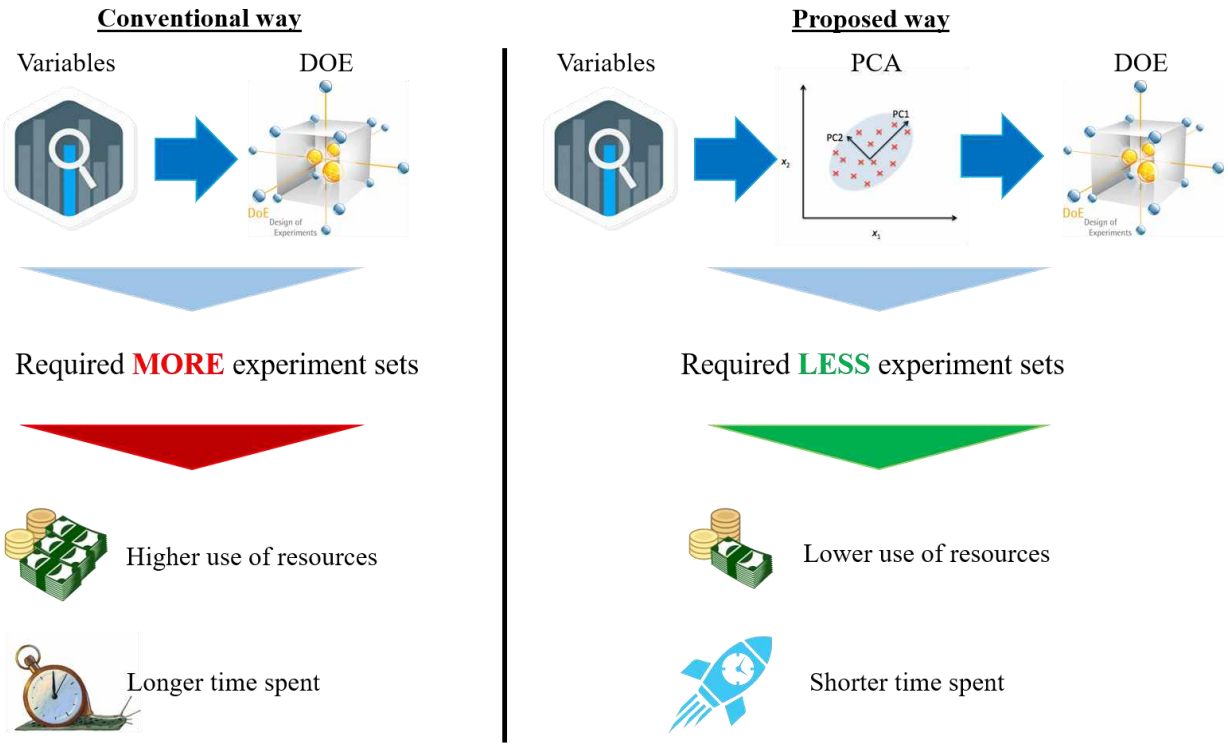


Figure 6.4: Conceptual idea of the PASPO framework

6.4.1 Principal Component Analysis (PCA)

PCA is a multivariate statistical technique capable of shrinking the dimensions of a data set consisting of innumerate correlated variables while ensuring most variations in the data set are captured (Pearson, 1901). Dimension reduction is achieved by converting the correlated variables from the original data sets into linearly uncorrelated variables, called *principal components* (PC). Traditionally, PCs' are determined by solving the inverse eigenvector of the covariance matrix. However, Jolliffe and Cadima (2016) discussed that traditional covariance based PCA will give a poor representation of data when applied to combinations of variables with different units of measurement. Thus, the correlation method which involves

normalisation of original data sets is performed instead of the covariance method (How and Lam, 2018b). The equation to evaluate the correlation between variables (Al-Sayed, 2015) is shown in Eqn. 6.1 below.

$$corr(x_\alpha, x_\beta) = \frac{1}{n-1}\sum_1^n \left(\frac{x_\alpha - \bar{x}_\alpha}{\sigma_{x_\alpha}}\right)\left(\frac{x_\beta - \bar{x}_\beta}{\sigma_{x_\beta}}\right) \quad (6.1)$$

In Eqn. 1, x_α and x_β are the comparative variables; $\bar{x}_\alpha$ and $\bar{x}_\beta$ are the average values of the corresponding variables; while σ_{x_α} and σ_{x_β} are the standard deviations of the corresponding variables, and n is the number of variable sets. The PC is evaluated from the correlational matrix, *A*, by solving an eigenvector-eigenvalue problem, as shown in Eqn. 6.2. The first PC (or PC1) is responsible for the majority of variance in the data, followed by second PC (PC2) and so forth.

$$A\,v = \lambda v \quad (6.2)$$

Where *A* refers to the correlation matrix, v refers to the eigenvector representing the regression coefficient of the principal components, while **λ** refers to the eigenvalue which represents process variance (Shlens, 2014). Since the variables are multivariate and direct compiling the variables into the matrix will result in a less accurate model due to weight disproportion. A normalisation of process variables is required as shown in Eqn. 6.3 below.

$$x^{normalized} = \frac{x - \bar{x}}{\sigma_x} \quad (6.3)$$

To determine the numbers of PC included, the scree plot method with a heuristic minimum cumulative variance of 90% is used (Jackson, 1993). This is to ensure that the captured information is significant and data loss is acceptable (Rea and Rea, 2016).

$$Cum.Var_k \geq 90\ \% \quad (6.4)$$

Next, the dimensions the multi-variable process inputs can be expressed as PC and be assessed using a scoring method (How and Lam, 2018b). The equation is shown in Eqn. 6.5, where X refers to the normalised process variable matrix.

$$PC\ Score = Xv \tag{6.5}$$

In this work, the contribution of each variable is evaluated using an absolute method.

$$Contribution\ Score_{b,z} = \frac{|e_{b,z}|}{\sum_b |e_{b,z}|} \qquad \forall b \in B,\ \forall z \in Z \tag{6.6}$$

To select critical variables as a representation of the full information, the contribution scores for each variable are sorted from large contribution to low contribution to plot a second scree plot. The cumulative contribution score is used as an indication for consideration of process variables for optimisation.

6.4.2 Design of Experiment (DoE)

The statistical significant process variables obtained from PCA in Chapter 6.3.3.1 are the input variables for the experiments, also known as factors, whereas the results from the experiments are recognised as responses. Subsequently, the predictive model is generated based on the changes in factors and responses from the experiments by regression analysis (Fisher, 1935). Response surface methodology (RSM) is used to generate multiple surface response plots which are used for the latter optimisation. A full factorial methodology is adopted for this framework, as the model includes complete information on the process data (Collins et al., 2009) for optimisation. Due to the inherent nature of process systems being highly complex (McKay et al., 1997), this work considers up to 4th order of interaction factor. Subsequently, an automatic selection algorithm with p-value as the criterion is used to remove terms that are detrimental (Anderson, 2018).

The multi-objective optimisation technique used is a two-step optimisation coupled with the desirability function. The desirability function approach is to convert each surface response into a desirability score d_i with a range of 0 to 1 (Derringer and Suich, 1980). The overall desirability can be expressed as the following, where m is the total number of responses.

$$D = (d_1 d_2 \cdots d_m)^{1/m} \tag{6.7}$$

The individual desirability function for a response, *y* with a maximum requirement is shown in Eqn. 6.8 below.

$$d = \begin{cases} 0 & y < T \\ \left(\frac{y-L}{T-L}\right)^r & T \leq y \leq U \\ 1 & y > U \end{cases} \quad (6.8)$$

For a response targeting minimum value, the equation will be as the following.

$$d = \begin{cases} 1 & y < T \\ \left(\frac{U-y}{U-T}\right)^r & T \leq y \leq U \\ 0 & y > U \end{cases} \quad (6.9)$$

For the above equations, *L* and *U* are the lower and upper limits respectively; *r* is the weight of the response. Setting $r>1$ prioritises the corresponding response while choosing $0<r<1$ makes the response less important. Commonly, *r* is set to be one of the five standard levels as shown in Table 6.6 below (Kraber, 2009).

Table 6.6: Importance level and r value of desirability function

Importance Level	**1**	**2**	**3**	**4**	**5**
Pulses	+	++	+++	++++	+++++
***r* value**	10^{-1}	$10^{-0.5}$	10^{0}	$10^{0.5}$	10^{1}

In this work, the importance level of each objective is prescribed by managerial decisions after evaluating market economics and product requirements. The pulses for importance level are presented in Table 6.7.

Table 6.7: Pulses for the importance of objectives

Properties	**SPE Product**	**WFE Product**
Yield	+++++	+++
Quality	+++	+++

Based on processing requirements, product yield is drastically more important than quality. Hence, a two-step optimisation method is applied to the desirability function for yield and quality sequentially as shown below.

$Step\ 1$: $Max\ D_{yield}$ $s.t.\ T_k \leq y_{quality,\ k} \leq U_k$ $\forall k \in K$ (6.10)

$Step\ 2$: $Max\ D_{quality}$ $s.t.\ T_j \leq y_{yield,\ j} \leq U_j$ $\forall j \in J$ (6.11)

In addition, factors should be manipulated within minima and maxima boundary conditions based on equipment capacities and safety limits for the desired responses. Lastly, product quality is compared before and after DoE is performed.

6.5 Process Cycle Assessment

Utilities, energy, and equipment performance are recorded prior to and after statistical process optimisation. Impact category is chosen based on the presence of indicators in the process and similarly for the characterisation model. Lastly, the scores for the selected impact category before and after statistical process optimisation are compared. Process Cycle Assessment evaluates the process based on impact categories which are considered in LCA (WBCSD Chemicals, 2013). They include global warming potential (GWP) and acidification potential (AP), which are evaluated by the following equation:

$$GWP = \sum_{j=1}^{Stream} M_{emission,j} \sum_{k=1}^{Component} x_{j,k}\, \psi_{GWP,k} \quad (6.12)$$

$$AP = \sum_{j=1}^{Stream} M_{emission,j} \sum_{k=1}^{Component} x_{j,k}\, \psi_{AP,k} \quad (6.13)$$

The assessment mainly considers mass flowrate of emission stream which is denoted by $M_{emission,j}$ for stream j. Mass fraction of contaminant component k in stream j is expressed as $x_{j,k}$, while the specific potential environmental impact is expressed as ψ. Moreover, statistical process optimisation only considers the operating conditions of the process thus, Process Cycle Assessment only assesses the performance of the equipment and the final product.

6.6 Results

Large sets of processing data are collected from the Supervisory Control and Data Acquisition (SCADA) system of the oil re-refinery plant which enabled the effective use of principal component analysis and design of experiments. The validated process simulation model is also used to assist the design of experiments and for the case of 99% coverage score benchmark in Chapter 6.4.3. The following chapters cover the detailed result of the principal component analysis, design and analysis of experiments, optimisation results of different coverage score and process cycle assessment.

6.6.1 Principal Component Analysis Result

Using the computed principal component (PC) as discussed in chapter 6.3.3.1, the scree plot can be generated. The PCs are sorted from the highest eigenvalue to the lowest, while the cumulative eigenvalue (also cumulative variance) is plotted.

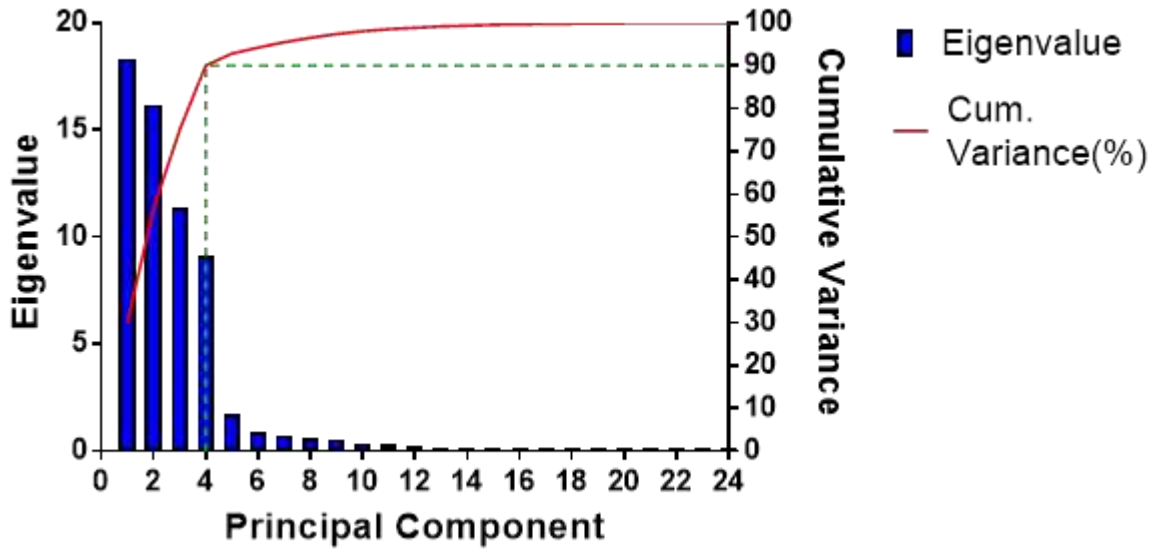


Figure 6.5: Scree plot of principal component

From Figure 6.5, there is an obvious eigenvalue drop in between the fourth PC and the fifth PC, this is commonly referred to as the “knee point”. This shows that considering the first four PC is representative of the full information, while further including extra PC gives less significant results. Additionally, having four PC also satisfies the criteria of 90% minimum cumulative variance. Therefore, it is determined that four PCs will be satisfactory for this work. Subsequently, the most significant four PCs are decoupled back into process variables that contribute to them and are plotted in Figure 6.6.

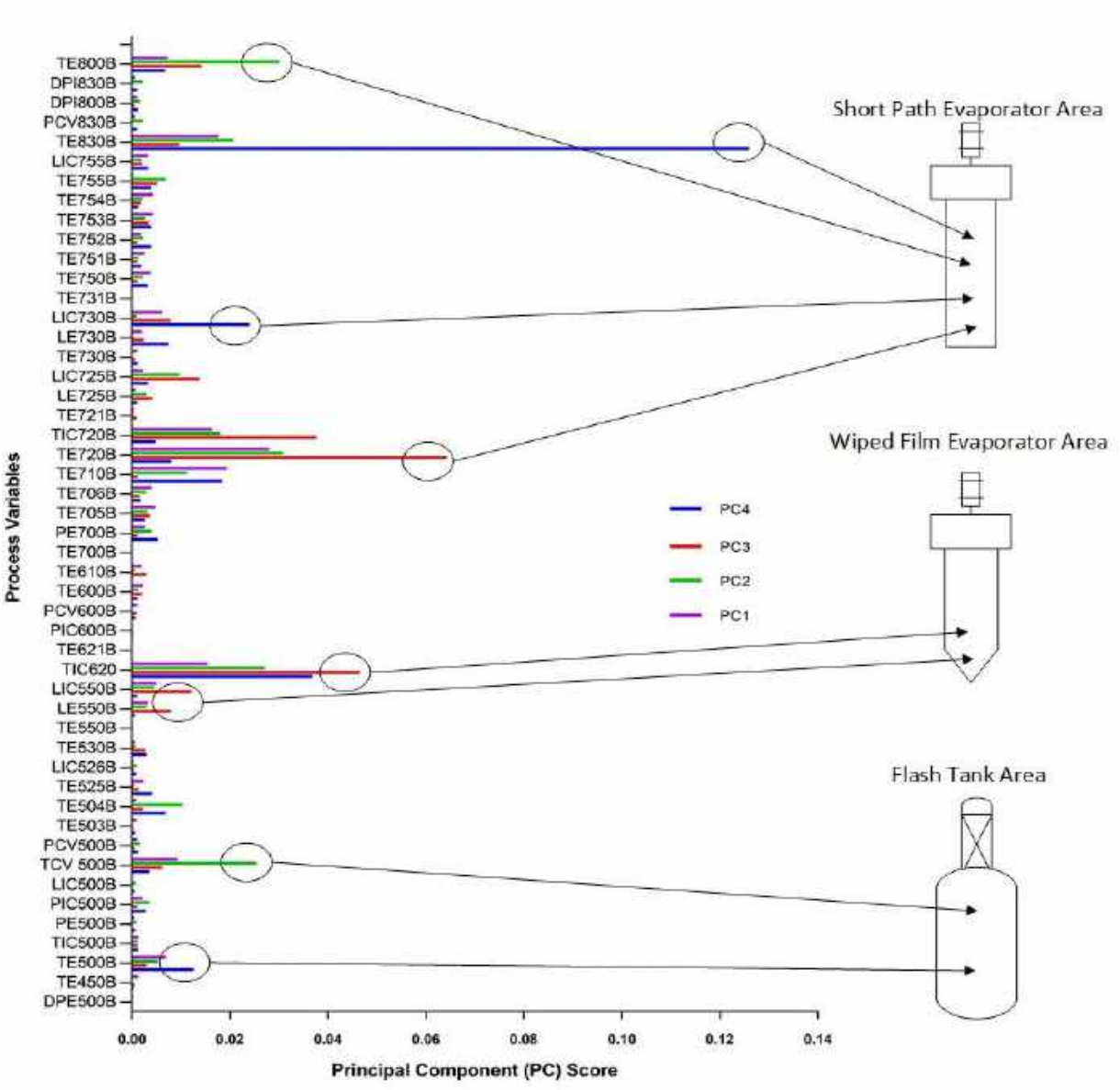


Figure 6.6: Contribution of Process Variables on Principal Components

Having the critical PC decoupled into process variables, some highly significant peaks are observed to highly contribute to the processing system. Variables such as flash tank temperature, WFE temperature and level, SPE temperature and level were immediately identified to be significant in the process. These parameters that are identified decoupling the PC are highly logical, as temperature and levels within separators are variables that affect the separation efficiency and thermodynamics of the system. However, lesser significant variables are difficult to be studied in Figure 6.6, as the contribution of each variable has not been combined for direct comparison.

Contribution scores of each process variables are reconstructed into a total contribution score. To ease the analysis, the process variables are sorted from the highest contribution score to the lowest with the cumulative contribution score (coverage score) plotted in Figure 6.7.

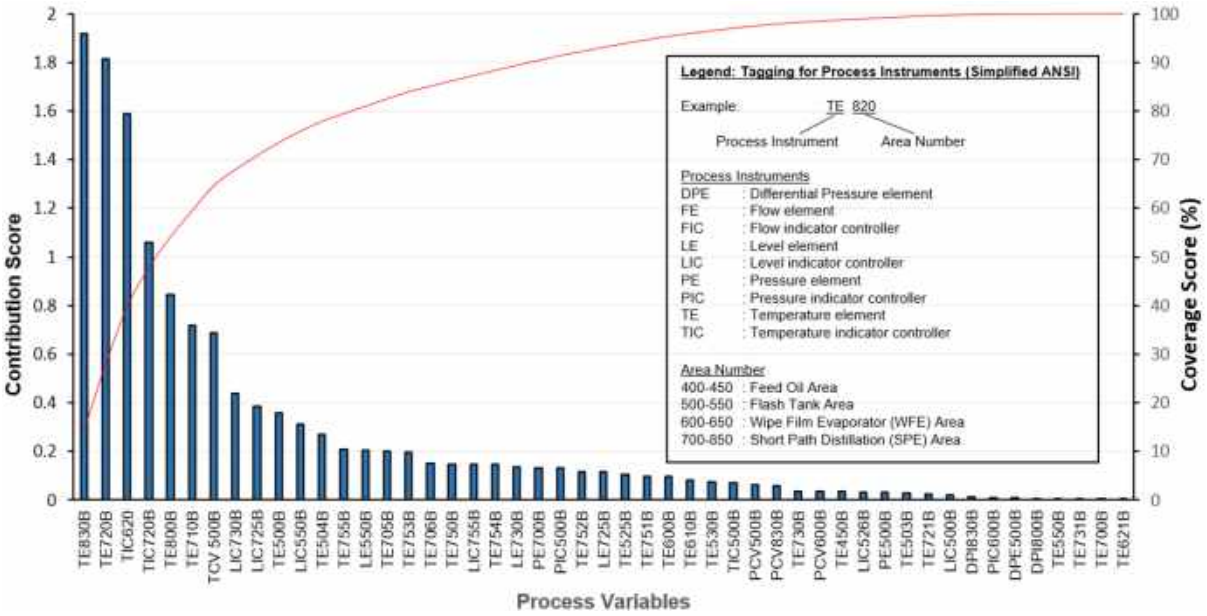


Figure 6.7: Prioritisation of process variables using contribution score

The factors that will be focused on during DoE are operating temperature of the flash tank, WFE, SPE, operating pressure in WFE and SPE, operating level in WFE tank and SPE. These variables are selected based on a 95% coverage score and can be directly manipulated in the processing system. Furthermore, a study of 80% and 99% coverage score was carried out as a comparison. The 80% cumulative score case considers the temperature of WFE, SPE and flash tank, while 99% coverage score case considers all variables in 95% case with the addition of decanter level, decanter pressure and flash tank pressure.

6.6.2 Design and Analysis of Experiments

According to PCA result, factors considered for DoE includes operating temperature of flash tank (A), temperature of WFE (B), temperature of SPE (C), operating pressure of SPE (D), level of WFE (E) and level of SPE (F), while WFE (y_1) and SPE (y_2) product flow, WFE (y_3) viscosity and SPE (y_4) viscosity are the corresponding response in this study. The factorial design is used to generate desired responses. In addition, the design matrix is shown in Table 8, the "+" and "–" signs represent treatment combinations for the factors. As illustrated, there

are seven degrees of freedom (DOF) for eight treatment combinations in which three DOF were associated with main effects of factor A, B, C, D and E (Table 6.8).

Table 6.8: The design matrix for factorial design considering six factors

Run	A	B	C	D	E	F	Labels
1	-	-	-	-	-	-	1
2	+	-	-	-	-	-	A
3	-	+	-	-	-	-	B
4	-	-	+	-	-	-	C
5	-	-	-	+	-	-	D
6	-	-	-	-	+	-	E
7	-	-	-	-	-	+	F
8	+	+	-	-	-	-	AB
9	+	-	+	-	-	-	AC
10	+	-	-	+	-	-	AD
11	+	-	-	-	+	-	AE
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	-	+	+	+	+	-	BCDE
54	-	+	+	+	-	+	BCDF
55	-	+	+	-	+	+	BCEF
56	-	+	-	+	+	+	BDEF
57	-	-	+	+	+	+	CDEF
58	+	+	+	+	+	-	ABCDE
59	+	+	+	+	-	+	ABCDF
60	+	+	+	-	+	+	ABCEF
61	+	+	-	+	+	+	ABDEF
62	+	-	+	+	+	+	ACDEF
63	-	+	+	+	+	+	BCDEF
64	+	+	+	+	+	+	ABCDEF

Subsequently, factors and the responses were fitted with the regression model as shown in Eqn. 6.14. Responses are denoted by y_j, the regression coefficient is indicated by β_j whereby j can

be any of the desired responses projected by factors considered. Thus, each response will have a dedicated regression model.

The regression model for each response is used to generate a surface response plot as illustrated and explained below. The considered factors are up to 4th order interaction factors, then detrimental factors are removed using hierarchical automatic model selection algorithm with p-values as the criterion. All the considered possibilities of interaction factors considered to give an optimal surface response for y_j is shown in the equation below.

$$\begin{aligned} y_j \;=\; & \beta_{0,j} + \beta_{A_j}A + \beta_{B_j}B + \beta_{c_j}C + \beta_{D_j}D + \beta_{E_j}E + \beta_{F_j}F + \beta_{AB_j}AB + \beta_{AC_j}AC \\ & + \beta_{AD_j}AD + \beta_{BC_j}BC + \beta_{BD_j}BD + \beta_{CD_j}CD + \beta_{CE_j}CE + \beta_{CF_j}CF + \beta_{DE_j}DE \\ & + \beta_{DF_j}DF + \beta_{EF_j}EF + \beta_{ABC_j}ABC + \beta_{ABD_j}ABD + \beta_{ACD_j}ACD + \beta_{BCD_j}BCD \\ & + \beta_{CDE_j}CDE + \beta_{CDF_j}CDF + \beta_{CEF_j}CEF + \beta_{DEF_j}DEF + \beta_{ABCD_j}ABCD \end{aligned} \tag{6.14}$$

Analysis of WFE product flow (LPM) in the assessment of experimental design ranged from 13.97 to 30.65 LPM. The response was fitted in a regression model as shown in Eqn. 6.15.

$$y_1 = 22.601 - 1.003A + 7.345B + 0.339E + 0.348F + 0.210AB + 0.339EF \tag{6.15}$$

The most significant effect on the WFE product flow is the WFE temperature (B), which has the largest coefficient of 7.34. An inversely proportional relation can be observed between flash tank temperature (A) and WFE product flow. This is highly possible from the perspectives of separation sciences, as a higher temperature at the flash tank will cause the oil fraction to evaporate into the vapour fraction, reducing the amount of WFE oil products. SPE temperature (C) and pressure (D) have minimal effects on the WFE product flow. This is not surprising since the operating conditions of a downstream separation unit should give minimal effects to the units beforehand. However, level in WFE (E) and SPE (F) contribute to giving a higher WFE product flow, demonstrating positive coefficients of 0.339 and 0.348, respectively. Significant interaction factors are the interaction factors between flash tank temperature, WFE temperature (AB) and WFE level and SPE level (EF). Technically, the contributing interaction factors represent temperatures and levels of consecutive processing units.

In addition, Figure 6.8 shows that the regression model for WFE product flow as a function of flash, WFE, SPE operating conditions was suitable to investigate the tendency of this response (detailed ANOVA table can be found in Appendix Table A6.17). Thus, a response surface is generated by plotting two of the main effects with the highest coefficient magnitude (flash and WFE temperature) with respect to the response. Evidently, WFE product flow is dependent on both flash and WFE temperature as illustrated in Figure 6.9 where the colour of the response surface is warmer when WFE temperature increased and cooler when flash temperature increased. In all, postulation made earlier from the regression model is validated.

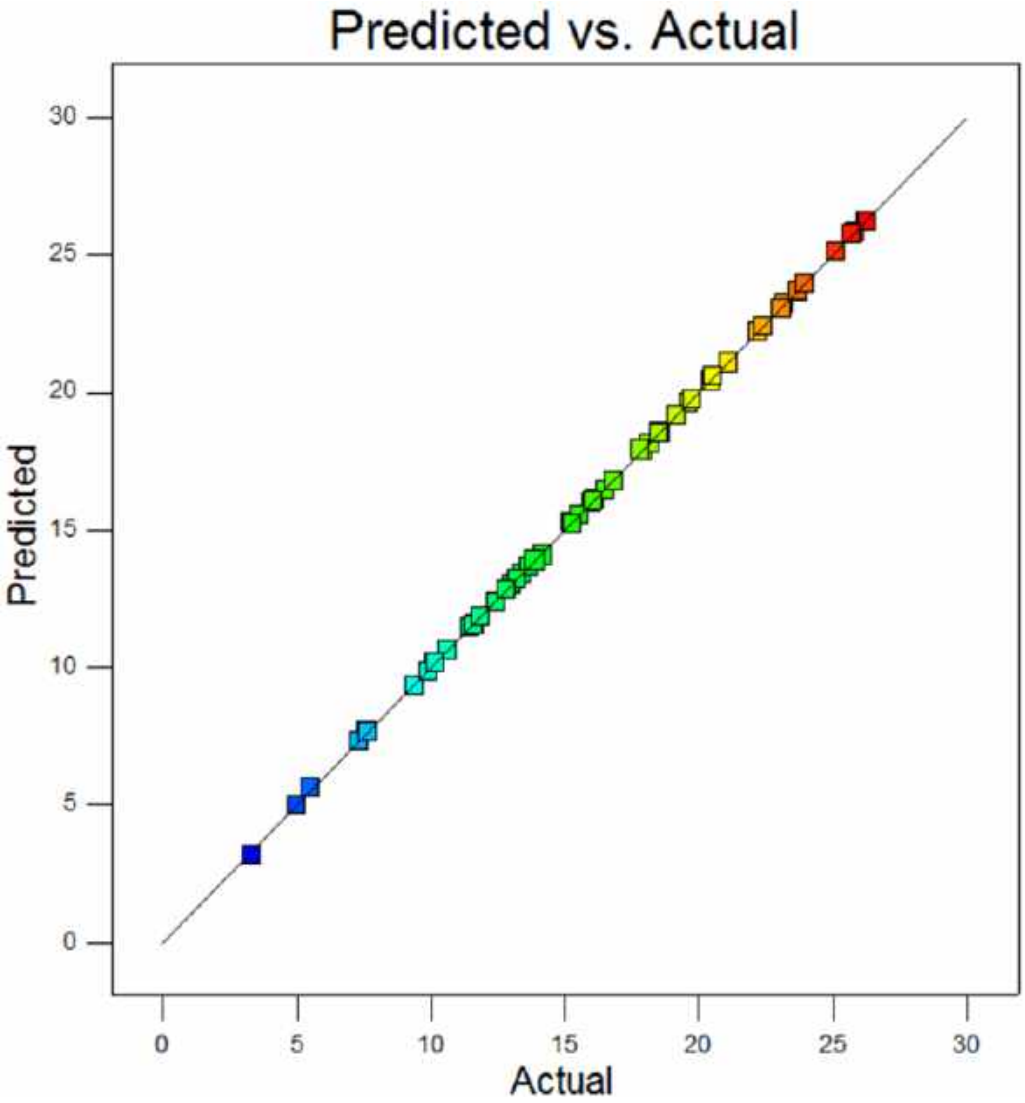


Figure 6.8: Experimental values for WFE flowrate (LPM) as a function of the values predicted by the fitted model

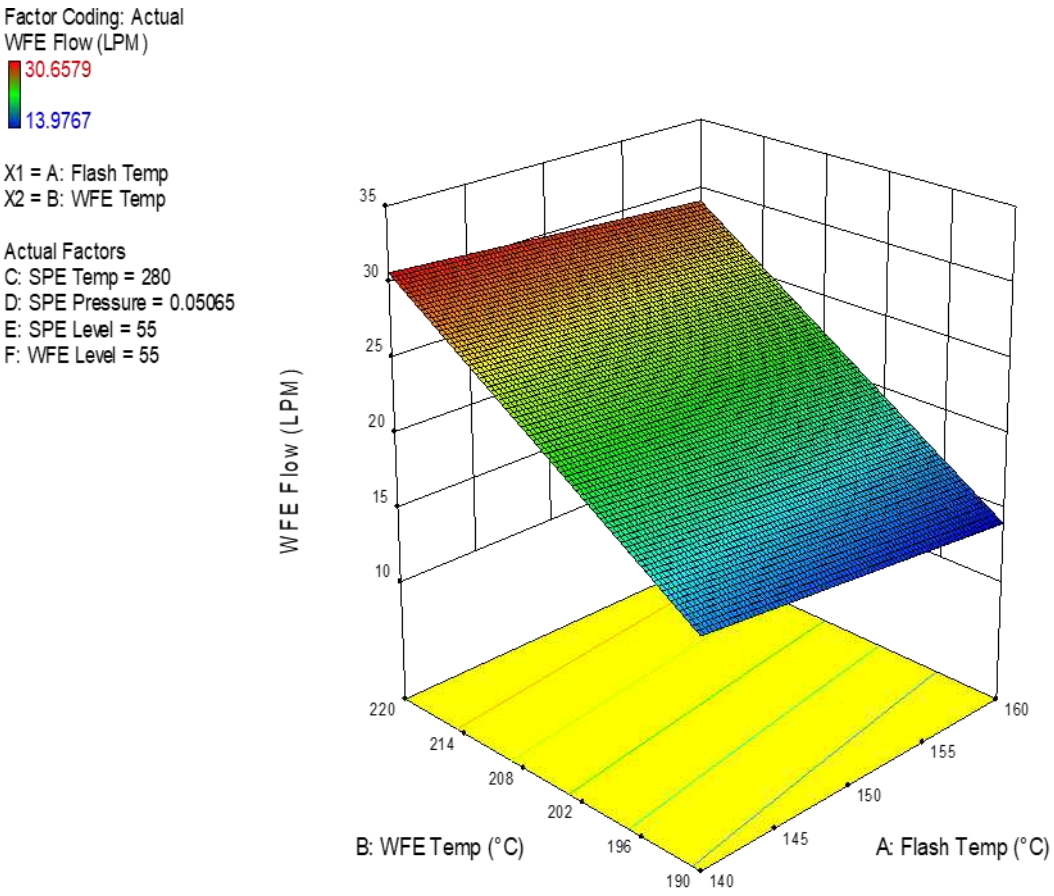


Figure 6.9: Response surface for WFE flowrate (LPM) as a function of WFE temperature (°C) and Flash tank temperature (°C)

Next, analysis of WFE product viscosity (cSt) in the assessment of experimental design ranged from 6.96 to 10.63 cSt is used and the response is fitted in the regression model as shown in Eqn. 6.16. (The detailed ANOVA analysis can be found in Appendix Table A6.18)

$$y_2 = 8.728 - 0.206A + 1.633B - 0.0214E + 0.0493AB \quad (6.16)$$

It shows that the relationship of WFE temperatures (B) is proportional to WFE product viscosity with a positive regression coefficient of 1.633. The viscosity of WFE (y_2) is found to be inversely proportional to flash temperature (A) and WFE level (E) with a negative coefficient of -0.206 and -0.0214 respectively. A higher temperature in the flash tank causes viscous oil additives within the oil is broken down, giving slightly lower WFE oil viscosity. Besides, at a significantly high level of WFE, the separation efficiency of the evaporator is affected and hence, WFE viscosity decreases. The significant interaction factor is the factor of the flash tank and WFE temperature (AB), which is the temperature interaction of consecutive

processing units. Similarly, the WFE flow response, the regression coefficients of the viscosity of the WFE product are not significantly affected by the later process unit (SPE).

In addition, Figure 6.10 shows that the regression model for WFE product viscosity is appropriate to explore the trend of this response. For validation and visualisation, a response surface is generated using the regression model and shown in Figure 6.11. From the plot, WFE product viscosity is validated to be dependent on both flash and WFE temperature with stronger dependency for the latter. As illustrated in Figure 6.11, the colour of the response surface is warmer at the axis of WFE temperature but less warm at the axis of the flash tank temperature.

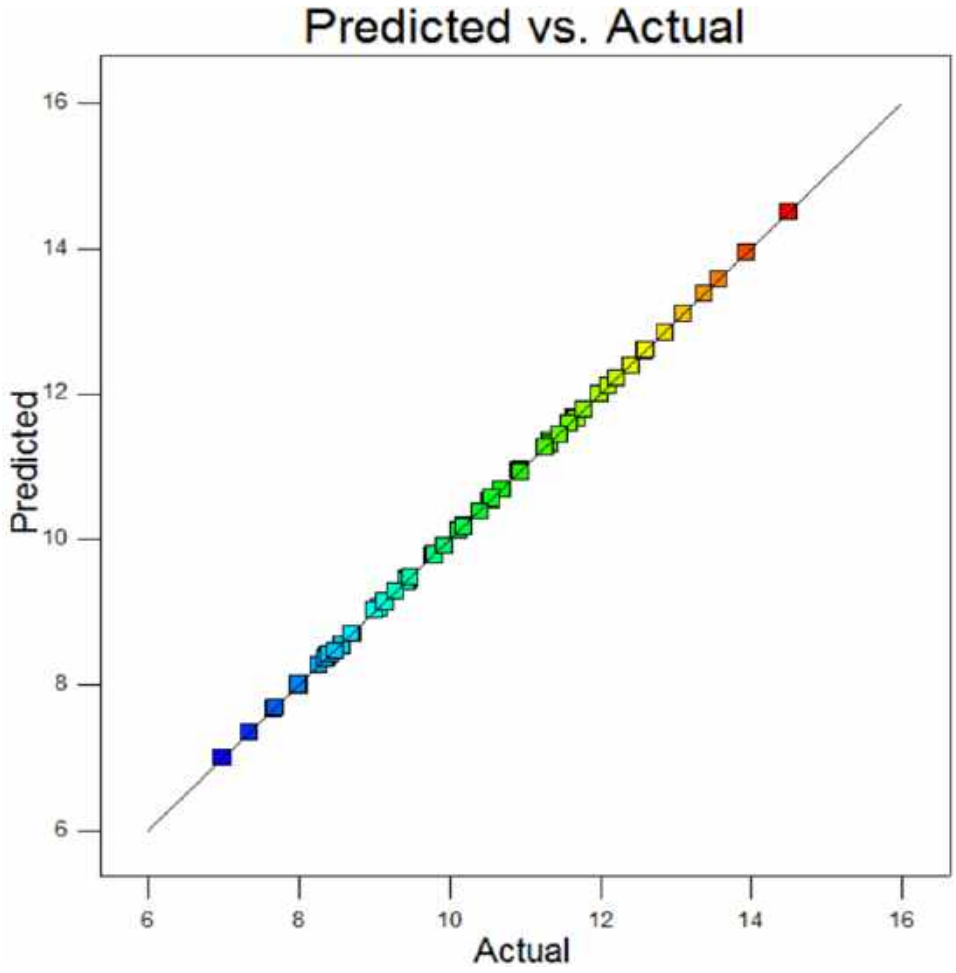


Figure 6.10: Experimental values for WFE product viscosity (cSt) against of the values predicted

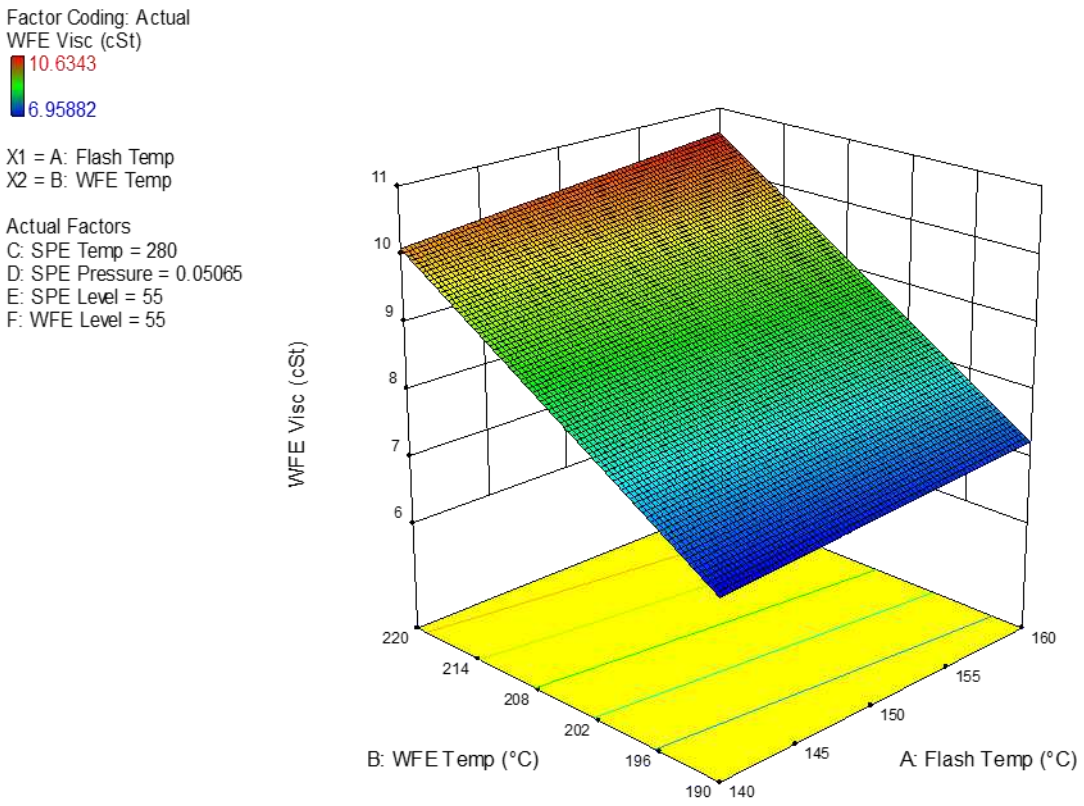


Figure 6.11: Response surface for WFE product viscosity (cSt) as a function of WFE temperature (°C) and Flash tank temperature (°C)

Analysis of SPE product flow (LPM) in the assessment of experimental design ranged from 0 to 30.54 LPM is used and the response was fitted in the regression model with regression coefficients as shown in Eqn. 6.17 below. (The detailed ANOVA analysis can be found in Appendix Table A6.19)

$$y_3 = 10.148 + 0.446A - 5.485B + 11.329C - 8.541D - 0.456E + 0.666F - 4.535BC + 2.095BC + 2.095BD + 4.905CD + 1.383CF - 2.095DE + 3.931DF - 4.113BCD + 5.189CDF \quad (17)$$

From Eqn. 6.17, the regression coefficients show that SPE product flow is affected by all main effects of Flash tank, WFE and SPE operating conditions. This is logical, as SPE is the final unit in the process and the product flow will depend on upstream units. In addition, Figure 6.12 shows that the regression model for SPE product flow as a function of flash, WFE, SPE operating condition is fitting to inspect the trend of this response. Thus, a response surface is generated using the regression model as shown in Figure 6.13. From Figure 6.13(a), the model

predicted that a higher SPE temperature (C) and higher WFE temperature (B) will improve the SPE product flow. Figure 6.13(b) and (c) also shows that having lower pressure in SFE improves the product flow, while the WFE level gives a slight proportional relationship with product flow. Moreover, increasing the WFE level (E) and SPE temperature (C) simultaneously give high SPE product flowrate.

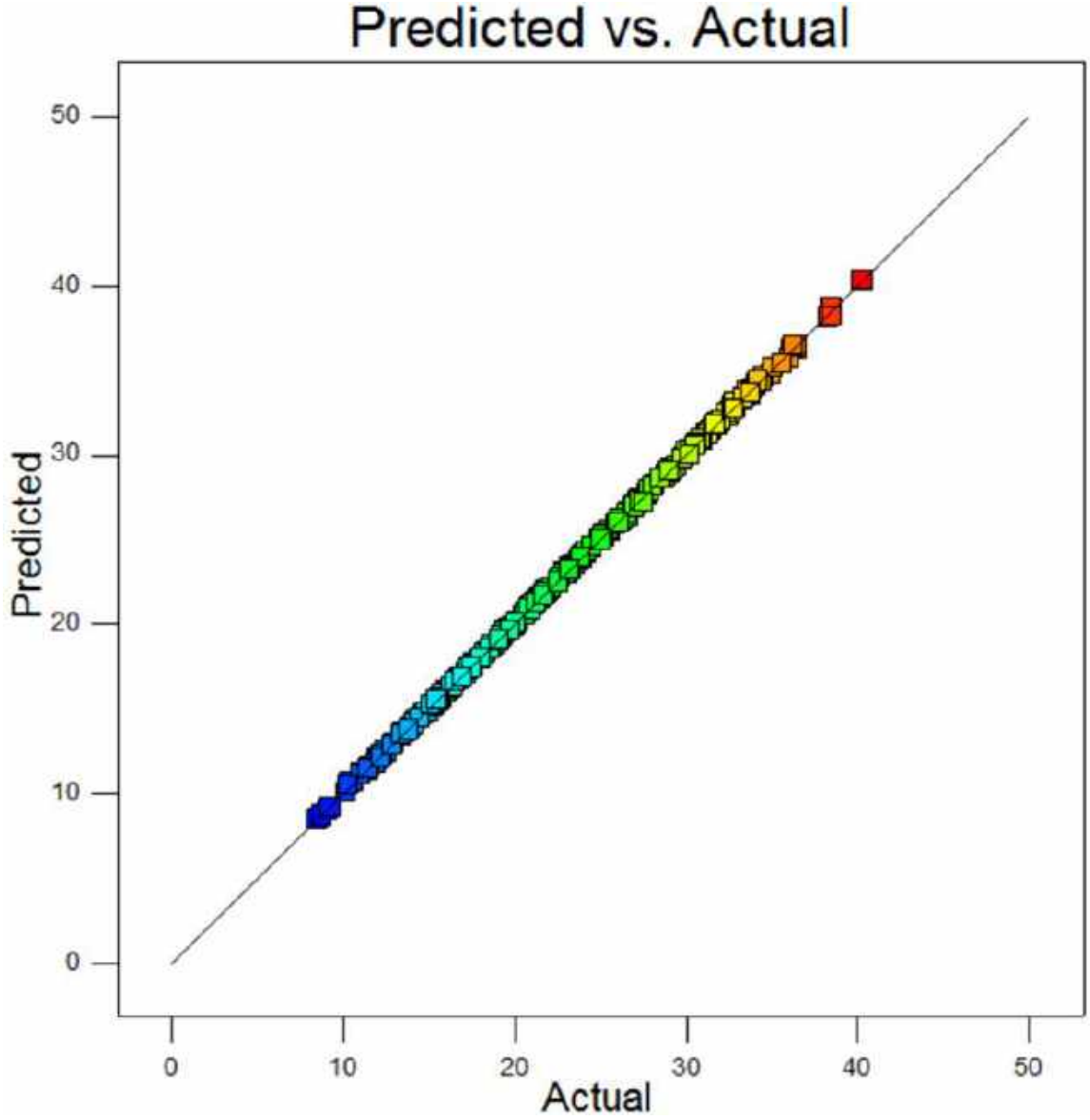
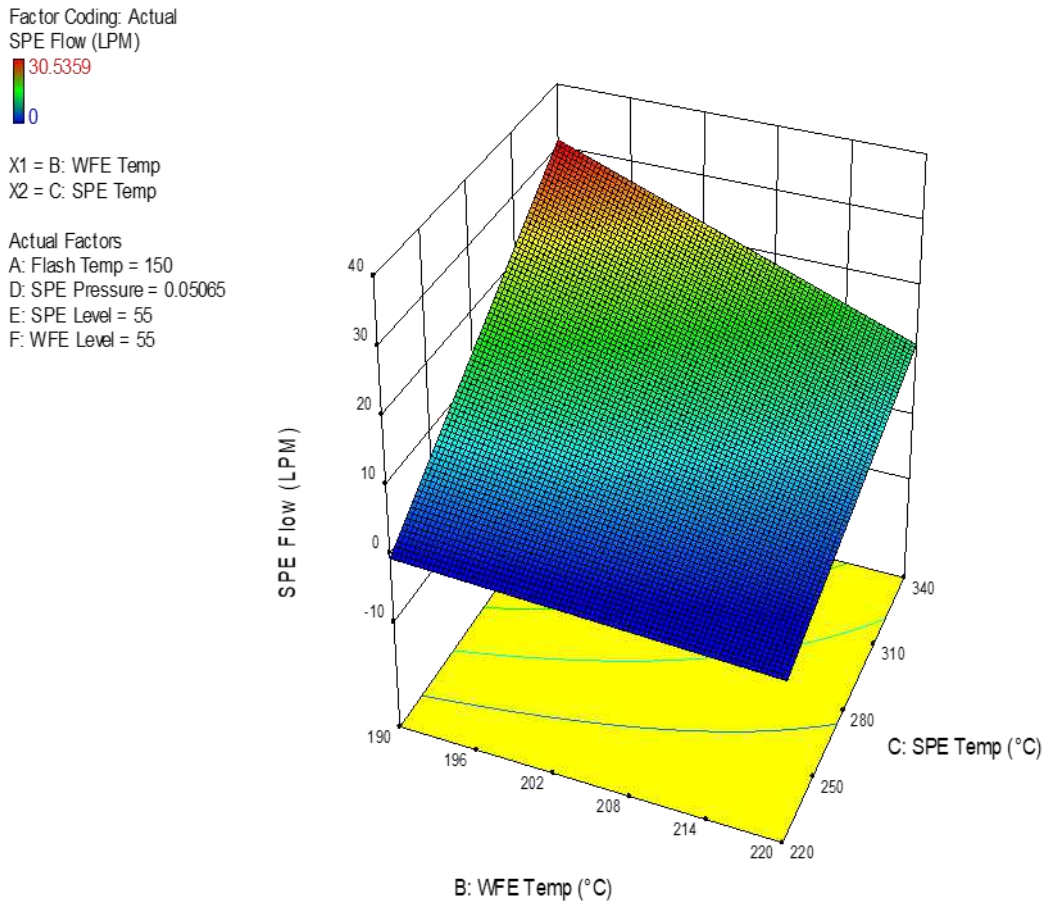
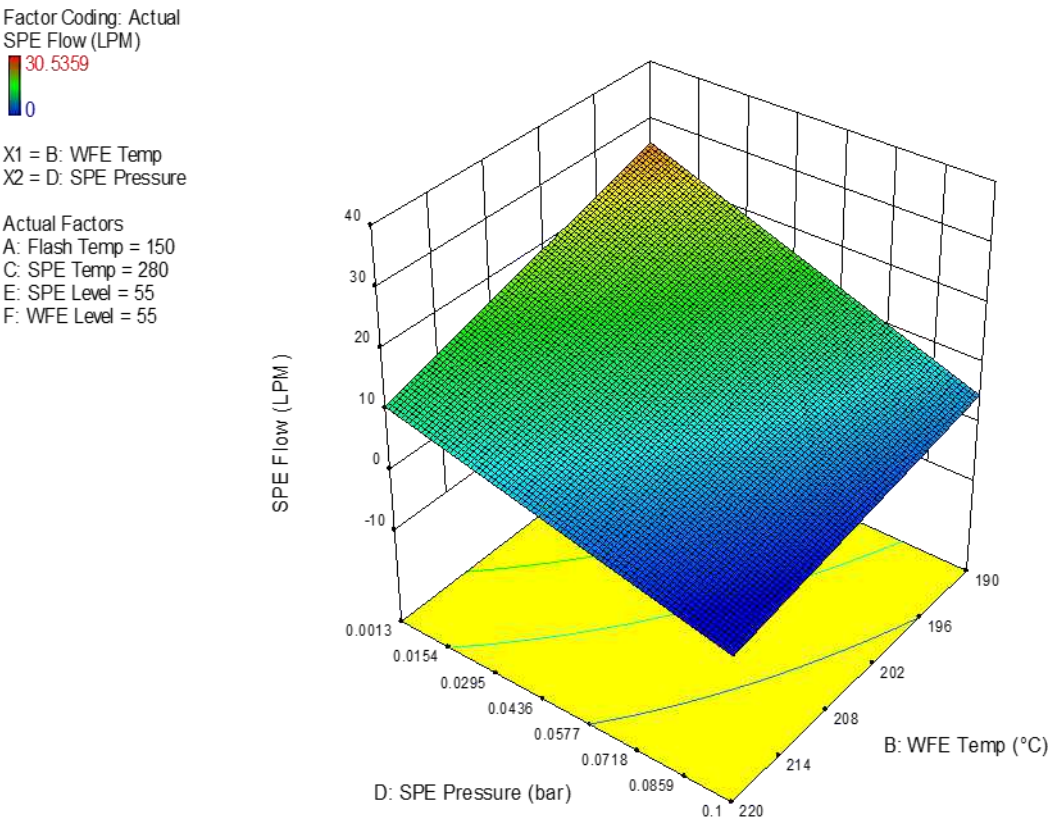


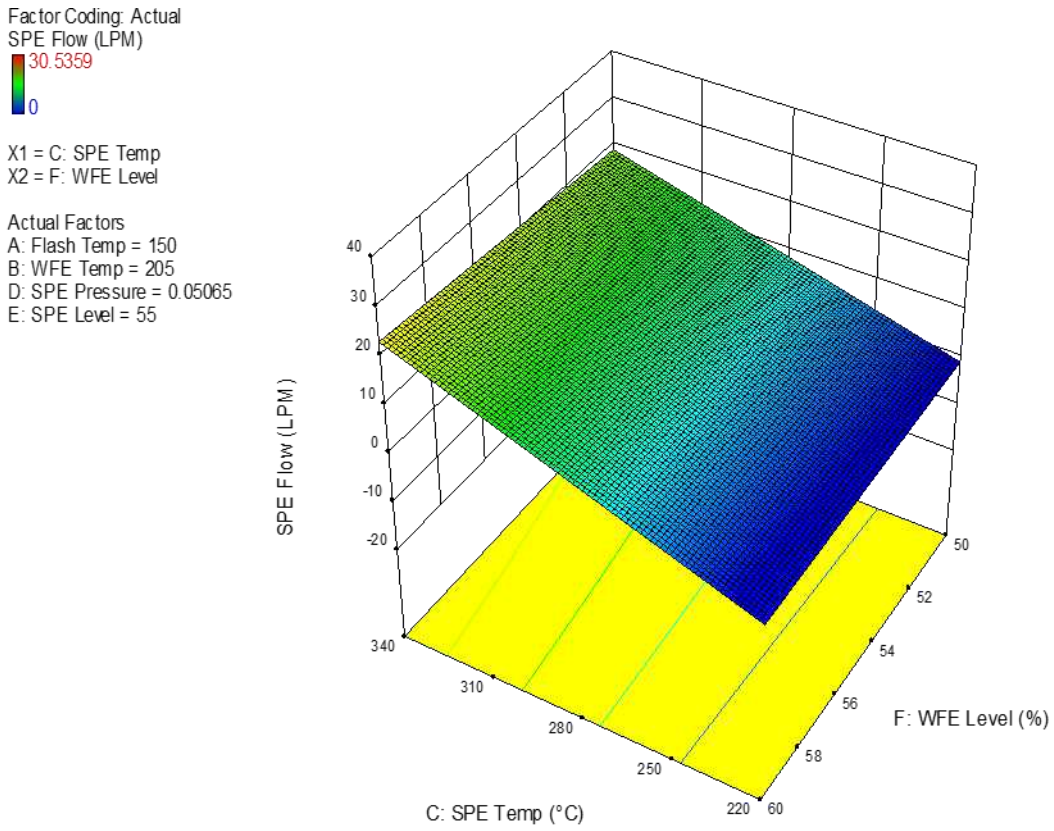
Figure 6.12: Experimental values for SPE product flow (LPM) against the values predicted


Factor Coding: Actual
SPE Flow (LPM)
30.5359
0
X1 = B: WFE Temp
X2 = C: SPE Temp
Actual Factors
A: Flash Temp = 150
D: SPE Pressure = 0.05065
E: SPE Level = 55
F: WFE Level = 55
SPE Flow (LPM)
40
30
20
10
0
-10
190
196
202
208
214
220
B: WFE Temp (°C)
220
250
280
310
340
C: SPE Temp (°C)

(a)


Factor Coding: Actual
SPE Flow (LPM)
30.5359
0
X1 = B: WFE Temp
X2 = D: SPE Pressure
Actual Factors
A: Flash Temp = 150
C: SPE Temp = 280
E: SPE Level = 55
F: WFE Level = 55
SPE Flow (LPM)
40
30
20
10
0
-10
0.0013
0.0154
0.0295
0.0436
0.0577
0.0718
0.0859
0.1
D: SPE Pressure (bar)
190
196
202
208
214
220
B: WFE Temp (°C)

(b)



(c)

Figure 6.13: Response surface for SPE product flow (LPM) as a function of:

(a) WFE temperature (°C) and SPE temperature (°C)

(b) SPE pressure (bar) and WFE temperature (°C)

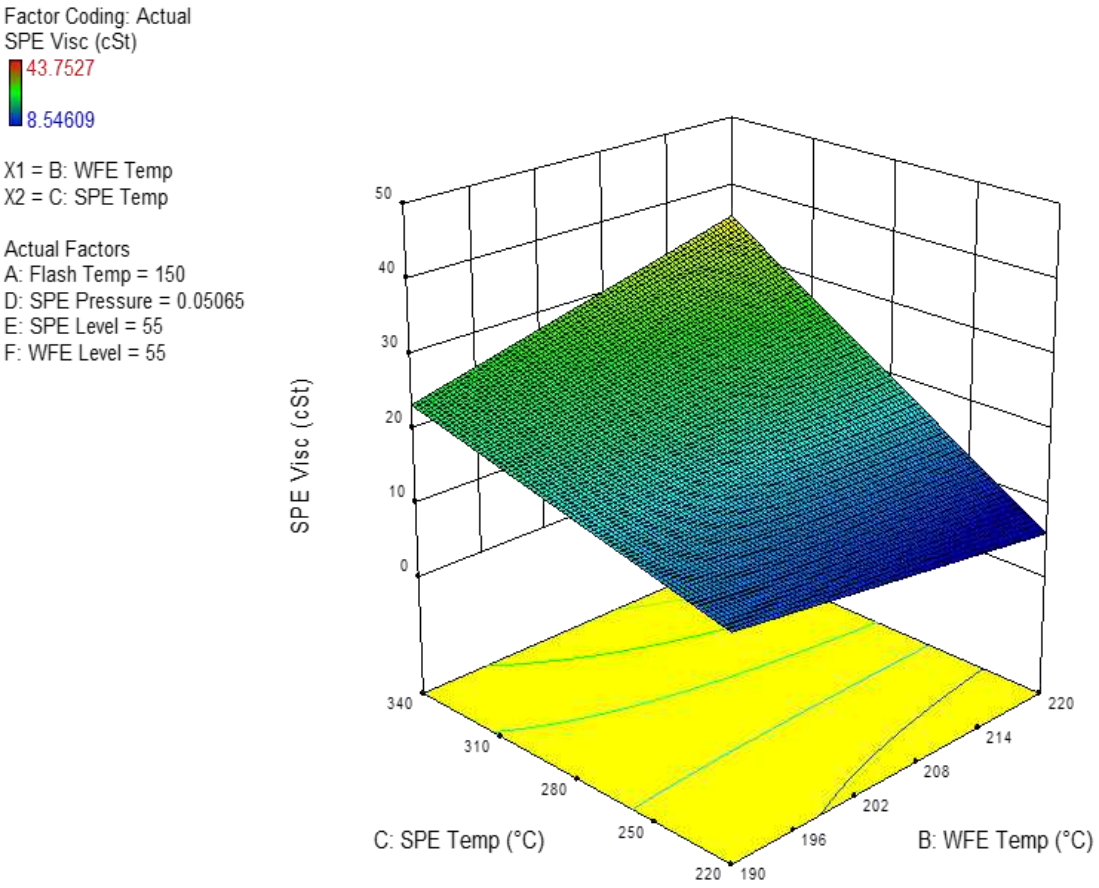
(c) SPE temperature (°C) and WFE level (%)

Analysis of SPE product viscosity (cSt) in the assessment of experimental design ranged from 8.5 to 43.8 cSt is used and the response is fitted in the regression model with regression coefficients listed in Eqn. 6.18.(The detailed ANOVA analysis can be found in Appendix Table A6.20)

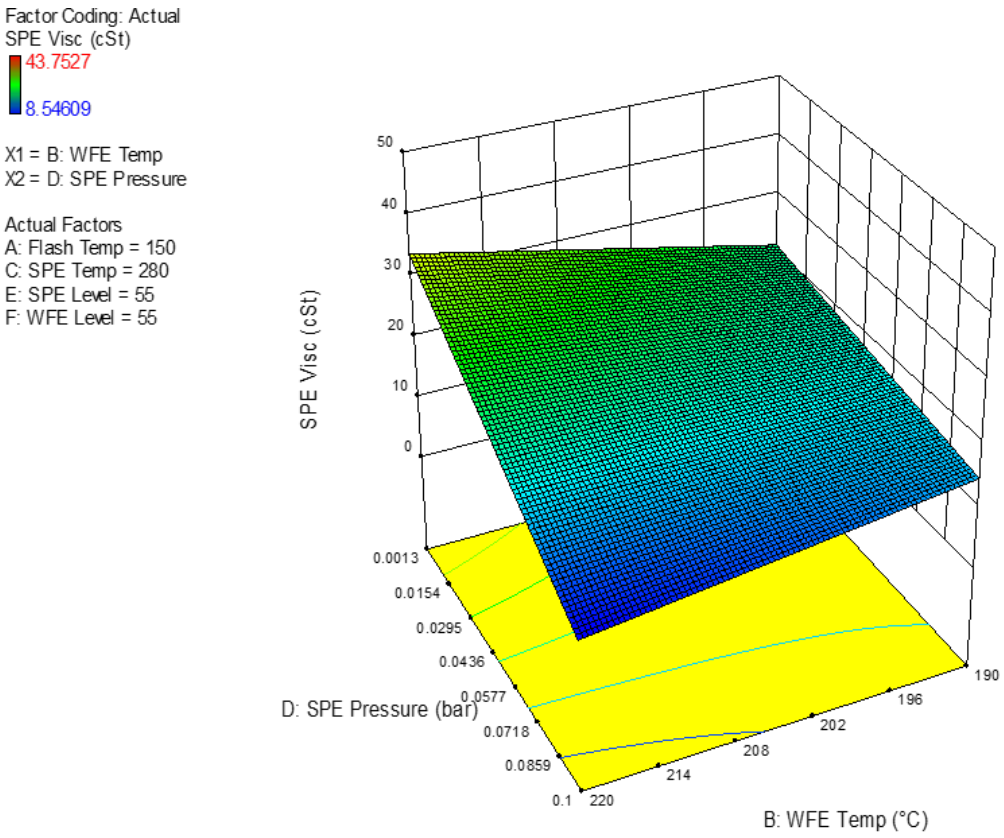
$$y_4 = 19.527 + 1.357B + 10.134C - 8.257D + 4.714BC - 4.842BD + 1.961CD + 7.806CF + 9.512CDF \quad (6.18)$$

From Eqn. 6.18, the SPE product viscosity mainly depends on SPE pressure (D) and temperature (C), as well as the WFE temperature (B). Thus, a response surface is generated using the regression model as shown in Figure 6.14. Subsequently, SPE product viscosity is found to improve with the simultaneous increase of SPE temperature and of WFE temperature

from Figure 6.14(a). A similar relation to the SPE flow is found with the SPE pressure and level from Figure 6.14(b), in which a lower SPE pressure and higher WFE temperature give better product viscosity. This shows that lowering pressure while maintaining a high temperature in both SPE and WFE can improve SPE separation quality and yield.



(a)



(b)

Figure 6.14: Response surface for SPE product viscosity (cSt) as a function of:

(a) WFE temperature (°C) and SPE temperature (°C)

(b) WFE temperature (°C) and SPE pressure (bar)

From the surface responses, the optimal operating temperatures of the flash tank, WFE and SPE are successfully established. To aid the understanding of the optimal combination of operating conditions for flash, WFE and SPE, an operating condition solution graph is plotted as illustrated in Figure 6.15. In details, Figure 6.15(a), (b), (c), (d), (e) and (f) represent factors, while Figure 6.15 (g), (h), (i) and (j) represent the responses. Moreover, the factors in Figure 6.15(a), (b) and (c) are varied between their respective minimum and maximum operating temperatures denoted by the vertical boundary based on the actual capabilities of processing equipment. Furthermore, the vertical boundaries in Figure 6.15(g) and (h) are the process limit for WFE and SPE product viscosity. Note that higher viscosity is preferred for products of WFE and SPE.

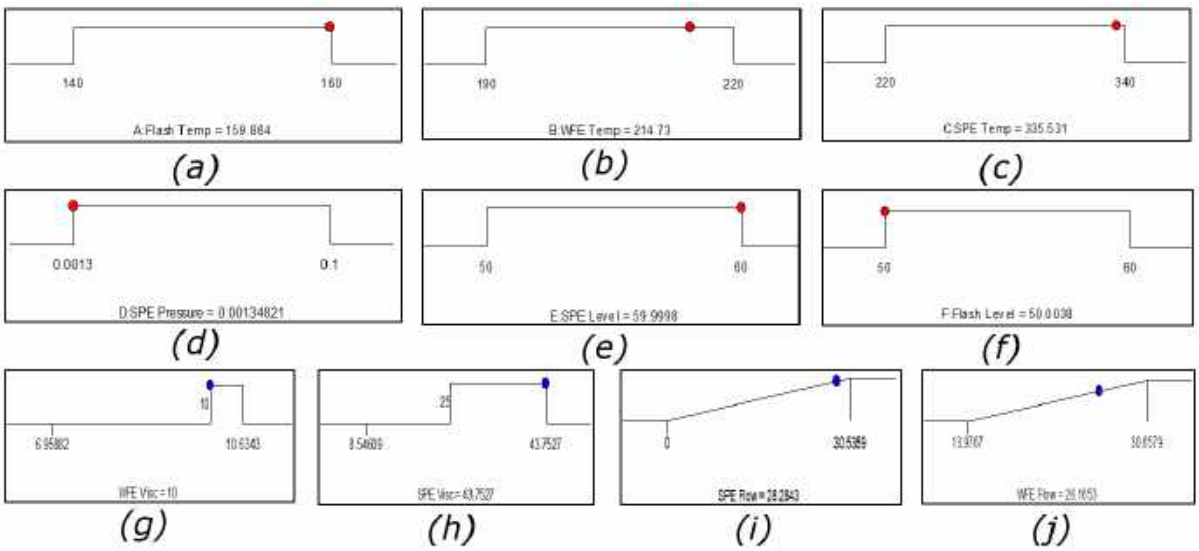


Figure 6.15: Optimisation solution diagram

(a) Operating temperature of flash tank
(b) Operating temperature of WFE
(c) Operating temperature of SPE
(d) Operating pressure of SPE
(e) Operating level of SPE
(f) Operating level of WFE
(g) WFE product flow (LPM)
(h) WFE product viscosity (cSt)
(i) SPE product flow (LPM)
(j) SPE product viscosity (cSt)

Finally, comparisons between the factors and responses before and after statistical optimisation are shown in Table 6.9. Tremendous improvements are clearly seen in SPE product for its throughput (flow) at 84.4% and quality (viscosity) at 46.5%. This is achieved by lowering operating temperatures of the flash tank as much as 28.2% and 5.8% for WFE but an increase of 45.2% for SPE. The pressure in SPE remains unchanged while the level in the WFE and SPE increased by 9.1% and 20% respectively (two-tailed t-test can be found in Appendix Table A6.22).

Table 6.9: Comparisons between the factors and responses prior to and after optimisation

Process parameters	**Units**	**Before**	**After**	**Change (%)**
Flash Tank Temperature	(°C)	222.6	159.86	28.2
WFE Temperature	(°C)	228	214.73	5.8
SPE Temperature	(°C)	231	335.5	45.2
SPE Pressure	(bar)	0.0013	0.0013	0.0
WFE Level	(%)	55	50.0	9.1

Table 6.9 (Cont.): Comparisons between the factors and responses prior to and after optimisation

SPE Level	(%)	50	60	20.0
WFE Product Flow	(LPM)	20.75	26.16	26.1
WFE Product Viscosity	(cSt)	10.56	10.0	5.3
SPE Product Flow	(LPM)	15.34	28.28	84.4
SPE Product Viscosity	(cSt)	29.86	43.75	46.5

6.6.3 Optimization Results of Different Coverage Scores

Different coverage scores from the PCA model is used for DoE optimisation. Coverage scores of 80%, 95% and 99% (100% coverage score requires 49 DoE factors and resulting in 5.6×10^{14} number of runs) are used to benchmark the optimisation results. Each design will require an increase in DoE factors which, in return increases the number of runs. The optimised results are benchmarked in Table 6.10.

Table 6.10: Comparisons between the 80%, 95% and 99% coverage score (C.S.) optimisation

Variables	**Unit**	**80%** C.S. **(A)**	**Change (%)**	**95%** C.S. **(B)**	**Change (%)**	**99%** C.S. **(C)**
DoE Factors	-	3	-	6	-	17
Minimum Number of Runs	-	8	-	64	-	262144
Coverage Variance	%	80	-	95	-	99
WFE Product Flow	LPM	20.74	26.13	26.16	10.28	28.85
WFE Product Viscosity	cSt	10.00	0.00	10.00	0.00	10.00
SPE Product Flow	LPM	23.66	19.53	28.28	2.58	29.01
SPE Product Viscosity	cSt	38.08	14.89	43.75	6.17	46.45

From the table above, Case C is taken as the comparison basis since the coverage score is nearest to including the full process system. Improving the optimisation model from a coverage score of 80% to 95% (Case A to B) gives an averaged deviation of 15.1%. However, this work recommends a higher quality of optimisation coverage, which is 95%. This is shown when the optimisation model is improved from 95% to 99% coverage, in which the averaged deviation is 4.75%. The small improvements in optimisation results are not justified for the incurring capital cost for 262080 extra experimental runs (4095 folds compared to Case B). Therefore, the 95% coverage score (Case B) used for this work is effective and efficient.

6.6.4 Process Cycle Assessment

The impact categories chosen to access environmental impacts for this case study are global warming potential (GWP) and acidification potential (AP). Standard indicators for GWP and AP which are carbon dioxide (CO_2) and sulphur dioxide (SO_2) equivalent are used respectively since it is common for oil products contaminated with hydrogen sulphide (H_2S). Thus, characterisation models for conversion of natural and electricity to CO_2 equivalent are shown in Table 6.11. Similarly, characterisation model for conversion of H_2S to SO_2 equivalent is shown in Table 6.12.

Table 6.11: Characterisation model for GWP (The Carbon Trust, 2011)

Energy	**Natural gas combustion**	**Grid Electricity**
Units	kWh	kWh
GWP ($kgCO_2^{-1}$)	5.38	0.52

Table 6.12: Characterisation model for AP (The Carbon Trust, 2011)

Acid Producer	**Hydrogen Sulphide (H_2S)**
Units	kg
AP ($kgSO_2^{-1}$)	1.88

With the appropriate characterisation models, data for utilities, energy consumption and fractions of H_2S in oil products are extracted from the simulation model and compared for both cases of before and after statistical process optimisation. Calculated results are shown in Table 6.13, Table 6.14 and compared in Table 6.15.

Table 6.13: GWP for combustion of natural gas in furnace and electricity consumption for pump

Units	Furnace			Pump		
	Before	After	% Change	Before	After	% Change
Energy (kWh)	93389.1	90139.6	3.54	763.9	757.3	86.6
GWP ($kgCO_2^{-1}$)	17146.2	16549.6		400.7	397.3	

Table 6.14: AP for by-products and main products in the process

Product	WFE			SPE		
	Before	After	% Change	Before	After	% Change
H_2S (kg)	0.0057	0.0007	156.2	0.00	0.00	-
AP ($kgSO_2^{-1}$)	0.01	0.00		0.00	0.00	
Product	**Lights**			**Oil water**		
	Before	**After**	**% Change**	**Before**	**After**	**% Change**
H_2S (kg)	62.11	4.80	171.3	1.02	0.95	7.1
AP ($kgSO2^{-1}$)	116.76	9.03		1.92	1.78	
Product	**Asphalt**					
	Before	**After**	**% Change**			
H_2S (kg)	0.00	0.00	0.00			
AP ($kgSO2^{-1}$)	0.00	0.00	-			

Table 6.15: Comparison of environmental performance for before and after statistical process optimisation

Impact category	GWP			AP		
	Before	After	% Change	Before	After	% Change
Value	17546.9	16946.9	3.42	118.7	10.8	90.89

Ultimately, statistical process optimisation has helped to achieve a leaner process by reducing the consumption of energy while improving product throughput and quality (Table 6.9). In addition, benefits extended to the environmental performances as GWP and AP assessed have witnessed immense improvement as shown in Table 6.15. Although it may be argued that GWP improvement is merely 3.42% but it is remarkable considering that no additional investment cost is required. Moreover, operating cost decreased due to reductions in energy and utility consumption. Furthermore, AP improved 90.89% which means that product or by-product sold to consumers will pose a significantly lesser environmental threat. To conclude, the process achieved lean and green after statistical process optimisation.

6.7 Conclusion

In this work, a novel framework (PASPO) which integrates the use of process simulation, PCA, DoE and process cycle assessment is developed. The PASPO framework aims to provide a cost-effective and systematic way for process optimisation in the real chemical plant with the consideration of environmental impacts. Data from plant data historian are normally very big data sets that are difficult to be analysed. The significance of the PASPO framework is that it transforms plant operational data as the driving force for a practical and sensible plant-wide optimisation study. To elucidate the applicability and effectiveness of the PASPO framework, a real industrial case study of a waste oil re-refining plant (Pentas Flora Sdn. Bhd.) is carried out. This framework utilizes a novel correlational-based PCA implementation, in which processing variables are dimensionally reduced by both principal component's cumulative variance and variable contribution score. After applying the PASPO framework, optimised operating conditions gave an increase in SPE product (main product) yield of 84.4% and WFE product improved by 26.1%. The quality of SPE also increased by 46.5%. In addition, results from DoE were assessed from the ecological point of view with process cycle assessment. The optimised processing conditions had improved Global Warming Potential (GWP) by 3.42% and Acidification Potential (AP) by 90.89%. With the aid of this framework, the waste oil re-

refining plant can transform towards a leaner and greener operation. Evidently, emissions are reduced simultaneously with the enhancement of product throughput and quality. In all, the novel PASPO framework proposed has high potential in guiding engineers to design a sustainable process that considers the allowance for future expansions.

The main limitation in implementing the PASPO framework within a process plant is it requires a considerable amount of historical data on the whole process system to be used as input to the framework. This can be a major challenge for the antiquated processing systems or plants, as sampling instruments may be not as readily available. Furthermore, this work only covers the improvement of process systems from an operational perspective. The approach only targets improvements by altering the operating conditions of processing systems. The PASPO framework can be extended to include the debottlenecking studies of process systems by modifying equipment capacities.

Nomenclature

Abbreviation	**Definition**
WFE	Wiped Film Evaporator
SPE	Short Path Evaporator
DoE	Design of Experiment
GWP	Global Warming Potential
PCA	Principal Component Analysis
H2S	Hydrogen Sulphite
CO2	Carbon Dioxide
SO2	Sulphur Dioxide
LCA	Life Cycle Analysis
LPM	Litres Per Minute
GCMS	Gas Chromatography–Mass Spectrometry
AP	Acidification Potential
GHG	Greenhouse Gases
ISO	International Standard Organisation

Appendix

The following is the paired t-test results for simulated and actual data. The results gave a p-value of 0.1384 (the difference is not statistically significant). The 95% confidence interval is from -8.6837 to 43.2157 with a mean of 17.2660.

Table A6.16: Paired t-test parameters (t=1.8473, dt=4, standard error of difference=9.346)

Group	Simulated Data	Actual Data
Mean	554.522	537.256
SD	483.6589	470.1995
SEM	216.2988	210.2796
N	5	5

The following are the regression coefficient and the statistical analysis of the design of experiment (DoE).

Table A6.17: Regression coefficients and ANOVA for the response of WFE product flow (LPM)

Source	Coefficient	Sum of Squares	F-value	p-value (Prob>F)
Model	-	16415.1560	116358.8793	<0.0001
Intercept	22.6013	-	-	-
A-Flash Temp	-1.0031	275.1277	11701.4628	<0.0001
B-WFE Temp	7.3449	11906.2998	506387.0382	<0.0001
E-SPE Level	0.3393	1.2691	53.9769	<0.0001
F-WFE Level	0.3483	0.6430	27.3466	<0.0001
AB	0.2099	5.7274	243.5931	<0.0001
EF	0.3393	0.5066	21.5483	<0.0001
Residual	-	12.5085	-	-
Cor Total	-	16765.5081	-	-

A-Flash temperature; B-WFE Temperature; C-SPE Temperature; D-SPE Pressure; E-SPE Level; F- WFE Level

Table A6.18: Regression coefficients and ANOVA for response of WFE product viscosity (cSt)

Source	Coefficient	Sum of Squares	F-value	p-value (Prob>F)
Model	-	786.9929	165126.4812	<0.0001
Intercept	8.7278	-	-	-
A-Flash Temp	0.2055	12.4209	10424.5553	<0.0001
B-WFE Temp	1.6331	638.0008	535460.1176	<0.0001
E-SPE Level	-0.0214	0.0108	9.0462	0.0028
AB	0.0493	0.3514	294.9067	<0.0001
Residual	-	0.6363	-	-
Cor Total	-	812.9685	-	-

A-Flash temperature; B-WFE Temperature; C-SPE Temperature; D-SPE Pressure; E-SPE Level; F- WFE Level

Table A6.19: Regression coefficients and ANOVA for the response of SPE product flow (LPM)

Source	Coefficient	Sum of Squares	F-value	p-value (Prob>F)
Model	-	52398.7500	236.2309	<0.0001
Intercept	10.1482	-	-	-
A-Flash Temp	0.4463	60.5451	3.8214	0.0300
B-WFE Temp	-5.4853	6667.8276	420.8508	<0.0001
C-SPE Temp	11.3287	23774.3998	1500.5601	<0.0001
D-SPE Pressure	-8.5408	11734.2456	740.6261	<0.0001
E-SPE Level	-0.4555	12.4843	5.1568	0.0492
F-WFE Level	0.6656	13.2176	7.2031	0.0452
BC	-4.5353	1975.1285	124.6635	<0.0001
BD	2.0954	367.1878	23.1757	<0.0001
CD	4.9047	1647.7490	104.0004	<0.0001
CF	1.3832	17.2610	10.4583	0.0499
DE	-2.0946	26.1982	6.6535	0.0199
DF	3.9307	53.9186	3.4032	0.0456
BCD	-4.1132	604.7040	38.1669	<0.0001
CDF	5.1890	53.5888	3.3823	0.0365
Residual	-	8302.0906	-	-
Cor Total	-	60782.8346	-	-

A-Flash temperature; B-WFE Temperature; C-SPE Temperature; D-SPE Pressure; E-SPE Level; F- WFE Level

Table A6.20: Regression coefficients for the response of SPE product viscosity (cSt) as a function of flash, WFE and SPE temperature (°C)

Factor	Coefficient	Sum of Squares	F-value	p-value (Prob>F)
Model	-	39819.2607	208.1621	<0.0001
Intercept	19.5272	-	-	-
B-WFE Temp	1.3569	472.7502	19.7711	<0.0001
C-SPE Temp	10.1338	19782.9625	827.3510	<0.0001
D-SPE Pressure	-8.2566	11825.1727	494.5452	<0.0001
BC	4.7141	2531.0360	105.8514	<0.0001
BD	-4.8424	2220.4271	92.8614	<0.0001
CD	1.9611	270.6337	11.3183	0.0008
CF	7.8064	308.4422	12.8995	0.0004
CDF	9.5120	272.2222	11.3847	0.0008
Residual	-	12672.9401	-	-
Cor Total	-	52630.4007	-	-

A-Flash temperature; B-WFE Temperature; C-SPE Temperature; D-SPE Pressure; E-SPE Level; F- WFE Level

Table A6.21: Model regression parameters

Model Properties	WFE Viscosity Model	SPE Viscosity Model	SPE Flow Model	WFE Flow Model
R^2	0.9992	0.9586	0.9632	0.9992
Adj R^2	0.9992	0.9549	0.9596	0.9992
Pred R^2	0.9992	0.9446	0.9478	0.9992

Table A6.22: Confidence interval of optimised solution (Two-tailed t-test with n=1)

Response	WFE Viscosity (cSt)	SPE Viscosity (cSt)	SPE Flowrate (LPM)	WFE Flowrate (LPM)
Predicted Mean	10.0000	43.7527	28.2843	26.1653
Predicted Median	10.0000	43.7527	28.2843	26.1653
Std Dev	0.0345	4.8899	3.9804	0.1533
n	1	1	1	1
SE Pred	0.0361	5.9735	5.9261	0.1899
95% PI low	9.9290	32.0180	16.6424	25.7923
95% PI high	10.0710	55.4874	39.9262	26.5384

Table A6.23: Confirmation of optimisation results by sampling and error analysis

Properties	WFE Viscosity (cSt)	SPE Viscosity (cSt)	SPE Flowrate (LPM)	WFE Flowrate (LPM)
Optimised Value	10.00	43.75	28.28	26.16
Sample 1	10.19	44.71	28.25	26.31
Sample 2	10.29	43.92	28.39	26.33
Sample 3	10.04	44.39	28.34	26.67
Signal	0.173333	0.59	0.046667	0.28
Pooled S.D.	0.088976	0.28098	0.050166	0.14089
Noise	0.072648	0.22942	0.040961	0.115036
T-value	2.385924	2.571708	1.139304	2.434016
P value (Type I Error)	0.075501	0.061867	0.318172	0.07167
P value (Type II Error)	<0.0001	<0.0001	0.0004	<0.0001

References

Abdi, H. and Williams, L. J., 2010. Principal Component Analysis. *John Wiley &Sons, Inc.*,2, 433-459.

Aida, S., Matsuno, T., Hasegawa, T., Tsuji, K., 2017. Application of principal component analysis for improvement of X-ray fluorescence images obtained by polycapillary-based micro-XRF technique. *Nuclear Instruments and Methods inPhysics Research B,* 402, 267-273.

Al-Sayed, A., 2015. Principal component analysis within nuclear structure. Nuclear Physics A, 933, 154-164.

Anderson, M., 2018. *Design Expert Documentation*. [online] Statease. Available at: https://www.statease.com/docs/v11/tutorials/combined-mix-process.html [Accessed 29 Oct. 2018].

Aspen Process Engineering Webinar, 2006. Aspen HYSYS Property Packages Overview and Best Practices for Optimum Simulations. Aspentech, Galaxis, Singapore.

Box, G. E. P., Behnken, D. W., 1960. Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics*, 2(4), 455-475.

Bratchell, N., 1989. Multivariate response surface modelling by principal components analysis. *Journal of Chemometrics*, 3(4), 579-588.

Ning, C. and You, F., 2018. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. *Computers and Chemical Engineering,* Volume 112, 190-210.

Collins, L., Dziak, J., Li, R., 2009. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods*, 14(3), 202-224.

Čuček, L., Klemeš, J.,Kravanja, Z., 2012. A Review of Footprint analysis tools for monitoring impacts on sustainability. *Journal of Cleaner Production*, 34, 9-20.

Chiang, L., Lu, B., Castillo, I., 2017. Big data analytics in chemical engineering. *Annual review of chemical and biomolecular engineering*, *8*, 63-85.

Derringer, G. and Suich, R., 1980. Simultaneous Optimization of Several Response Variables.*Journal of Quality Technology*, 12(4), 214-219.

Drovandi, C.C., Holmes, C., McGree, J.M., Mengersen, K., Richardson, S., Ryan, E.G., 2017. Principles of experimental design for Big Data analysis. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *32*(3), p.385.

Fisher, R. A., 1935. *The Design of Experiments.* Edinburg: Oliver & Boyd.

Gunst, R. F. and Mason, R. L., 2009. Fractional factorial design. *WIREs CompStat,*1, 234-244.

Halvorsen, M., Elseth, G.,Naevdal, O., 2012. Increased oil production at Troll by autonomous inflow control with RCP valves. SPE Annual Technical Conference and Exhibition.

Hostelling, H., 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441.

How, B.S. and Lam, H.L., 2018a. Sustainability evaluation for biomass supply chain synthesis: Novel principal component analysis (PCA) aided optimisation approach. Journal of Cleaner Production, 189, 941-961.

How, B.S. and Lam, H.L. 2018b. PCA Method for Debottlenecking of Sustainability Performance in Integrated Biomass Supply Chain. (In Press) Process Integration and Optimization for Sustainability, doi.org/10.1007/s41660-018-0036-3.

Hu, Y., Li, G., Chen, H., Li, H., Liu, J., 2016. Sensitivity analysis for PCA-based chiller sensor fault detection. International Journal of Refrigeration, 63, 133-143.

International Organization for Standardization (ISO), 2006. *ISO 14044 :Environmental management - Life cycle assessment - Requirements and guidelines.* 1st ed. Geneva: ISO.

Jackson, D., 1993. Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches.*Ecology*, 74(8), pp.2204-2214.

Jolliffe, I. and Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.

Joly, M., Moro, L., Pinto, J., 2002. Planning and scheduling for petroleum refineries using mathematical programming. Brazilian Journal of Chemical Engineering, 19(2), 207-228.

Klemeš, J. and Kravanja, Z., 2013. Forty years of Heat Integration: Pinch Analysis (PA) and Mathematical Programming (MP). *Current Opinion in Chemical Engineering*, 2(4), 461-474.

Kraber, S., 2009. *Response Surface Optimization*. Stat-Ease Inc.

Krishna Madhavi, S., Sreeramulu, D., Venkatesh, M., 2017. Evaluation of Optimum Turning Process of Process Parameters Using DOE and PCA Taguchi Method. *Materials Today: Proceedings*, 4(2), 1937-1946.

McKay, B., Willis, M., Barton, G., 1997. Steady-state modelling of chemical process systems using genetic programming. *Computers & Chemical Engineering*, 21(9), 981-996.

Morita, T., Yogo, S., Koike, M., Hamaguchi, T., Jung, S., Koshijima, I., Hashimoto, Y., 2013. Detection of cyber-attacks with zone dividing and PCA. Procedia Computer Science, 22, 727-736.

Murray, P. and Forfar, L., 2017. The Application of Advanced Design of Experiments for the Efficient Development of Chemical Processes. *Chemical Informatics*, 03(02).

Nobi, A., Alam, S., Lee, J.W., 2017. Dynamic of consumer groups and response of commodity markets by principal component analysis. Physica A: Statistical Mechanics and its Applications, 482, 337-344.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine,2,559-572. http://pbil.univ-lyon1.fr/R/pearson1901.pdf

Rea, A. and Rea, W., 2016. How Many Components should be Retained from a Multivariate Time Series PCA? *arXiv preprint arXiv:1610.03588*.

Schneider, D. F., 1997. *Debottlenecking Options and Optimization,* Houston: Stratus Engineering, Inc.

Shlens, J., 2014. A Tutorial on Principal Component Analysis. *eprintarXiv:1404.1100.*

Smith, R. and Linnhoff, B., 1988. Design of separators in the context of overall processes. *Chemical Engineering Research and Design,* 66(3), 195-228.

Telford, J. K., 2007. A brief introduction to design of experiments. *JohnsHopkins APL Technical Digest (Applied Physics Laboratory),* 27(3), 224-232.

The Carbon Trust, 2011. *Energy and Carbon conversion factors,* London: The Carbon Trust.

Toyota, H., Asai, T. and Oku, N., 2017. Process optimization by use of design of experiments: Application for liposomalization of FK506. *European Journal of pharmaceutical Sciences,* 102, 196-202.

WBCSD Chemicals, 2013. *Life Cycle Metrics for Chemical Products: A guideline by the chemical sector to assess and report on the environmental footprint of products, based on life cycle assessment,* Washington: WBCSD Chemicals.

Zhang, Y. and Edgar, T., 2008. PCA Combined Model-Based Design of Experiments (DOE) Criteria for Differential and Algebraic System Parameter Estimation. *Industrial & Engineering Chemistry Research,* 47(20), 7772-7783.

CHAPTER 7 ON DEEP DIMENSION REDUCTION FOR MULTI-UNIT PROCESS OPTIMIZATION

This work has been peer-reviewed and published in Chemical Engineering Transaction.
Teng S.Y., Máša V., Stehlík P., Lam H.L., 2019, Deep Learning Approach for Industrial Process Improvement, Chemical Engineering Transactions, 76, 487-492.

Abstract

The full operation of an industrial processing facility with artificial intelligence has been the holy grail of industry 4.0. One of the inherent difficulties is the enumerate and complex nature of processing information within an industrial plant. Hence, such data should be processed efficiently. This chapter demonstrates the effectiveness of a deep auto-encoder neural network for the dimensionality reduction of industrial processing data. The deep auto-encoder neural network functions to intake all possible processing data from the process system by sending it into an encoder neural network. Subsequently, the encoder condenses the data into highly compressed encoded variables. The network is trained in an unsupervised manner, where a decoder neural network simultaneously attempts to revert the encoded variables to their original form. Such deep learning approach allows data to be highly compressed into lower dimensions. The coded variables retain critical information of the process system, allowing reconstruction of the full process data. Auto-encoder neural networks are also able to provide noise removal for encoded data. For application, the encoded variable is demonstrated to be an effective dimension-reduced variable that can be used for plant-wide optimisation. This chapter also discusses about the further applications of auto-encoded variables for industrial equipment design improvements using industrial internet of things (IIoT).

7.1 Introduction

Deep learning is a specialized strain of machine learning which utilizes computation models with multiple processing layers to learn representations of data with multiple levels of abstraction (LeCun et al., 2015). The term “deep learning” was first introduced by Dechter (1986) for the applications of learning while searching. One of the core-enabling technology of deep learning is the invention of deep neural networks. Neural networks originate from the fields of neuroscience where McCulloch and Pitts (1943) first formalized the mathematical model of the biological neural activity inside the brain. The first model of this neural graph consists of three main components, which are the dendrites, soma, and axon. This simple model

of the neuron is later referred to as the McCulloch-Pitts neuron and has opened the doors to the popular deep learning as it is today. Although behaviours of nervous activity can be explained using McCulloch and Pitts' approach, there was one key dynamic factor that was missing, being the simulation of neural learning. This piece of irreplaceable technology for the development of deep learning is later implemented using the theory of backpropagation. Backpropagation is based on the method of automatic differentiation which was first published by Seppo (1970) in his master's thesis. The coupling of backpropagation method with the gradient descent method (Cauchy,1847) enabled useful real-world applications of these algorithms. Later, Rumelhart et al. (1988) popularized the use of the backpropagation method by applying it to learning representations in a self-organizing neural network. Today, deep learning has demonstrated remarkable success in applications of computer vision, medical diagnosis, phonetic identification, voice recognition, feature coding, natural language processing, robotics, computer games, molecular and drug analysis (Deng and Yu, 2014).

The pivotal factor that differentiates deep learning as a subfield of machine learning is the action of learning intrinsic representations in a hierarchical manner (Bengio et al., 2013). Some research works such as Mhaskar and Poggio (2016) classifies single hidden layer neural networks as shallow networks while neural network with multiple hidden layers are classified as deep neural network. Even for practical applications, deep learning has substantially outperformed other good old-fashioned Artificial Intelligence (GOFAI) in the aspects of expressibility, efficiency and learnability (Lin et al., 2017). Although there occasionally implementations of shallow learning that can attain performance close to deep learning (Ba and Caruana, 2014), the state-of-art record holder for application performances are always either deep, or very deep neural networks (Szegedy et al., 2017).

The first step to transfer deep learning techniques to process system engineering (PSE) is not to start from the commonly used supervised learning. Statistical methods dominate the field of predictive analysis in PSE, as shown in the works of Boukouvala et al. (2010) where surface response methodology was used to predict missing and noisy data. Although deep learning methodologies often outperforms statistical methods (Máša et al., 2018), these statistical methods have already proven their robustness and firmed their roots in the grounds of PSE predictive analysis. Dimension reduction using deep learning (Hinton and Salakhutdinov, 2006) is one of the most promising and implementable directions for PSE. Traditionally, dimension reduction techniques are already applied in PSE, such as work of Li and Wang (2002), which

utilized independent component analysis (ICA) for dimension reduction of dynamic trends. Principal component analysis (PCA) was utilized by How and Lam (2018) to debottleneck integrated biomass supply chain. With the requirements of a leaner and greener manufacturing process (Leong et al., 2019), there is a need for more advanced dimension reduction techniques (Lam et al., 2011). Máša et al. (2016) also shown that modern data-driven and neural network approaches can effectively optimize real-world energy systems, which further validates the application of deep learning in industrial processes.

Main problems of implementing deep and reinforcement learning in the real world are: (i) Rewards in a real-world setting are non-direct, sparse and sometimes imperfect (ii) Models are often inaccurate or non-robust for the use of model-based method (iii) Other technical challenges such as model seedings and network architecture (Henderson et al., 2018). Despite these challenges, three main topics within PSE are the most potential areas for the implementation of deep learning techniques, which are process integration, process optimization and process intensification. By insightful speculation, the later topic is the more difficult area to implement techniques of deep learning.

The contribution of this chapter is to extend the ideologies and methods of deep feature learning towards PSE. It is believed that by combining deep learning algorithms with traditional process engineering methodologies, the data-driven analysis will achieve new heights in the processing industries.

7.2 Methods and Theory

The overall framework of this chapter is shown in Figure 7.1. The framework is implemented in a continuous process improvement cycle in order to consistently optimize and debottleneck the process system of interest.

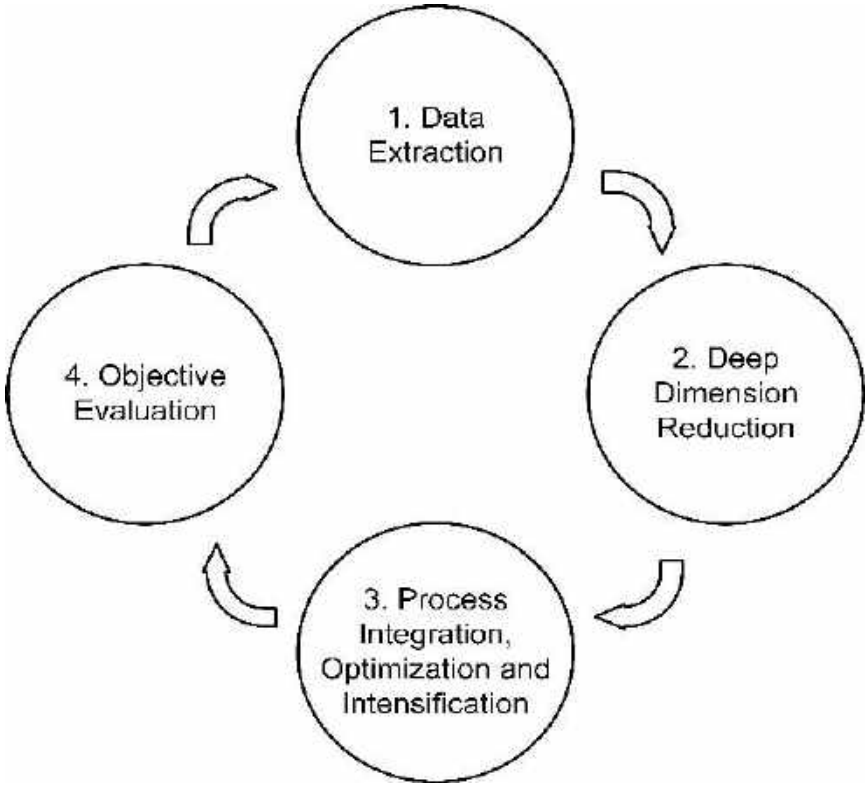


Figure 7.1: Overall framework of Deep process improvement

7.2.1 Learnable Data Extraction in the Processing Industry

The availability of learnable data varies for each industrial facility at different phase (see Figure 7.2). From experience of industrial projects, the availability of useful data is maximal during initial data extraction. The most expensive phase for data collection is during the calibration and testing phase as bad implementation can directly damage processing equipment. This highlights the criticality of front-end loading by doing data analysis using deep learning before calibration and testing.

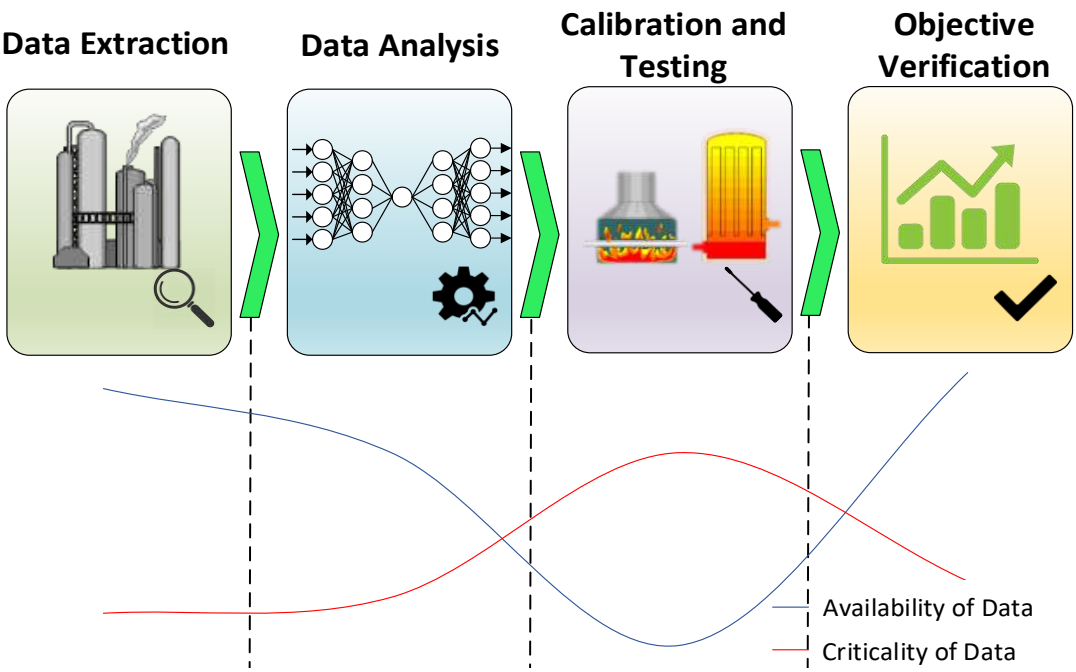
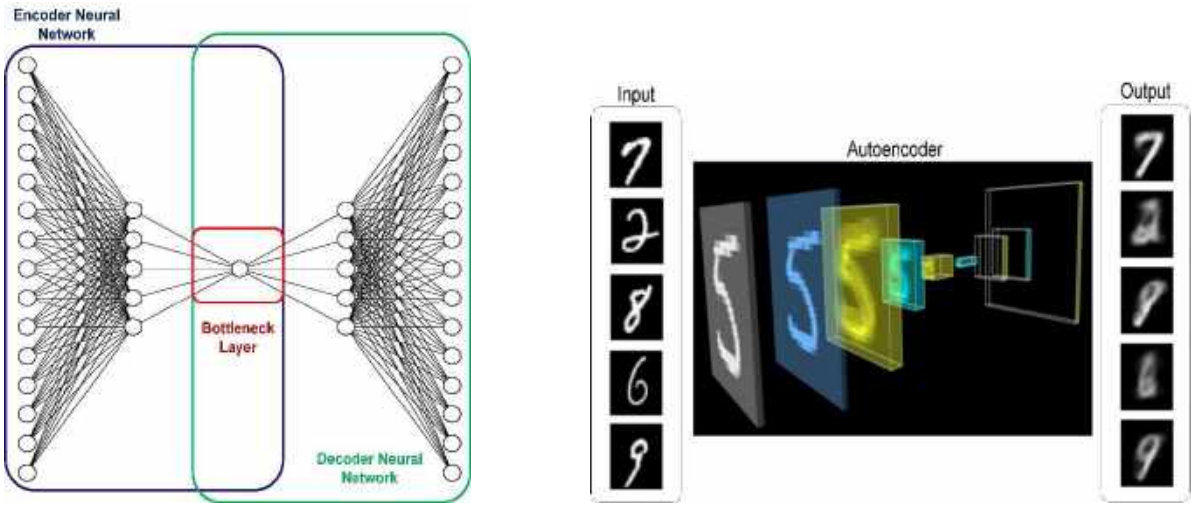


Figure 7.2: Availability and criticality of data at different phases of industrial process improvement projects

7.2.2 Deep Dimension Reduction

This work utilizes a deep *autoencoder* to process large and sparse data collected from industrial facilities. An *autoencoder* is a special type of neural network that is use for feature learning in an unsupervised manner. Information from a high dimensional input can be hierarchically encoded into network layers of lesser neurons (See Figure 7.3(a)). The layer with the least number of neurons is called the bottleneck layer, in which information is forced through a small-dimensional latent space. To ensure that the latent variables retains a large representation of the data, a decoder neural network attempts to map the latent variables to the original input.



(a) (b)

Figure 7.3: (a) Structure of an *autoencoder* (b) *Autoencoder* retaining information under dimension reduction (60k MNIST dataset)

A visual example to visualize the mechanisms of an *autoencoder* can be observed in Figure 7(b). A convolutional *autoencoder* is trained on the MNIST handwritten digit benchmark dataset. Handwritten digits from the input can be compressed into a small latent space, and then be dimensionally expanded to retain critical information of the input. Although the reconstruction of information is not perfect, the *autoencoder* learns critical features of the full data set and generalizes the learning output.

In this work, an evolutionary *autoencoder* (also known as EvoAE) is implemented using Elitist Strategy Genetic Algorithm (ESGA) with Adaptive learning rate optimization algorithm (Adam) to reduce the dimension of the processing system. The uniqueness of evolutionary *autoencoders* are such that both the activation functions and neural architecture are also hyper-optimized in the learning process. All computation work was implemented in Python and C/C++ language using custom codes and libraries such as *sklearn* and *Tensorflow*. Prior to dimension reduction, all data were normalized using a Min-Max normalization routine to standardize data between 0 and 1. Mean absolute error (%) was chosen as the benchmark criteria for dimension reduction accuracy because it gives a proportionate indication to the variability of data in the processing system.

7.3 Case Study

For the purpose of demonstration, a case study which consist of an area of the oil refinery plant is simulated. The process consists of oil being heated and then flashed in a flash tower with recycle loop. The vapor top product is cooled into light oil and collected, while bottom oil product is stored for further processing (See Figure 7.4).

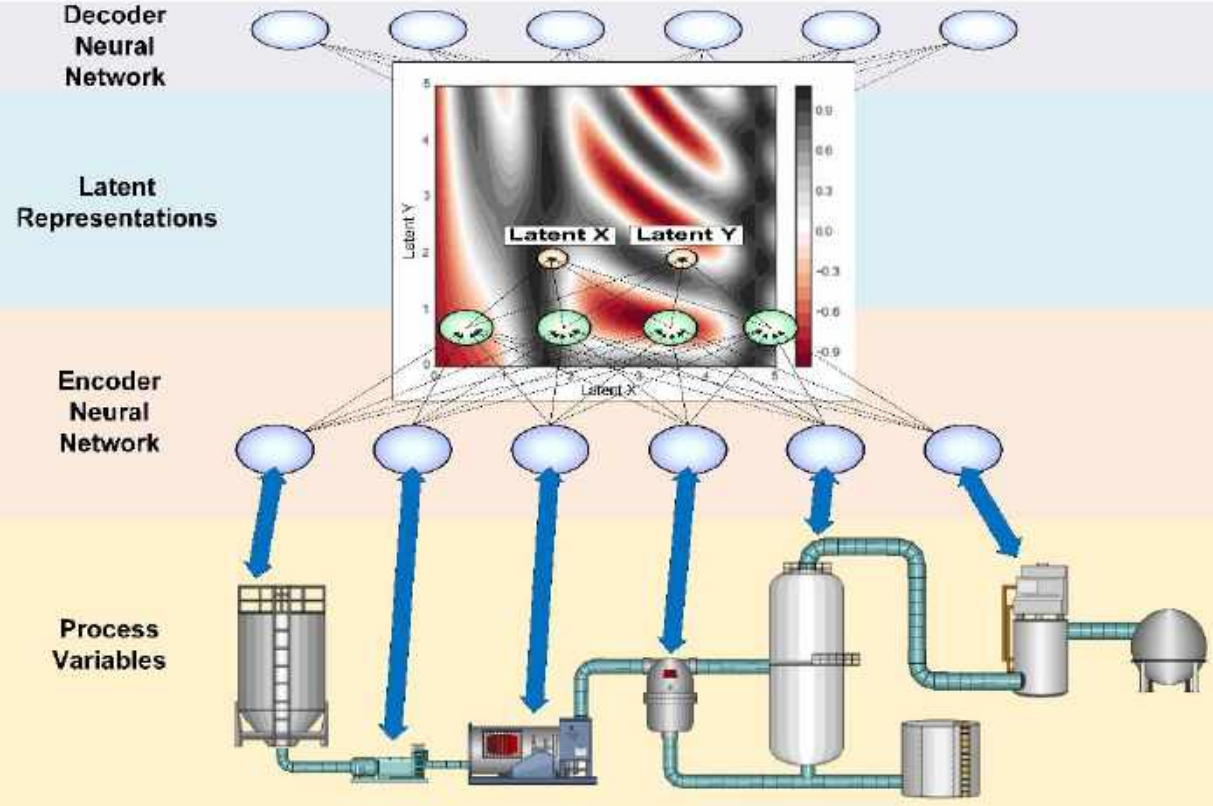


Figure 7.4: Case study and usage of *autoencoder* for process improvement

Monte Carlo simulation was implemented on the simulation model with Gaussian noises to stochastically generate process variables (See Eqn. 7.1 and 7.2). A total of 13,000 data points was simulated using this approach. 10,000 data points were allocated for training, while the remaining 3,000 data points were used for validation. 15 dimensions of process variables were recorded, which includes temperature, pressure, flowrate, density, and viscosity of processing fluid at different parts of the process.

$$y(s) = m(s) + \varepsilon \tag{7.1}$$

$$\varepsilon \sim N(0, \sigma^2) \tag{7.2}$$

Where y(s) is the stochastic process variables at a discrete state "s", m is the deterministic ground truth at discrete state "s", and ε is Gaussian noise which follows a normal distribution at 99% confidence interval.

7.4 Results and Discussion

By using the processing information dataset that contains noise and stochasticity, an *autoencoder* was trained using the neuro-evolution approach allowing variability of neural architectures. A latent dimension of 2 was specified to enable for effective interpretability of reduced variables. Neuro-evolution has shown no significant improvements of performance at 100^{th} generation with a structure of [15, 25, 10, 2, 2, 13] and training was stopped. The detailed architecture of the neural network is shown in Figure 5.

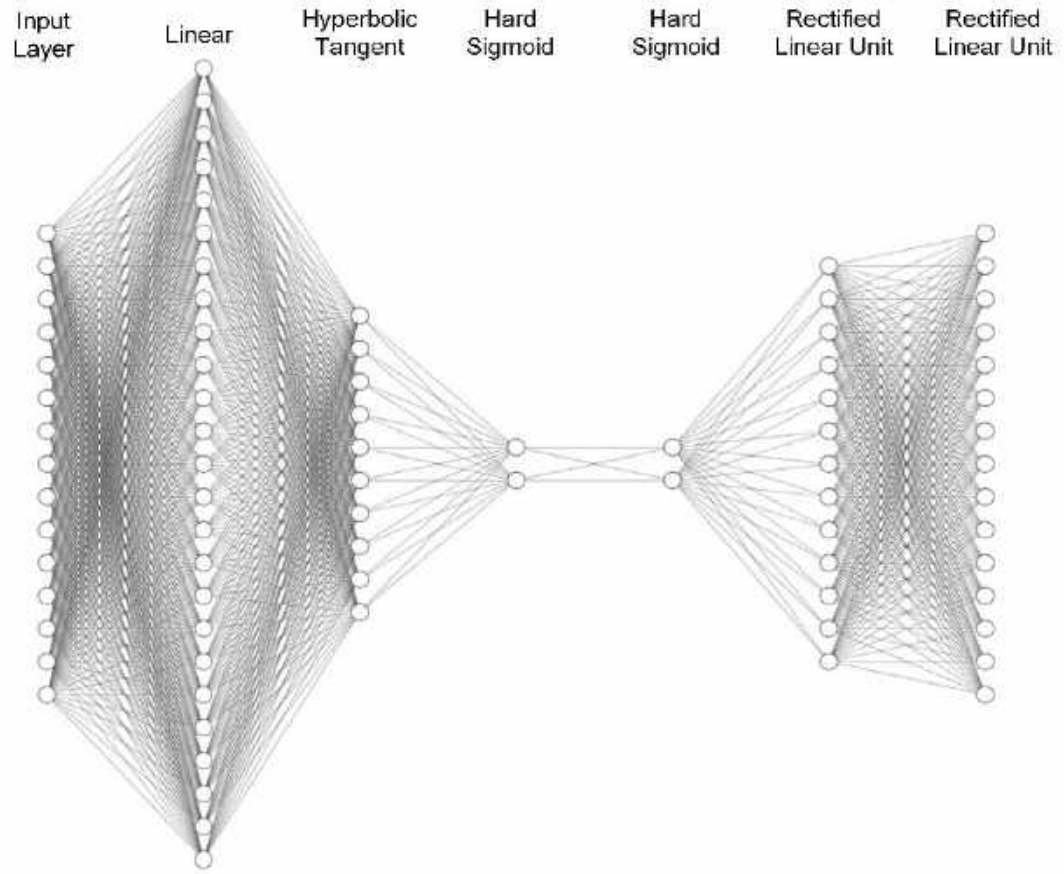


Figure 7.5: Optimal Structure of Evolutionary Deep *Autoencoder*

The performance of the evolutionary deep *autoencoder* is compared with state-of-art methodologies in dimension reduction and was shown to out-perform all prior methods (See Table 7.1).

Table 7.1: Comparison between state-of-art dimension reduction methods using 2 latent dimensions

Dimension Reduction Method	Implementation Reference	Validation Mean Absolute Error (%)
Principal Component Analysis (PCA)	Tipping and Biship, 1999	12.07
Linear Kernel PCA	Schölkopf el al., 1997	12.07
RBF Kernel PCA	Schölkopf el al., 1997	11.93
Polynomial Kernel PCA	Schölkopf el al., 1997	11.52
Sigmoid Kernel PCA	Schölkopf el al., 1997	14.03
Independent Component Analysis (ICA)	Hyvärinen and Oja., 2000	12.07
Non-negative Matrix Factorization (NMF)	Févotte and Idier, 2011	16.52
Evolutionary Deep *Autoencoder*	This work	**9.54**
Benchmark *Autoencoder*	See footnote*	16.75

*The architecture of this *autoencoder* is determined by approximately equalizing change in number of neurons in between layers. The architecture is [15,11,7,2,7,11,15] and activation functions are all sigmoid. This is a conventional heuristic.

7.5 Conclusions

This work has demonstrated the use of novel evolutionary deep *autoencoder* for the dimension reduction of industrial processing systems. The *autoencoder* neural network was applied to reduce noisy process information within an oil refinery processing system and achieved 9.54% validation MAE with only two latent dimensions. By comparing with methods such as ICA, NMF and various types of PCA, the proposed approach significantly outperformed the state-of-art methods. This demonstrates that the use of evolutionary deep *autoencoder* is a promising approach when compared with traditional dimension reduction methods.

References

Ba, J., Caruana, R., 2014. Do deep nets really need to be deep? In Advances in neural information processing systems, 2654-2662.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives, IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828.

Boukouvala, F., Muzzio, F.J., Ierapetritou, M.G., 2010. Predictive modeling of pharmaceutical processes with missing and noisy data, AIChE journal, 56(11), 2860-2872.

Cauchy, A., 1847. Méthode générale pour la résolution des systemes d'équations simultanées, Comp. Rend. Sci. Paris, 25(1847), 536-538.

Dechter, R., 1986. Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory, 178-183.

Deng, L., Yu, D., 2014. Deep learning: methods and applications. Foundations and Trends in Signal Processing, 7(3–4), 197-387.

Févotte, C., Idier, J., 2011. Algorithms for nonnegative matrix factorization with the β-divergence, Neural computation, 23(9), 2421-2456.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D., 2018. Deep reinforcement learning that matters, In Thirty-Second AAAI Conference on Artificial Intelligence.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks, science, 313(5786), 504-507.

How, B.S., Lam, H.L., 2018. PCA method for debottlenecking of sustainability performance in integrated biomass supply chain, Process Integration and Optimization for Sustainability, 1-22.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications, Neural networks, 13(4-5), 411-430.

Lam, H.L., Klemeš, J.J., Kravanja, Z., 2011. Model-size reduction techniques for large-scale biomass production and supply networks, Energy, 36(8), 4599-4608.

LeCun, Y., Bengio, Y., Hinton, G. 2015. Deep learning. Nature, 521(7553), 436-444.

Leong, W.D., Lam, H.L., Ng, W.P.Q., Lim, C.H., Tan, C.P. and Ponnambalam, S.G., 2019. Lean and Green Manufacturing—a Review on its Applications and Impacts, Process Integration and Optimization for Sustainability, 1-19.

Li, R.F., Wang, X.Z., 2002. Dimension reduction of process dynamic trends using independent component analysis, Computers & Chemical Engineering, 26(3), 467-473.

Lin, H.W., Tegmark, M., Rolnick, D., 2017. Why does deep and cheap learning work so well? Journal of Statistical Physics, 168(6), 1223-1247.

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018. Key pillars of successful energy saving projects in small and medium industrial enterprises, Energy, 158, 293-304.

Máša, V., Touš, M., Pavlas, M., 2016. Using a utility system grey-box model as a support tool for progressive energy management and automation of buildings, Clean Technologies and Environmental Policy, 18(1), 195-208.

McCulloch, W., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics, 5(4), 115-133.

Mhaskar, H.N., Poggio, T., 2016. Deep vs. shallow networks: An approximation theory perspective, Analysis and Applications, 14(6), 829-848.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning representations by back-propagating errors, Cognitive modeling, 5(3), 1.

Schölkopf, B., Smola, A., Müller, K.R., 1997. Kernel principal component analysis, In International conference on artificial neural networks, 583-588.

Seppo, L., 1970. The representation of the cumulative rounding error of an algorithm as a taylor expansion of the local rounding errors, Master's thesis.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, In Thirty-First AAAI Conference on Artificial Intelligence.

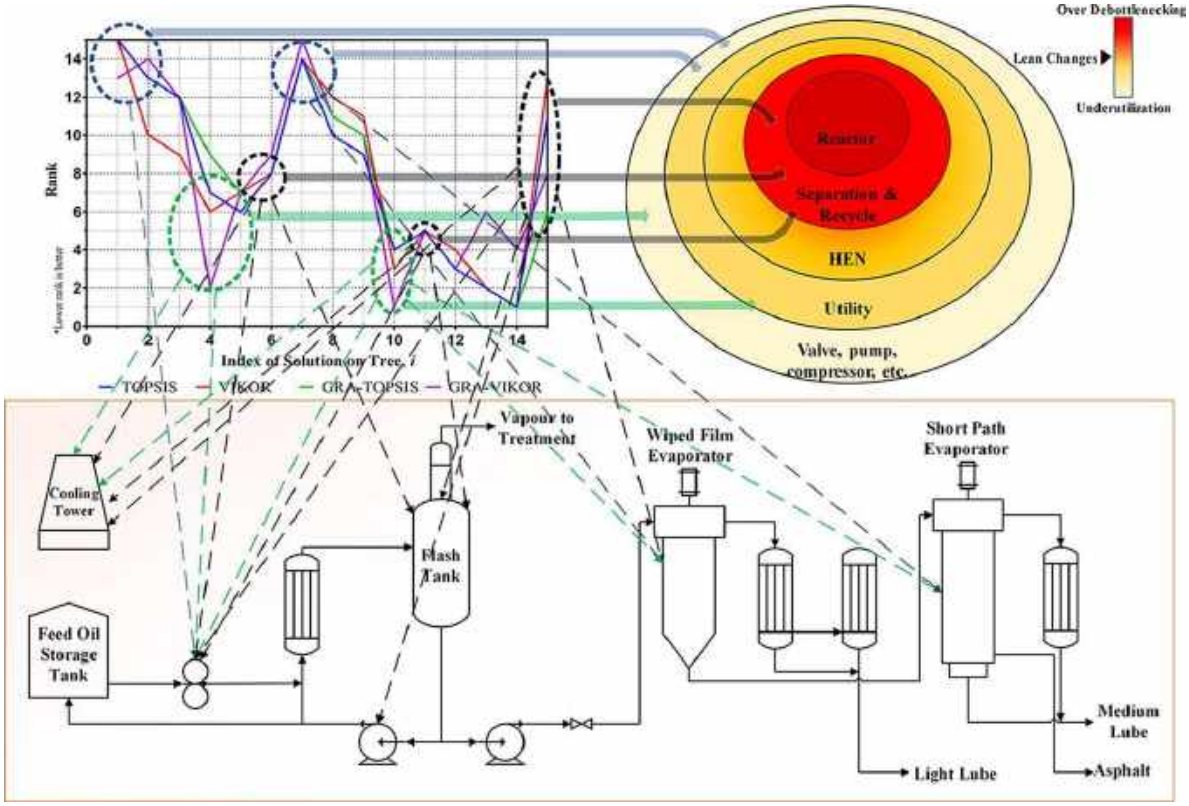
Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 611-622.

CHAPTER 8 BOTTLENECK TREE ANALYSIS: A NOVEL MULTI-OBJECTIVE METHOD TO DEBOTTLENECK PROCESS CAPACITIES

This chapter has been peer-reviewed and published in Chemical Engineering Sciences.

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H. and Lam, H.L., 2020. Bottleneck Tree Analysis (BOTA) with green and lean index for process capacity debottlenecking in industrial refineries. Chemical Engineering Science, 214, 115429.

Graphical Abstract



Abstract

This chapter presents a novel Bottleneck Tree Analysis (BOTA) to cope with the increasing capacities within industrial refineries. BOTA is a clear and concise heuristic graphical debottlenecking method that can accurately pinpoint the process capacity bottlenecks. Coupled with BOTA, four multiple criteria decision-making methods are used for an ensemble with Spearman's correlation to assess the Green and Lean Index (GLI) of industrial processes. The decision-making tool is formulated to improve operational performance with the consideration for environmental conditions. Empirically, BOTA demonstrated a debottleneck stopping mechanism which can be theoretically explained with the reversed onion model. An additional

advantage is that retrofit projects can be guided by BOTA with effective scheduling. From an industrial case study, BOTA improved normalized Global Warming Potential by 94.43 %, normalized energy consumption by 93.09 % and return on investment by 58.36 %. Project implementation by scheduling also reduced payback period from 85 to 66 months.

Keyword: Capacity Debottlenecking, Lean and Green, Multiple Criteria Decision Making, Expert System, Reversed Onion Model, Bottleneck Tree Analysis (BOTA)

8.1 Introduction

From a process retrofit and debottlenecking perspective, the major KPI that can be applied are money, material, machine, environment, and methods (defined as debottleneck 5M). Dües et al. (2013) suggested that financial (money) indicator should include return on investment (ROI) or return on sales (ROS). When making decisions of equipment replacement within lean manufacturing firms, Sullivan et al. (2002) suggested the consideration of defects in products, quality of products and production capacity which relates to poor delivery performance. Teng et al. (2019) also highlighted the importance of considering product yield and quality for a continuous chemical process. For the consideration of machine and equipment, Weng (1996) discussed that faster production lead time is directly correlated with higher equipment utilization. For the application of a batch cellular manufacturing process, Metternich et al. (2013) demonstrated the effectiveness of the average utilization factor of machine tools using machine time as a basis. However, for continuous processes and general manufacturing, Ray (2015) recommended the measurement of utilization factor in terms of capacity. This approach of maximal capacity utilization is also industrially tested in Sony, Oita where equipment is constantly fully utilized and investment in equipment is provided when required (de Haan et al., 2001). Another critical consideration for L&G manufacturing is on the assessment of environmental performances particularly under the International Organization for Standardization's (ISO) guideline (King and Lenox, 2001). For a chemical process, a comprehensive for environmental indicator can be found under the Waste Reduction Algorithm (WAR GUI, 2011) designed by United States Environmental Protection Agency where eight categories of environmental impacts were considered: (i) Human toxicity potential by ingestion (ii) Human toxicity potential by exposure (iii) Aquatic toxicity potential (iii) Terrestrial toxicity potential (iv) Global warming potential (v) Ozone depletion potential (vi) Smog formation potential and (vii) Acidification potential. Although there are many categories of

environmental impacts, only the relevant categories should be chosen during a specific environmental assessment (Bey et al., 2013).

The methods for process debottlenecking are derived from the roots of process system design. Realistic synthesis and design of process networks have mainly relied on combinations of heuristic and algorithmic frameworks (Mizsey et al., 1990). Mizsey et al. (1990) had proposed integrating heuristic approach with algorithmic mixed-integer non-linear programming (MINLP) approach for conceptual flowsheet design. The work also proposed using rigorous process simulations tools such as Aspen HYSYS for detailed design. In general, heuristic frameworks for process synthesis are mainly derived from two famous models, the first being the Onion model (Linnhoff et al., 1982) and the second being the hierarchy of decision (Douglas, 1985). The Onion diagram proposed by Linnhoff et al. (1982) was then modified by Smith and Linnhoff (1988) to generalize for different processes. Čuček et al. (2017) have also extended the Onion diagram for a biogas production network. The hierarchy of decision by Douglas (1985) mainly considered the mode and type of the processing system following by the heat exchanger network.

Systematic work on process debottlenecking has been carried out by various researchers such as Tan et al. (2018) which utilized an inoperability input-output model for the debottlenecking of the palm oil milling process. Fundamentally, Diaz et al. (1995) had applied an MINLP approach for the debottlenecking of an ethane production plant. Pinch process network technology was reviewed in the work of Dunn and El-Halwagi (2003) to debottleneck energy and waste materials in a chemical process. Liang et al. (2016) proposed a flexible design method for the optimization and debottlenecking of multi-period hydrogen networks. A level-by-level debottlenecking approach was proposed by Zhang et al. (2001) which carried out the analysis by first modifying utility networks and then retrofitting other bottlenecks in the second stage. Koulouris et al. (2000), on the other hand, proposed throughput analysis and debottlenecking of a batch chemical process via batch process simulator (SuperPro Designer software). Utilizing similar computer-aided process design methods and critical path concepts, Tan et al. (2004, 2006) had carried out research for the debottlenecking of batch pharmaceutical cream product. Alshekhli et al. (2011) have also utilized similar computer-aided batch process design methods for debottlenecking of cocoa manufacturing. A stair-step chart was proposed by Litzen and Bravo (1999) which can identify the capacity of major equipment that needs to be debottlenecked. More recently, Kasivisvanathan et al. (2014) had proposed that a process

debottlenecking framework that requires the consideration of cost to benefit analysis. To-date, P-graph (Process graph) framework is extended for debottlenecking purposes. With the aid of the P-graph Studio software, How et al. (2018a) incorporated the analytical hierarchy process (AHP) and waste reduction (WAR) algorithm for debottlenecking purposes. Lately, the effectiveness P-graph-aided debottlenecking method is benchmarked with another novel debottlenecking method which integrates the concept of principal component analysis (PCA) (How and Lam, 2018b). Aside from this, P-graph has also been combined with Monte Carlo simulations to debottleneck the carbon network by Tan et al. (2017).

A brief comparison of the topics covered by this work and other relevant works can be found in Table 8.1. In general, most of the works merely consider single criterion during the debottlenecking process. For instance, Andiappan et al. (2014) had developed a heuristic framework to retrofit an existing plant. In their work, benefit-cost ratio (BCR) was used as a sole indicator to justify whether the debottlenecking process should be carried out. On top of that some of these works were focused on the debottlenecking at conceptual design phase instead of on the retrofit scenario. As an example, Heitmann et al. (2017) had developed a design framework for the early stages of plant design while Kasivisvanathan et al. (2014) proposed a step-by-step framework which was able to identify bottlenecks of an integrated biorefinery plant and subsequently debottleneck the process at the design stage. Besides, it is worth mentioning that none of these works have covered the project scheduling which the rigorous debottlenecking activities can be decomposed into multiple smaller stages based on certain financial constraints. These situations highlight a research gap and the needs for a more comprehensive retrofit phase debottlenecking framework.

Table 8.1: Comparison of current work with previous works

Work	Industrial phase	Consider capacity with multiple stages	Consider multiple criteria	Stopping mechanism for debottlenecking	Implemented with scheduling
Litzen and Bravo (1999)	Retrofit phase	Yes	Not specified	No	No
Liang et al. (2016)	Retrofit phase	Yes	No	No	No
Kasivisvanathan et al. (2014)	Design phase	No	No	Yes	No
Seifert et al. (2015)	Design phase	Yes	No	No	No
Andiappan et al. (2017)	Retrofit phase	No	No	Yes	No
Heitmann et al. (2017)	Design phase	No	Yes	Not applicable	No
This work	Retrofit phase	Yes	Yes	Yes	Yes

Besides capacity debottlenecking, researchers have used modifications and intensifications of individual units to debottleneck the overall process. Such a concept has been demonstrated in the works of Gai et al. (2018) where coal-gasification technology was analysed and retrofitted for debottlenecking. Rigorous computational fluid dynamic (CFD) methodologies were also used by Lee et al. (2008) for the debottlenecking of the high-pressure and low-pressure separator. Van Duc Long et al. (2010) had improved a dividing wall column for the debottlenecking of an acetic acid purification process. Van Duc Long and Lee (2011) also utilized a fully thermally coupled distillation column for process energy debottlenecking.

All the reviewed works are admirable, however, most of the work only considers costs for debottlenecking (comparison can be found in Table 8.1). In the 21st century, aspects such as environmental impacts, energy consumption and product quality are increasing in concern.

Economic costs only act as a short-term objective. For this issue, Walker (2011) provided evidence that many plants that prioritize costs and neglect environment impacts are to downsize or decommission by regulation. Long-term sustainability of processing plants must be considered by using lean and green indicators (Leong et al., 2019b). Furthermore, most works did not highlight the debottlenecking hierarchy based on the design layers in the onion model. To address all these gaps, this work takes a fundamentally different concept, theory, and implementation. The novelty of this work is:

1. The utilization and novel formulation of Green and Lean Index (GLI) for process plant debottlenecking based-on an ensemble of MCDM methods.
2. The formulation of a novel graphical method for the schedule implementation and visualization of debottlenecking projects.

In terms of contribution towards debottlenecking analysis within the fields of process system engineering, this work contributed in:

1. Providing algorithmic focus on long-term sustainability objectives in debottlenecking projects.
2. Developing a concrete method on knowing when to stop debottlenecking (theoretically explained by reversed onion model) and to visualize the project roadmap.
3. Efficient implementation of stage-by-stage debottlenecking projects with scheduling.

8.2 Problem Statement

Onion model is a graph-based diagram which provides an explicit illustration of the hierarchy of process design (Linnhoff et al., 1982). There are four layers in the early design of the onion model. Starting from the core of the onion model, it involves reactor design which concerns the selectivity and capacity, whereas the separation and recycling subsystems are considered in the second layer. Since the reaction and separation layers define the energy requirement, the third layer, therefore, considers the heat exchanger network (HEN). Finally, to enable the heat recovery system into the process, utility system design is included in the outermost layer. The model had been modified and improved to consider other elements in process design, e.g., valves, pumps, compressors, etc. (see Figure 8.1(a)).

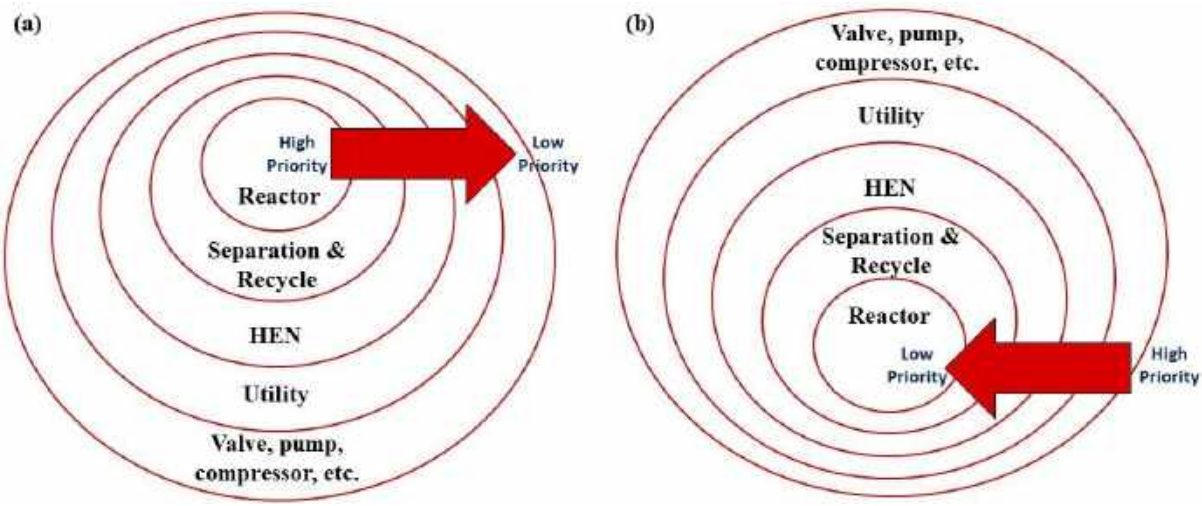


Figure 8.1: (a) Onion model for process design (modified from Kemp(2007)); (b) Reversed onion model for process debottlenecking

However, the direction of the hierarchy may be reversed when decision-makers are seeking for strategies in process improvement, in other words, debottlenecking. This work suggests that the general strategy of the process improvement model is expected to follow the Reversed Onion model (as illustrated in Figure 8.1(b)), where the bottleneck in the outermost layer should be prioritized in order to ensure sufficient economic effectiveness. With the Reversed Onion model heuristic, one can achieve the same debottlenecking outputs with the least requirements in terms of changes and investments. Fundamentally, removal of one bottleneck may create a new bottleneck (can be in the same or different layer) (Kasivisvanathan et al., 2014). If the debottlenecking process is applied without appropriate and systematic "stopping mechanism" (a way to terminate the debottlenecking process when its resulting performance is no longer attractive), the model will suggest building an identical plant or plants eventually. To ensure the suggested solutions are feasible in terms of economic performance and overall sustainability, a *Green and Lean Index (GLI)* is introduced and used as an indicator for "stopping mechanism". The proposed reversed onion model concept is demonstrated using a real industry case study (refer to Chapter 4.0 for detailed for the case study).

8.3 Method

A novel framework is proposed to systematically debottleneck the process while ensuring the attainment of green and lean performance of the system (see Figure 8.2). Firstly, process data including operating conditions (pressure, temperature, etc.) and raw material specification (concentration, etc.), as well as the unit capacity of each equipment are systematically collected.

With the aid of these data, a simulation model can be constructed through Aspen Hysys to determine the material and energy flow of the entire process. The process capacity in the process simulation is then incrementally increased from its original operating conditions to determine the utilization factors of each equipment and various important variables in the process (i.e. product flowrate, product yield, energy consumption, product quality, waste flowrate, waste concentration). Then, a bottleneck tree is drawn to analyse the "bottleneck stages" of the current process (refer to Chapter 8.3.1 for the detailed description of the bottleneck tree). In short, the "bottleneck stages" indicate the debottleneck hierarchy of the process. Debottlenecking will be applied from outside to inside of the reversed onion diagram (see Figure 8.1(b)). The green and lean index (GLI) filled into the bottleneck tree, acting as an indicator to show whether the debottlenecking strategy is desirable (refer to Chapter 8.3.2 for the detailed description of GLI). The point on the tree at which improving the capacity of the process in anyway does not improve the GLI, is called the bottleneck stopping point (BSP). A project pathway is constructed from the current operating conditions to the BSP by using the bottleneck tree. Finally, implementation of the retrofit project can be carried out by considering scheduling with the constructed bottleneck pathway. In this work, a waste oil re-refinery plant case study from Pentas Flora Sdn. Bhd. is employed to demonstrate the proposed method (see Chapter 8.4).

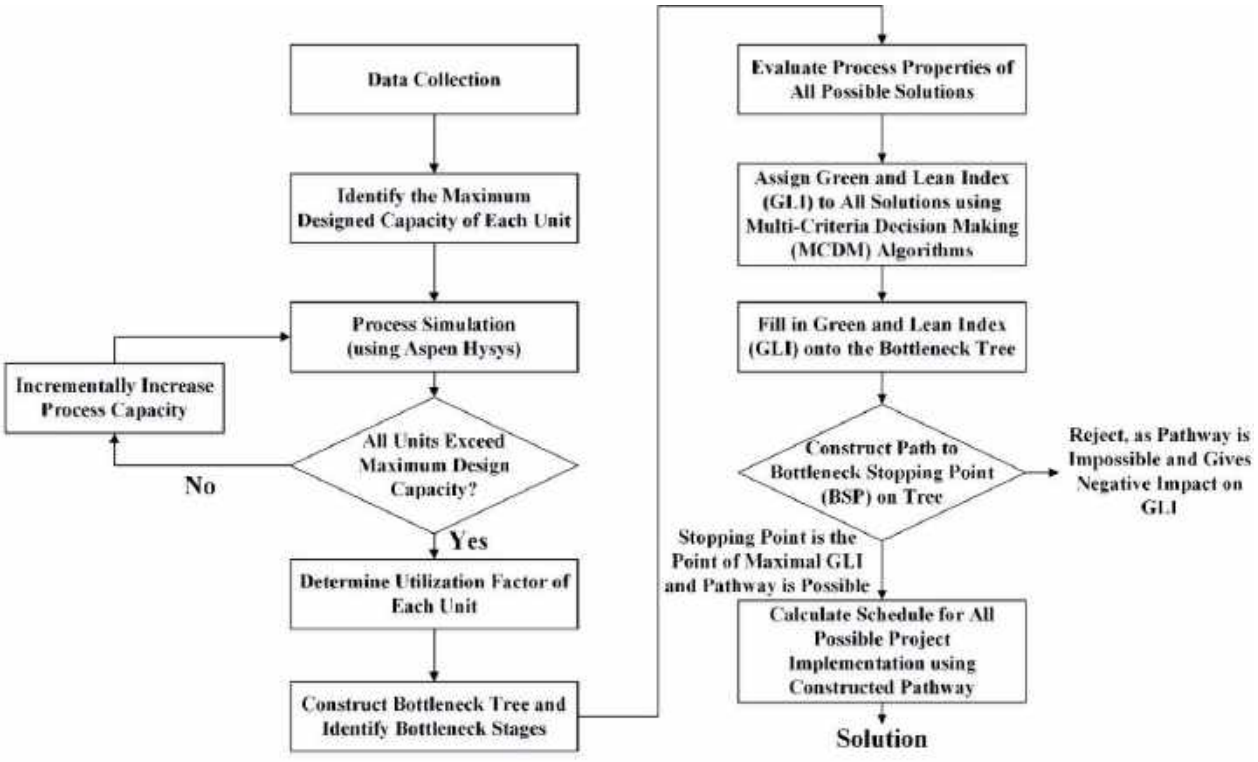


Figure 8.2: Procedure of bottleneck tree analysis

8.3.1 Conceptual Idea of a Bottleneck Tree

For the current process plant design practice, individual units within a processing system are designed for a required size with an additional safety margin (Kasivisvanathan et al., 2014). However, business expansion and supply-demand gap variation in the plant often require the processing system to operate at a varied production capacity (Kasivisvanathan et al., 2014). This framework recommends debottlenecking the process after operating conditions have been optimized to identify capacity bottlenecks within the process to maximize the utilization of each unit. Hence, the design and utilization of each equipment are studied.

From the work of Ahmad and Hui (1991), bottlenecks within the process were proposed to be addressed by decoupling the problem into regions of integrity. Later, this concept was extended in the works of Matsuda et al. (2009) to solve heat exchanger network bottleneck issues by decoupling the case study into "sites". This decoupling technique can be also found in the works of Bagajewicz and Rodera (2002), where a total site was decoupled into a few separated processes for debottlenecking. Hence, this work also proposed to use a similar decoupling technique to separate the process into two regions for region-wise debottlenecking. Next, Kasivisvanathan et al. (2014) had identified that debottlenecking works can be carried out in stages, which was then related to cost to benefit ratio (CBR) in their work. Litzen and Bravo (1999) had also observed this phenomenon of multiple debottleneck opportunities. Therefore, this work further clarifies the stages of opportunities within a capacity debottlenecking analysis. Thus, a concise bottleneck tree analysis (BOTA) is proposed in this work to systematically identify the stage-wise debottlenecking opportunities of different process regions.

In this work, multiple stages of opportunity for debottlenecking are studied at a different feed rate in the process. The initially designed capacity of each unit was collected in the data collection phase. After the initial stage, the maximum design capacity is set to the capacity at which the utilization factor of the next unit within the processing region reaches 100%. With this, the utilization factor of each equipment (indicate the portion of capacity which is currently utilized) can then be computed using Equation (8.1):

$$Utilization\ factor = \frac{Current\ Operating\ Capacity}{Maximum\ Design\ Capacity} \tag{8.1}$$

Note that giving that the nature of each equipment is different, the capacity unit used in Eqn. 8.1 will be different as well (see Table 8.2).

Table 8.2: Equipment capacity indicator and the field unit

Equipment	**Equipment Capacity Indicator**	**Unit**
Heat Exchanger	Heat Transfer Area	m^2
Separator/Evaporator	Inlet Flow Rate	lpm
Pump	Electricity Power	kW
Cooling Tower	Cooling Power	RT
Heater/Furnace	Heating Power	MMBTU/hr

By gradually increasing the input flow rate of the process, the utilization factor of one unit will gradually approach unity (i.e. 1), indicating that it is fully utilized. The respective inlet flow capacity is noted as the first bottleneck stage. To debottleneck this, the current fully utilized equipment is replaced with larger-size equipment that provides the same functionality which can accommodate the next bottleneck stage. The process can be repeated to remove the new bottleneck until debottlenecking is no longer provide positive benefits (GLI is used as the indicator in this work; refer to Chapter 8.3.2 for more detail). It is worth to note that, all these can be illustrated in a novel tree diagram, called a *bottleneck tree*.

At each stage, the unit with full utilization factor (100%) is suggested to be resized to the operating capacity of the next bottleneck stage. A novel tree diagram is plotted to demonstrate the capacity bottleneck of a process. The tree diagram can directly identify process equipment that is being the bottleneck of the process. An example bottleneck tree diagram is shown in Figure 8.3.

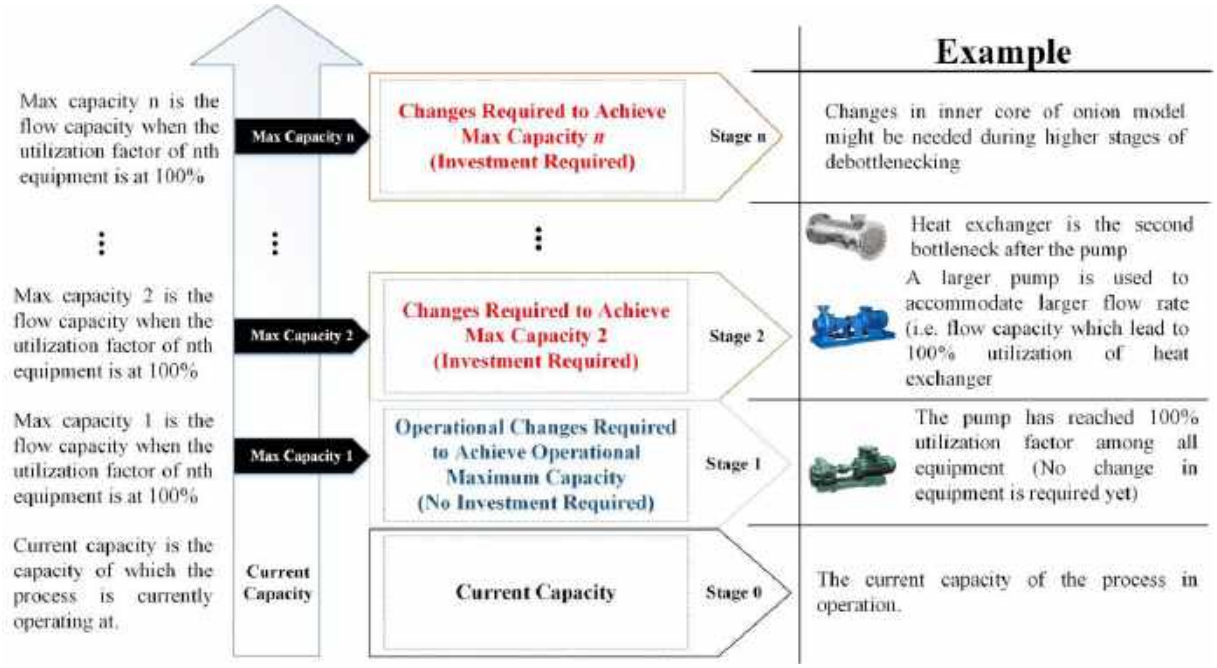


Figure 8.3: An example of a bottleneck tree diagram

The bottom level of the tree diagram shows the current capacity and the current critical unit capacity. It is followed by max capacity 1, max capacity 2 and so on (up to max capacity n). Note that max capacity 1 refers to the maximum capacity that the process can achieve without additional changes of equipment. At this flow capacity, one of the equipment in the process had operated at its maximal. Usually, in a well-designed plant, equipment at the outer layer of the onion model such as pumps and valves will hit the cap before others. The process can be debottlenecked to achieve maximum capacity 2 (i.e., flow capacity when the utilization factor of the second equipment = 1). This can be done by replacing the first bottleneck to one with a higher capacity which provides the same functionality. For instance, in the example shown in Figure 3, after identifying the pump as a bottleneck, the next bottleneck after that is the heat exchanger. In order to fully operate the heat exchanger, the first bottleneck (i.e. pump) has to be replaced with another larger pump. As mentioned, this bottleneck analysis and debottlenecking steps can be repeated continuously, from the outer layers to the inner layers until debottlenecking works are no longer justified in terms of GLI. The use of GLI as an indicator to justify the debottlenecking process ensures that the debottlenecking process can be employed with the consideration of both economic and environmental sustainability.

For a process with more than one region, the bottleneck trees can be cascaded to form a bottleneck tree matrix. The dimension of the bottleneck tree matrix is equal to the number of

regions, while the size of the matrix is equal to the stages of bottlenecks within each respective region plus 1. For example, a process has 2 regions with 2 and 1 stages of bottlenecks respectively, the corresponding bottleneck tree matrix is a 2-dimensional matrix with a size of 3x2 (See Figure 8.4).

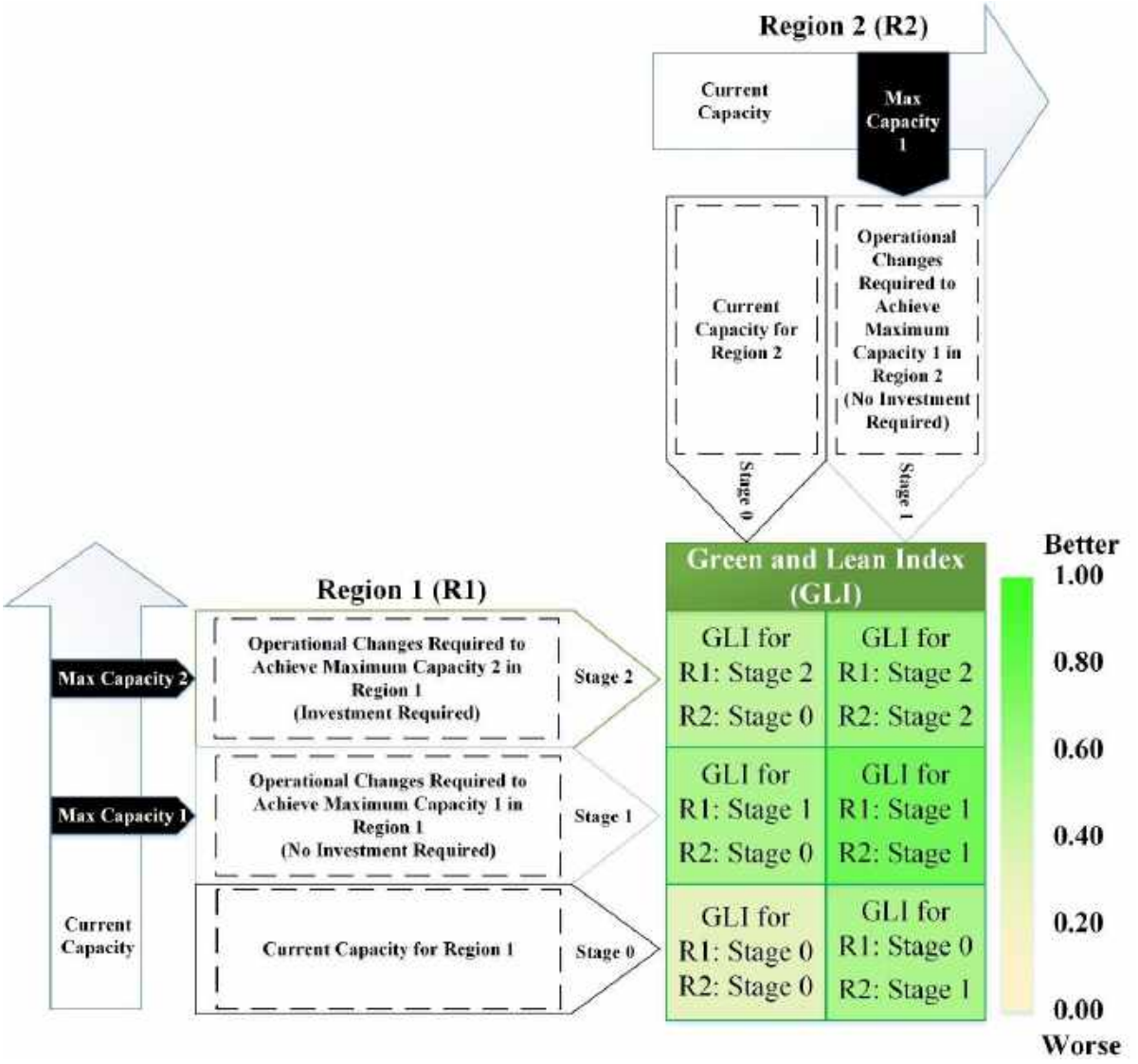


Figure 8.4: An example of a bottleneck tree matrix

The method for obtaining these values of GLI will be discussed in Chapter 8.3.2. Note that it is possible that a single cell of GLI cannot be calculated where there is violation of thermodynamics or mass conservation in the process, in such cases, GLI is ignored and treated as 0 or "-". Since GLI is a direct indicator of the sustainability of the process, the most desirable point of debottlenecking would be the global maximum point within the bottleneck tree matrix. This is the point where retrofit project cannot improve the sustainability of the process by

further debottlenecking activities. This point is defined as the *bottleneck stopping point (BSP)*, with the mathematical representation in Equation (8.2).

$$BSP = global\ max\ M \tag{8.2}$$

Where M is the bottleneck tree matrix. The *BSP* is also the point at which the process can achieve maximal benefits over time if enough investment funds are provided for its immediate construction. Note that this global maximum procedure is not computationally expensive (and does not require a solver most of the time) as it is just finding the maximum value in a matrix.

However, in practice, retrofit projects are carried out in stages due to financial concerns, risk mitigation and the requirements for step-by-step feasibility evaluation. Thus, a schedule to meet the point of maximal benefits (i.e. BSP) should be carried out with the intermediate phases of the project stopping on cells within the bottleneck tree matrix with sufficiently high GLI. This work defines the procedure as the bottleneck pathfinding. The bottleneck path is asserted to have the longest path possible from origin (original operating point) to the BSP provided that each step taken is an improvement in GLI and the capacity does not decrease during steps. Note that the number of steps is maximized (Equation 8.3) so that there can be more combinations of project stages during scheduling, which allows for incremental debottlenecking of process to gain an increasing amount of profits. The longest path determination is a heuristic for solution space reduction in this case. This procedure can be mathematically represent as following.

Suppose $M[i, j]$ is the GLI at which the stage for Region 1 and Region 2 is i and j respectively, $M[0,0]$ is the origin and $M[BSP_i, BSP_j]$ is the BSP. Each step taken moves the origin to an equal or higher capacity (refer to Equation (2) where x_n and y_n are either 0 or positive integers). Moreover, the subsequent step should arrive at a cell in the bottleneck tree matrix where the GLI is higher than the original cell (see Equation (8.5)). It is set that the starting point is the origin, and the final point is the BSP (see Equation (8.6)). Under these conditions, it is guaranteed to find a path from origin to BSP that is increasing in process benefits (in terms of GLI) and does not revert in capacity (i.e. not removing or idling retrofitted equipment from process). The bottleneck path is defined as the longest path possible taken, hence, the final number of project stages, S is maximized (see Equation (8.3)). This allows for maximal combinations of project stages during scheduling. In the case of multiple bottleneck pathway,

the bottleneck pathway is taken for the path which contains the cell of stage 1 in all region, *M[i, j]* (refer to Equation (8.7)).

$$\max S \tag{8.3}$$

$$M_{n+1}[i,j] = M_n[i + x_n, j + y_n] \quad \forall n \in [0,S] \quad where\ x_n, y_n \in \boldsymbol{N} = \{0,1,2,3 \dots\} \tag{8.4}$$

$$M_{n+1}[i,j] > M_n[i + x_n, j + y_n] \tag{8.5}$$

$$M_0[i,j] = M[0,0]\ \ and\ \ M_S[i,j] = M\left[BSP_i, BSP_j\right] \tag{8.6}$$

$$if\ count\big(argmax(S)\big) > 1 \quad then \quad S = S_p \quad where\ \ M[1,1] \in Path(S_p) \tag{8.7}$$

where $M_n[i, j]$ represents GLI at stages for Region 1 and Region 2 of *i* and *j* respectively during "*n*" project stage. x_n and y_n are the manipulated steps for the path in *i* and *j* direction during "*n*" project stage and are strictly natural numbers.

These steps (Equation (8.3-8.6)) may look sophisticated in terms of mathematical representation. However, they can be easily described in 5 rules on the decision tree matrix:

Rule 1: The bottleneck pathway should start from origin and end at BSP.
Rule 2: The bottleneck pathway can only move from low capacity to high capacity.
Rule 3: The bottleneck pathway can only move from low GLI to high GLI.
Rule 4: Jumping through a cell is allowed, however always take the longest possible pathway. (Note that it is possible to have two longest pathway)
Rule 5: If there is more than one longest pathway, only accept pathways that pass-through cell of stage 1 in all region. (Otherwise, accept all pathways)

An example of these rules is shown in Figure 8.5. Figure 8.5(a) shows the position of origin and BSP. The BSP is the maximum value in the matrix. Figure 8.5(b) shows that each path must move from high GLI to low GLI, and not move backward (not move from high capacity to low capacity). Figure 8.5(c) shows that there are two possible longest pathways, and both are considered valid. All the bottleneck pathways are then used as reduced pathways for scheduling.

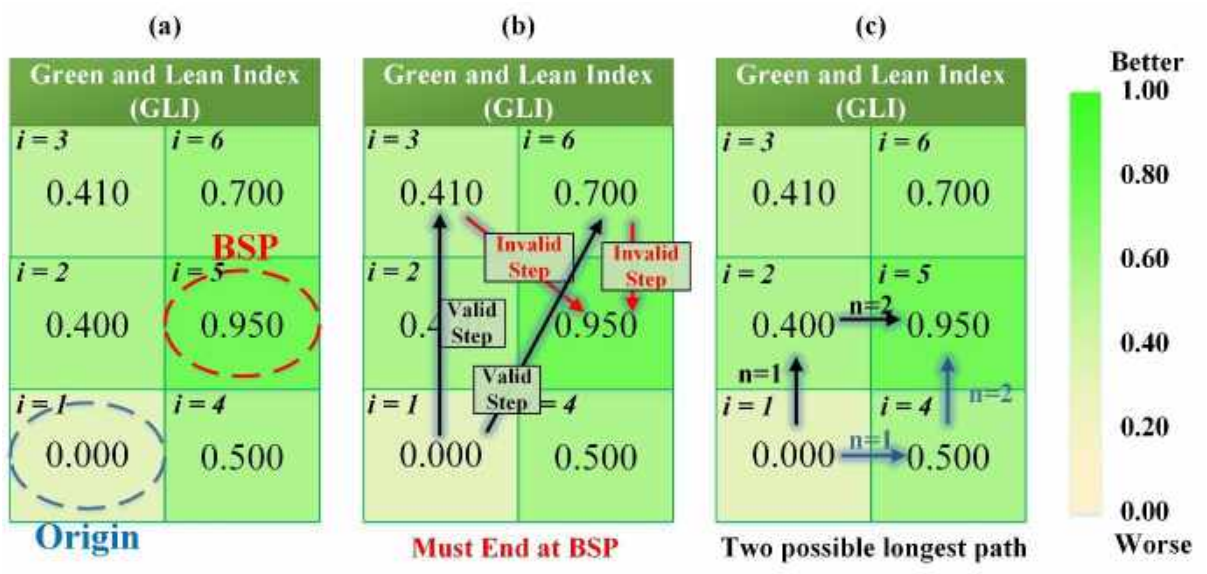


Figure 8.5: Examples of rules in bottleneck tree matrix (*i* is assigned on cells for easier reference)

Since the number of possible pathways has been significantly reduced by only considering the bottleneck pathway (from Figure 8.5(c)), all the possible sub-pathway of the bottleneck pathway can be derived (see Figure 8.6). A sub-pathway can be obtained by removing one or a few intermediate steps from the bottleneck pathway(s), while maintaining the same starting point and ending point.

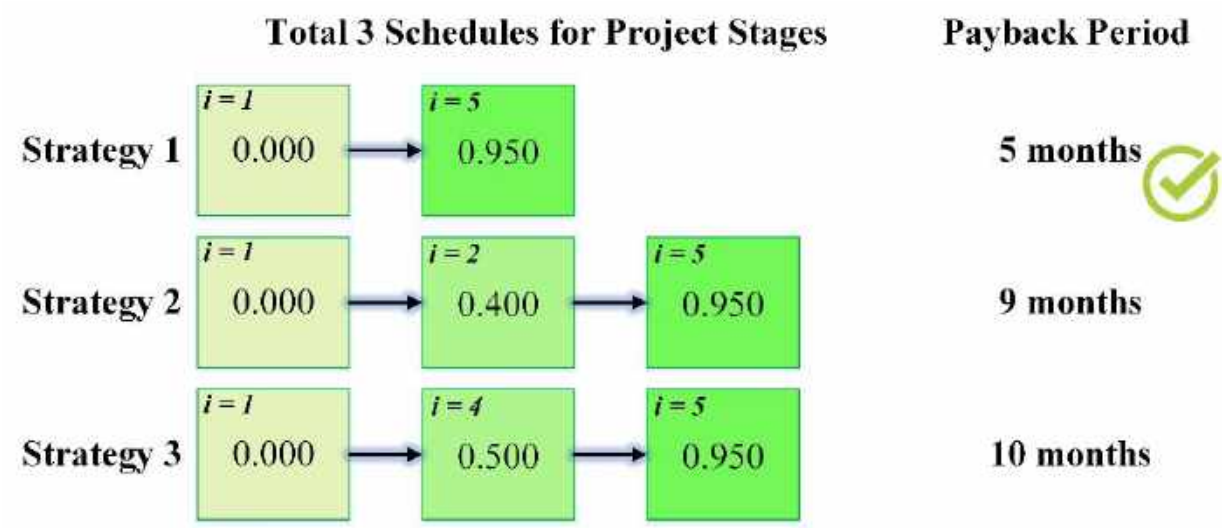


Figure 8.6: Examples of evaluating all possible sub-pathways and strategies

Subsequently, the payback period of each strategy (sub-pathways) can be calculated based on the economics and situation of the specific retrofit plant. This work recommends the use of a simple net present value calculation using Equation 8.8 and 8.9.

$$NPV_S(t) = II + NP_s(t) - IC_s(t) \quad (8.8)$$

$$NPV_0(t) = II + NP_0(t) \quad (8.9)$$

where $NPV_s(t)$ is the time-dependent net present value for strategy S while II is the initial investment costs. $NP_s(t)$ is the time-dependent net profit for strategy S and $IC_s(t)$ is the time-dependent investment costs required for strategy S. When strategy $S=0$, no debottlenecking is carried out. When $NPV_s(t)=NPV_0(t)$, t is the payback period. For the case study, this is solved graphically (See Figure 8.17). Since the number of strategies is always significantly lesser, the payback period can be directly compared for each solution, and the minimal can be selected.

8.3.2 Formulation of Green and Lean Index (GLI)

Generally, the motivation of L&G approach is to provide value-added elements for an organization. This work considers using the L&G concept in reducing non-value-added elements by capacity debottlenecking. As reviewed in Chapter 8.1, the relevant five indicators (debottleneck 5M) for debottlenecking a process are money, material, machine, environment, and methods. For a typical processing plant, the main consideration for money is the return on investment (ROI) while considerations for materials are product quality and production capacity. Moreover, the machine utilization factor is also shown to be important for both L&G manufacturing (Metternich et al., 2013) and process debottlenecking (Litzen and Bravo, 1999). For the aspect of environmental performance, a standardized and comprehensive method called the Waste Reduction Algorithm (WAR GUI, 2011) developed by the United States Environmental Protection Agency is recommended in this work.

Nevertheless, the biggest dilemma for L&G manufacturing is on balancing the desirability between economics (money), production (material and machine) and environmental performances. Although researchers such as Dieste et al. (2019) suspect that these performances can be aligned together with the activities of L&G manufacturing, there are still undeniably cases where they conflict (Dües et al., 2013). This commonly results in a scenario where decision making must be carried out with some compromises. Research such as

Carvalho et al. (2019) developed an integrated Green and Lean Index (GLI) to evaluate the performance of lean tools organization performance. To solve the L&G conflict dilemma, Leong et al. (2019b) also developed an adaptive GLI that focuses on continuous process improvement. Furthermore, the work has shown that the development of GLI is highly dependent on the nature of the implemented industry.

The GLI is a composite index which measures multidimensional elements which could not be represented by a single indicator (Joint Research Centre-European Commission, 2008). To maximize the debottlenecking outcome, performance elements which include aspects of money, material, machine, and environment must be considered into the developed of GLI. Prior to the formulation of the GLI, the relationship between each element shall be established to develop a better understanding of the motivation of the index (Sufiyan et al. 2019). In order to comprehend the development of GLI, multi-criteria decision-making (MCDM) algorithms are used to determine the relationship between elements. The MCDM is a complex decision-making tool that takes into consideration of both quantitative and qualitative factors for better decision making (Mardani et al. 2015). It also aims at selecting the best alternative option amongst the available options. Vinogradova et al. (2018) expressed that the elements in MCDM are commonly represented by weights. The weights represent the element's quantitative significance and influence on the desired results. The weights of elements are conventionally normalized, ranging from 0 to 1, with a total sum of 1. Therefore, elements that have higher weight reflect higher priority or ranking when compared to other elements. Moreover, a solution with higher composite index (i.e. GLI) also reflects higher desirability.

Historically, multiple criteria decision-making (MCDM) within debottlenecking studies has been carried out using analytical hierarchy process (AHP) by How et al. (2018a) for biomass industry and Tan et al. (2016) for treatment of microalgae. Principal component analysis (PCA) was also used How et al. (2018b) for debottlenecking studies in biomass networks. AHP utilizes a pair-wise scoring method and predefined steps to convert human intuition into relative scores (Saaty, 1987). However, AHP is prone to the human bias of different groups and does not draw focus on success-critical factors (Ahmad et al., 2010). On the other hand, PCA is a dimension reduction method that utilizes an orthogonal transformation to convert studied variables into linearly uncorrelated variables, called principal component. From the PCA study, the covariance of each studied variable is then used for MCDM scoring (Pearson, 1901). In one of the recent works, Lever et al. (2017) claimed that PCA is only effective on complete high

dimensional datasets with linear correlations. However, within a process plant, the decision-makers and stakeholders always consist of human organization of different backgrounds, discipline, and motivation (Pechmann, 1948). Hence, the use of AHP is fundamentally flawed for this application. On the other hand, the use of PCA also possess problems for process plant debottlenecking, as process plant systems commonly exhibit nonlinear behaviors (Carr and Shearer, 2007). Thus, an MCDM method that is not reliant on human input, while is able to analyse incomplete nonlinear information, is required. Authors would like to highlight that there is a significant likelihood of subjectivity when using only a single MCDM method for assessment. To minimize the effects of subjectivity, multiple MCDM method is used as an ensemble for the formulation of GLI. Spearman's correlation is also used to draw statistical significance using p-value, so that the GLI can be statistically supported.

In this work, four different MCDM approaches, namely TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), VIKOR (Višekriterijumsko Kompromisno Rangiranje); methods without requirement of weight input: GRA (Grey Relational Analysis)-based TOPSIS and GRA-based VIKOR are proposed to rank the solution based on their green and lean performance. These four methods are conventional powerful methods which are commonly used in multiple-criteria decision making (MCDM) problems. In this work, the similarity between ranks obtained from each MCDM method will show the effectiveness of BOTA study coupled with the respective MCDM method. Moreover, the respective weights used to rank the solutions are served as GLI in this work. The general mathematical formulation stages are illustrated in Figure 8.7. The scenarios of each cell in the bottleneck tree matrix with different regional capacity (see Figure 8.4 for example) can be simulated by process simulation (see Figure 8.8) to obtain their elements related to the bottleneck 5M. All the elements of each cells in the bottleneck tree is then pre-processed with a min-max data normalization to standardize their values. As discussed, this work utilizes two weight determination method that does not rely on human input but mathematical information theory, the methods are Shannon Entropy and Grey Relational Analysis (GRA). Both the entropy weights and the GRA weights are then used in an ensemble of MCDM algorithm which includes TOPSIS, VIKOR, GRA-TOPSIS and GRA-VIKOR. Spearman's correlation test is used to remove the results of MCDM method that do not follow the majority of the MCDM algorithm within ensemble. The ranking index of the selected MCDM algorithm will then be taken as the GLI and used on the bottleneck tree. Authors used Visual Basic for Application in Excel for implementing these procedures, however, they can be easily implemented regardless of software and platform, even

on pen and paper. The weight determination methods and MCDM algorithms are further described in the subchapters below:

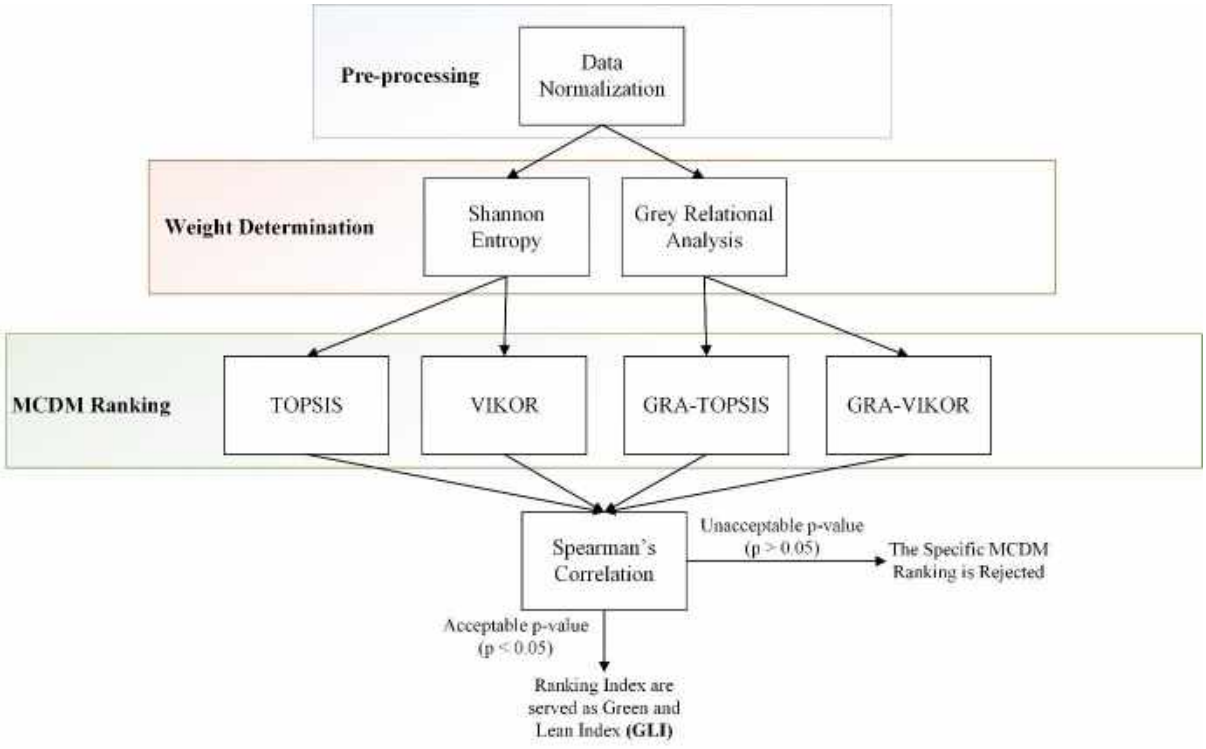


Figure 8.7: Mathematical formulation steps for green and lean evaluation

8.3.2.1 Data Normalization

Assume a decision matrix G of an MCDM problem is defined by y_{ij}, the performance of j criteria of alternative i (see Equation (10)), where n and m refer to the total number of criteria and possible alternatives respectively. Note that in this work, the term "alternative" refers to the different stages of process debottlenecking.

$$G = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \tag{8.10}$$

Since the data are expressed in a different scale, normalization is carried out to adjust the value of each criterion to a notionally common scale which enables the fair comparison between them. In this work, the normalized performance, x_{ij} is calculated as Equation (8.11).

$$x_{ij} = \frac{y_{ij} - {y_{ij}}^{min}}{{y_{ij}}^{max} - {y_{ij}}^{min}} \qquad \forall i \in I, \forall j \in J \tag{8.11}$$

where ${y_{ij}}^{min}$ and ${y_{ij}}^{max}$ refer to the minimum and maximum values of y_{ij}. As a result, the ${y_{ij}}^{max}$ will be normalized as "1", while "0" on the other hand, will be the normalized value for ${y_{ij}}^{min}$.

8.3.2.2 Shannon Entropy

This work proposed the use of a systematic approach for formulating prioritization weights during debottlenecking, called the entropy method. Entropy method has been commonly used in thermodynamics and has started to emerge into the information theory (Shannon, 1948). Entropy is the measure of randomness where the more inconsistent the variable of observation, the higher the measure of entropy. Based on the Shannon entropy, fundamentally, a smaller entropy of one criterion indicates that its corresponding information is more useful and meaningful (or lower uncertainty) compared to others (Zou et al., 2006). Thus, this criterion will be prioritized (higher value of weight). The measure of Shannon Entropy is formulated as an averaged probabilistic negative logarithmic function of the criterion of interest. Equation (8.12) is used to define the Shannon Entropy, of criterion *j*.

$$H_j = -\frac{1}{\ln m}\sum_{i=1}^{m} x_{ij} \ln\left(x_{ij}\right) \quad \forall j \in J \tag{8.12}$$

where -1/ln m is a constant that constraint H_j in between the value of 0 to 1. Note that suppose when x_{ij} is equal to "0", $x_{ij} ln\left(x_{ij}\right)$ is equal to "0" as well.

Finally, the weight of Shannon entropy for criterion *j*, w_j is defined as Equation (8.13). This value is used as the relative priority scale for each criterion. Using this prioritization weight, criteria that can be more significantly affected by debottlenecking are prioritized from their variability. This allows the MCDM ranking to assign higher GLI to a debottlenecking solution that has greatly improved the criteria, instead of one that has not made many changes (refer to Appendix for example of calculation).

$$w_j = \frac{1-H_j}{n-\sum_{j=1}^{n} H_j} \quad \forall j \in J \tag{8.13}$$

8.3.2.3 TOPSIS

TOPSIS (or Technique for Order Preference by Similarity to Ideal Solution) was initially developed by Hwang and Yoon (1981) and currently being used to evaluate the performance of alternatives through the similarity with the ideal solution. The core concept of TOPSIS is to determine the best solution based on the "closeness" between each feasible solution and the ideal solution (Li et al., 2011). In short, a solution which is closer to the positive ideal solution (minimize criteria with a beneficial attribute; minimize criteria with harmful attribute) and farther from the negative-ideal solution (minimize criteria with a beneficial attribute; maximize criteria with a harmful attribute), is considered as a better solution. The formulation steps are given as follows:

The weighted normalized matrix, z_{ij} is determined via multiplication of the normalized matrix (in Equation (8.11)) and Shannon weight w_j (in Equation (8.13)):

$$z_{ij} = x_{ij} w_j \qquad \forall i \in I, \forall j \in J \tag{8.14}$$

The "closeness" between each feasible solution and the positive-ideal solution, S_i^+ can be computed using Equation (8.15), while Equation (8.16), on the other hand, is used to determine the "closeness" between the feasible solution and negative-ideal solution, S_i^-.

$$S_i^+ = \sqrt{\sum_{j=1}^{n} (z_{ij} - z_j^+)^2} \qquad \forall i \in I \tag{8.15}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (z_{ij} - z_j^-)^2} \qquad \forall i \in I \tag{8.16}$$

where z_j^+ and z_j^- refer to the positive-ideal and negative-ideal values of criteria j. In short, these values are obtained based on Equations (8.17) and (8.18). Note that benefit-criteria are the criteria which should be maximized (e.g., profit), while the criteria which need to be minimized (e.g., environmental impacts) are noted as cost-criteria.

$$z_j^+ = \begin{cases} \max z_{ij}\big|_{j=benefit-criteria} \\ \min z_{ij}\big|_{j=cost-criteria} \end{cases} \qquad \forall j \in J \tag{8.17}$$

$$z_j^- = \begin{cases} \min z_{ij}\big|_{j=benefit-criteria} \\ \max z_{ij}\big|_{j=cost-criteria} \end{cases} \qquad \forall j \in J \tag{8.18}$$

Finally, the feasible solutions are ranked based on the corresponding relative closeness, C_i which is determined through Equation (8.19). The higher the value of C_i indicates that the respective solution is higher in rank.

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \qquad \forall i \in I \tag{8.19}$$

8.3.2.4 VIKOR

VIKOR (or Višekriterijumsko Kompromisno Rangiranje) is known as a compromise ranking method which was introduced by Opricovic (1990). It helps decision-makers to rank the feasible solutions based on the measure of "closeness" to the "ideal" solution (Opricovic, 1998). In general, it maximizes the group utility and minimizes the individual dissatisfactory to deal with MCDM problems (Yang et al., 2015). It is a powerful tool for MCDM problems, especially when decision-makers are not able to express their preferences initially (Opricovic and Tzeng, 2007). The formulation steps are given as follows:

First, the value of E_i and F_i are computed via Equations (8.20) and (8.21) respectively, where $(y_{ij})^{\text{best}}$ and $(y_{ij})^{\text{worst}}$ refer to the best and the worst attainable value of y_{ij}.

$$E_i = \sum_{j=1}^{n} w_j \frac{(y_{ij})^{best} - y_{ij}}{(y_{ij})^{best} - (y_{ij})^{worst}} \qquad \forall i \in I \tag{8.20}$$

$$F_i = Max^n \left\{ w_j \frac{(y_{ij})^{best} - (y_{ij})^{worst}}{(y_{ij})^{best} - (y_{ij})^{worst}} \right\} \qquad \forall i \in I \tag{8.21}$$

Note that E_i and F_i are used to compute the ranking index, R_i. Generally, the smaller value of R_i indicates the better ranking of the respective solution. It is defined as follows:

$$R_i = v \frac{E_i - {E_i}^{min}}{{E_i}^{max} - {E_i}^{min}} + (1 - v) \frac{F_i - {F_i}^{min}}{{F_i}^{max} - {F_i}^{min}} \qquad \forall i \in I \tag{8.22}$$

where v refers to the weight of the strategy of the "majority of criteria"; E_i^{min} and E_i^{max} refer to the minimum and maximum value of E_i respectively; while the maximum and minimum value of F_i are expressed as F_i^{min} and F_i^{max} respectively. Note that the value of v is taken as 0.5 in this work (it can be any value between 0 and 1).

8.3.2.5 Grey Relational Analysis (GRA)

Grey theory was first proposed by Deng (1982), where a mathematical methodology was invented to effectively analyze chaotic and incomplete systems. Later, the grey problem is defined as a problem where causing technology is unknown or unconfirmed (Offord, 2011). Grey problem is also commonly referred to problems that result in outcomes of the unknown exact value, but in a known range (Lin and Liu, 2006). Yang and Xue (2017) showed that grey theory is efficacious when the information is only partly known, or data samples are incomplete. The work also showed that appropriate estimation of the real world can be achieved by combining grey theory with other tools. For MCDM, a subfield of the grey theory, the Grey Relational Analysis (GRA) can be applied (Deng, 1989). Researchers such as Kuo et al. (2008), Hsu et al. (2015) and Hasani et al. (2012) had used GRA to solve MCDM problems in the manufacturing industry, construction industry and fabric industry, respectively. This work proposes that capacity debottlenecking of process plants with consideration of lean and green manufacturing is categorized under grey problem. This is evident, as this work had observed that data collected for debottlenecking studies are always innumerable and incomplete. This work then introduces two GRA-based approaches:

8.3.2.6 GRA-Based TOPSIS

Researchers have proven that GRA and TOPSIS standalone cannot accurately embody the approximation of the relationship between the performance of alternative solutions and the ideal solution (Wang and Peng, 2010). To address this issue, and combine the individual advantage of each method, a hybrid GRA-based TOPSIS method is introduced. The major differences of this hybrid method from conventional TOPSIS are the exclusion of need of weight input, whereas GRC is used to define the closeness between each feasible solution and the ideal solution (Ran and Wang, 2015).

First, the grey relational coefficient (GRC), ξ_{ij} is used to identify the closeness between the feasible solution and the ideal solution (for positive-ideal and negative-ideal solutions). Note

that ξ_{ij}^{+} is used to express the closeness between x_{ij} and x_{0j} (i.e., positive-ideal value), where x_{0j} refers to the best possible value for x_{ij} (i.e., smallest value for cost-criteria case; largest value for benefit-criteria case); $\Delta_{ij}^{+}{}^{max}$ and $\Delta_{ij}^{+}{}^{min}$ refer to the maximum and minimum values of Δ_{ij}^{+}; while ς is the distinguish coefficient which usually set as 0.5 (Zhang and Li, 2006). It is expressed as Equation (8.23), while the component Δ_{ij}^{+} is defined as Equation (8.24):

$$\xi_{ij}^{+} = \frac{\Delta_{ij}^{+}{}^{min} + \varsigma\, \Delta_{ij}^{+}{}^{max}}{\Delta_{ij}^{+} + \varsigma\, \Delta_{ij}^{+}{}^{max}} \qquad \forall i \in I, \forall j \in J \tag{8.23}$$

$$\Delta_{ij}^{+} = \left| x_{0j} - x_{ij} \right| \qquad \forall i \in I, \forall j \in J \tag{8.24}$$

Subsequently, the grey relational grade of each feasible solution *i*, γ_i^{+} is computed by the mean value of its accumulated positive-ideal grey relational coefficients (see Equation (8.25)).

$$\gamma_i^{+} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}^{+} \qquad \forall i \in I \tag{8.25}$$

The same procedure can be carried out to determine the negative-ideal values, i.e., ξ_{ij}^{-} and γ_i^{-} (see Equations (8.26-8.28)):

$$\xi_{ij}^{-} = \frac{\Delta_{ij}^{-}{}^{min} + \varsigma\, \Delta_{ij}^{-}{}^{max}}{\Delta_{ij}^{-} + \varsigma\, \Delta_{ij}^{-}{}^{max}} \qquad \forall i \in I, \forall j \in J \tag{8.26}$$

$$\Delta_{ij}^{-} = \left| x_{1j} - x_{ij} \right| \qquad \forall i \in I, \forall j \in J \tag{8.27}$$

$$\gamma_i^{-} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}^{-} \qquad \forall i \in I \tag{28}$$

where Δ_{ij}^{-} is the counterpart of Δ_{ij}^{+} in Equation (8.26) which represents the distance between x_{ij} and x_{1j} (x_{ij} is the worst possible value of x_{ij}); while $\Delta_{ij}^{-}{}^{min}$ and $\Delta_{ij}^{-}{}^{max}$ refers to the minimum and maximum value of Δ_{ij}^{-} respectively.

Then, the relative closeness in Equation (8.19) is redefined as Equation (8.29). Similarly, a solution with a higher value of C_i should be prioritized.

$$C_i = \frac{\gamma_i^-}{\gamma_i^+ + \gamma_i^-} \qquad \forall i \in I \tag{8.29}$$

Fundamentally, the order of the grey relational grades reflects the rank of each solution (higher values indicate the better in ranking).

8.3.2.7 GRA-Based VIKOR

GRA-based VIKOR was introduced by Kuo and Liang (2011) to reflect both subjective judgment and objective information when dealing with MCDM problems. Similar to the GRA-based TOPSIS method, the original measurement of "closeness" in normal VIKOR is now replaced by the GRC. A recent work claimed that GRA-based VIKOR can overcome the deficiencies of GRA-based TOPSIS which might fail to reflect the relative proximity of an integrated system (Sun et al., 2017).

In this method, E_i and F_i (from Equations (8.20) and (8.21)) are redefined to include the GRC values determined from Equation (8.21):

$$E_i = \sum_{j=1}^{n} \xi_{ij} \qquad \forall i \in I \tag{8.30}$$

$$F_i = Max^n\{\xi_{ij}\} \qquad \forall i \in I \tag{8.31}$$

Note that ξ_{ij} in Equations (8.30) and (8.31) refer to ξ_{ij}^+ (setting for conventional GRA method). Finally, Equation (8.22), the same equation which is used to determine the ranking index for conventional VIKOR, is reapplied here. Note that the better solutions are those which couple with smaller R_i values.

8.4 Case Study

In order to demonstrate the effectiveness of the proposed method, a real industrial case study from Pentas Flora Sdn. Bhd. (extended from Teng et al. (2019)) is employed in this study. Pentas Flora Sdn. Bhd. focuses on providing waste oil refining process to reduce waste oil disposal into the environment. Firstly, the waste oil is collected from various industry users in Peninsular Malaysia. Waste oils from different sources are blended prior to sending into a

series of the pretreatment process, i.e., flash tank dehydration unit remove moisture and light hydrocarbons from waste oil.

Next, the remaining oil is fed into a vacuum wiped film evaporator (WFE) which has high efficiency and minimal production degradation for further processing. From there, oil is separated into light lube oil (WFE product) and heavy oil. Following is a vacuum short path evaporator (SPE) that further separates the heavy oil into the asphalt and medium lube oil (SPE product) which is the main product of this waste oil refinery process. Figures 8.8 and 8.9 demonstrate the overall process flow in Pentas Flora Sdn. Bhd. To achieve higher lean and green attainments, a lean and green debottlenecking framework is therefore presented.

In the context of lean and green manufacturing, the use of the waste reduction (WAR) algorithm has been used in various lean manufacturing as an environmental assessment tool (WAR GUI, 2011). For example, D'Errico et al. (2009) have used global warming potential (GWP) to assess the environmental impact of magnesium products. In this case study, the relevant indicators from the WAR algorithm include global warming potential (GWP) and acidification potential (AP), which are evaluated by the following equation:

$$GWP = \frac{\sum_{a=1}^{Stream} \sum_{b=1}^{Component} F_a f_{b,a} \psi_{GWP,b}}{Feed\ Mass\ Flowrate} \tag{8.32}$$

$$AP = \frac{\sum_{a=1}^{Stream} \sum_{b=1}^{Component} F_a f_{b,a} \psi_{AP,b}}{Feed\ Mass\ Flowrate} \tag{8.33}$$

The assessment mainly considers the mass flow rate of emission stream which is denoted by F_a for stream a. Mass fraction of contaminant component b in the stream a is expressed as $f_{b,a}$, while specific potential environmental impact score for GWP and AP are expressed as $\psi_{GWP,k}$and $\psi_{AP,k}$ respectively.

Hegedic et al. (2018) also highlighted the importance of financial indicators, productivity, and product quality in lean and green manufacturing. Therefore, this work proposes the use of return on investment (ROI), product yield, and product quality as the respective lean indicators (note that the changes in variable costs refers to the additional net present value after debottlenecking activity is employed; while the investment costs refers to the additional capital

costs required for the purpose of debottlenecking). The mentioned indicators are calculated as the following:

$$ROI = \frac{Change\ in\ variable\ costs}{Investment\ costs} \tag{8.34}$$

$$Product\ Yield = \sum_{p=1}^{Products} \frac{Product\ i\ mass\ flow}{Total\ inlet\ mass\ flow} \tag{8.35}$$

$$Product\ Quality = \sum_{j=1}^{Indicator} f(q_j) \tag{8.36}$$

$$f(q_j) = \begin{cases} \frac{q_j - {q_j}^{low}}{{q_j}^{max} - {q_j}^{low}}, if\ q_i\ is\ desired\ to\ be\ high \\ \frac{q_j - {q_j}^{ideal}}{{q_j}^{ideal}}, if\ q_i\ is\ desired\ to\ be\ at\ ideal\ value \\ \frac{{q_j}^{high} - q_j}{{q_j}^{high} - {q_j}^{min}}, if\ q_i\ is\ desired\ to\ be\ low \end{cases} \tag{8.37}$$

For the product quality calculations, q_j is the j^{th} quality indicator, ${q_j}^{low}$ is the low control limit, ${q_j}^{high}$ is the high control limit, ${q_j}^{ideal}$ is the ideal control value, ${q_j}^{max}$ is the historical-high value, and ${q_j}^{min}$ is the historical-low value. For this case study, the only quality indicator is the kinematic viscosity of the lube oil product, which is desired to be high. Thus, the first equation in the set of Equation (8.29) is used.

All the above environmental impact potentials and lean indicators can be calculated based on the process simulation results in the Aspen HYSYS (see Figure 8.8), the model is adapted from Teng et al. (2019).

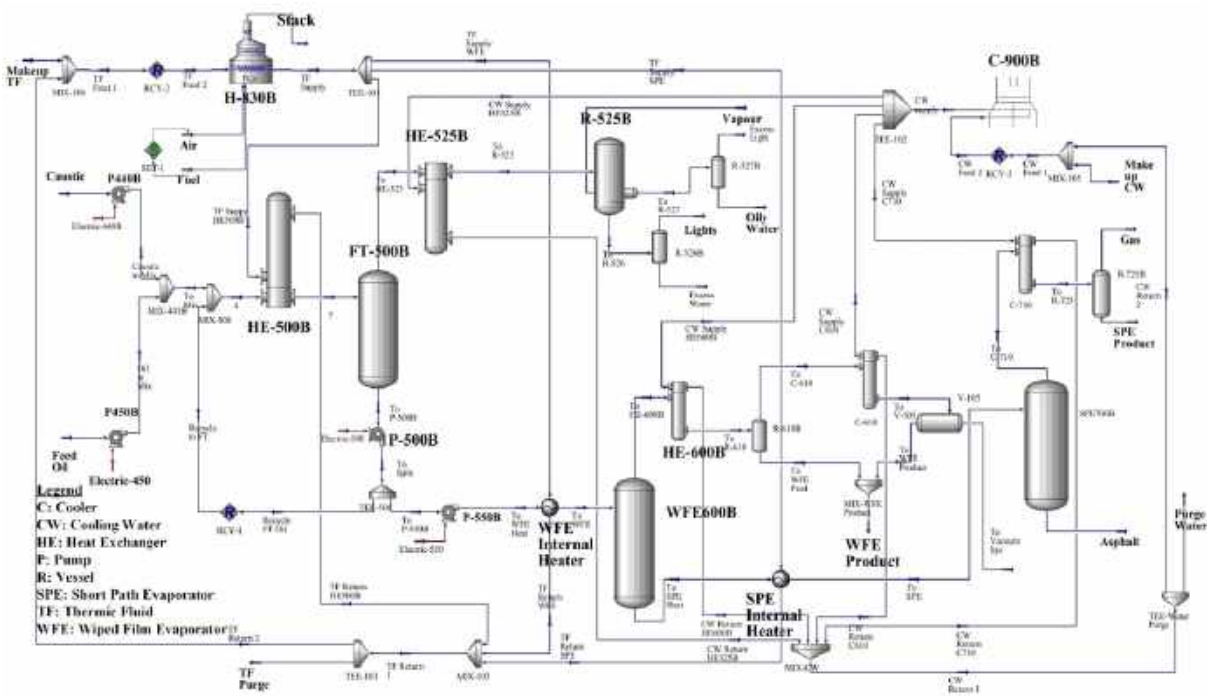


Figure 8.8: Process flow diagram of simulation in Aspen Hysys (adapted from Teng et al., 2019)

8.5 Results and Discussion

The subsequent chapters will cover the result and discussion on implementing the full debottlenecking framework on the real industrial case study in Pentas Flora Sdn. Bhd., Malaysia. Chapter 8.5.1 discusses the utilization of Bottleneck Tree Analysis (BOTA), Chapter 8.5.2 discusses the evaluation of Green and Lean Index (GLI) and obtaining the bottleneck path. Lastly, Chapter 8.5.3 discusses the implementation of the project by scheduling different project stages.

8.5.1 Bottleneck Tree Analysis (BOTA)

By evaluating the maximum design capacity of each equipment, multiple levels of the capacity bottleneck within the process are being identified. In this case study, the oil re-refinery process is decoupled into two regions. The first region contains the flash tank, while the second region contains the Wiped Film Evaporator (WFE) and Short Path Evaporator (SPE). The two regions are separated by a capacity-controlled valve with a recycle loop, which splits the process into two non-capacity interrelated regions. The debottleneck regions of the process are shown in the process flow diagram below (Figure 8.9).

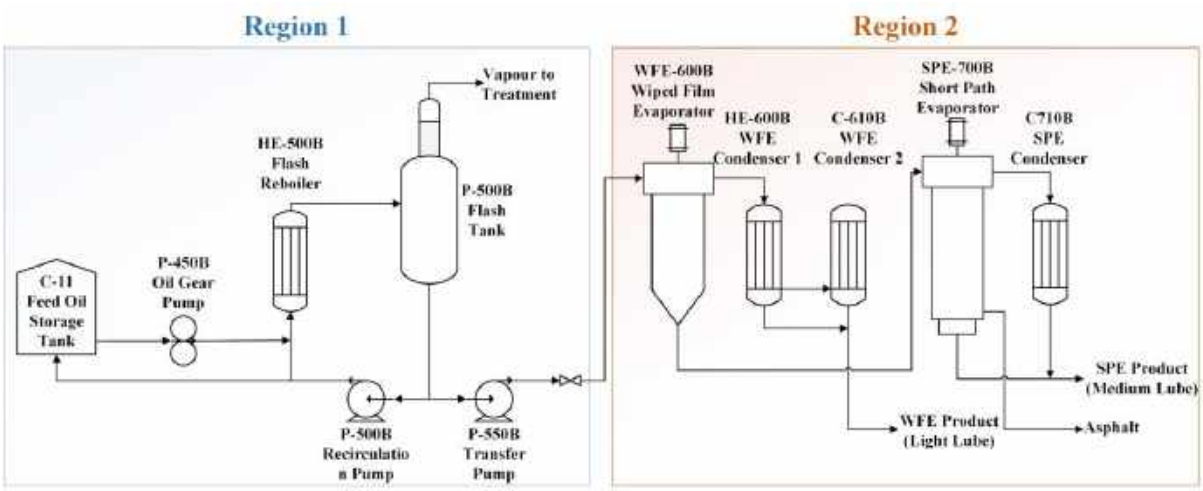


Figure 8.9: Process flow diagram and the region of study

The utilization factor is used to identify debottlenecking opportunities clearly and systematically. Initially, region 1 has an inlet flow of 70 lpm. This capacity is gradually increased by 5 lpm per interval until the utilization factor for any single equipment is equal to 1. From Figure 8.10(a), the first unit to achieve a utilization factor of 1 is the oil gear pump (P-450), hence this unit is the first stage of the bottleneck. Increasing the capacity of P-450 can immediately improve the capacity of region 1 in the plant. Next, the inlet flow of region 1 is increased to identify the next units that achieve a utilization factor of 1. Subsequently, the units are the cooling tower (C-900B & C-910B), flash tank (FT-500B) and centrifugal pump (P-500B). Note that this analysis is performed with the aid of the process simulation in Aspen Hysys (see Figure 8.8). The identical procedure was also undergone for region 2, in which the bottlenecking units were identified as Wiped Film Evaporator (WFE-600B) and Short Path Evaporator (SPE-700B) sequentially (see Figure 8.10(b)).

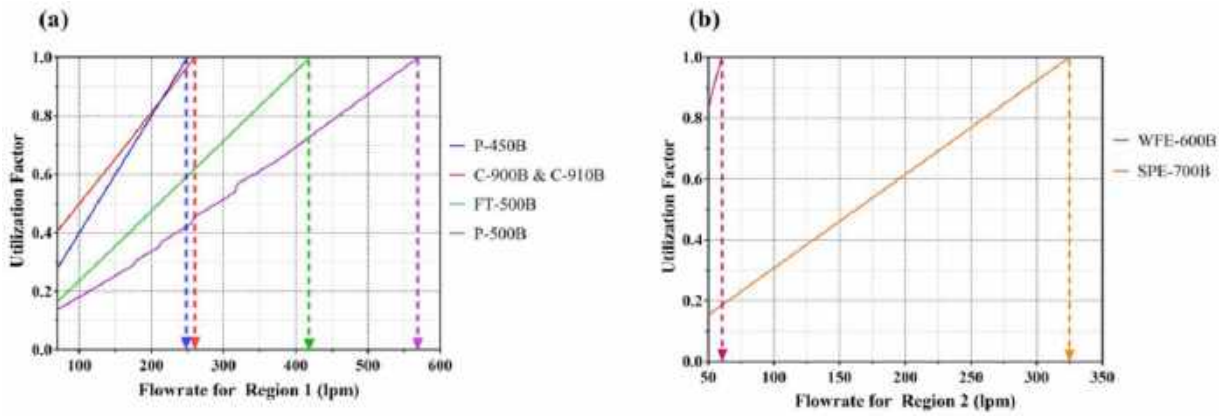


Figure 8.10: Utilization factor against inlet flow rate (lpm) for (a) region 1; (b) region 2

The debottleneck tree diagram for the flash tank system (Region 1) is shown in Figure 8.11. The process is currently operating at 70 lpm capacity with feed pump (P-450) at 2.2 kW capacity, Flash Tank (FT-500B) at 420 lpm capacity and transfer pump P-500B at 42.61 kW capacity. From Figure 8.10, the first bottleneck (stage 1 in Figure 11) for the processing capacity is at a feed rate of 250 lpm, which the feed pump needs to be increased to 2.3 kW. At the second bottleneck (stage 2 in Figure 11), which is 260 lpm feed rate, the redundant cooling tower C-900B and C-910B will be required to be turned on, while the feed pump P-450B needs to increase to 3.72 kW. The process can be increased to 430 lpm by further increasing the flash tank and subsequently to 570 lpm by increasing pump P-450B to 5.09 kW.

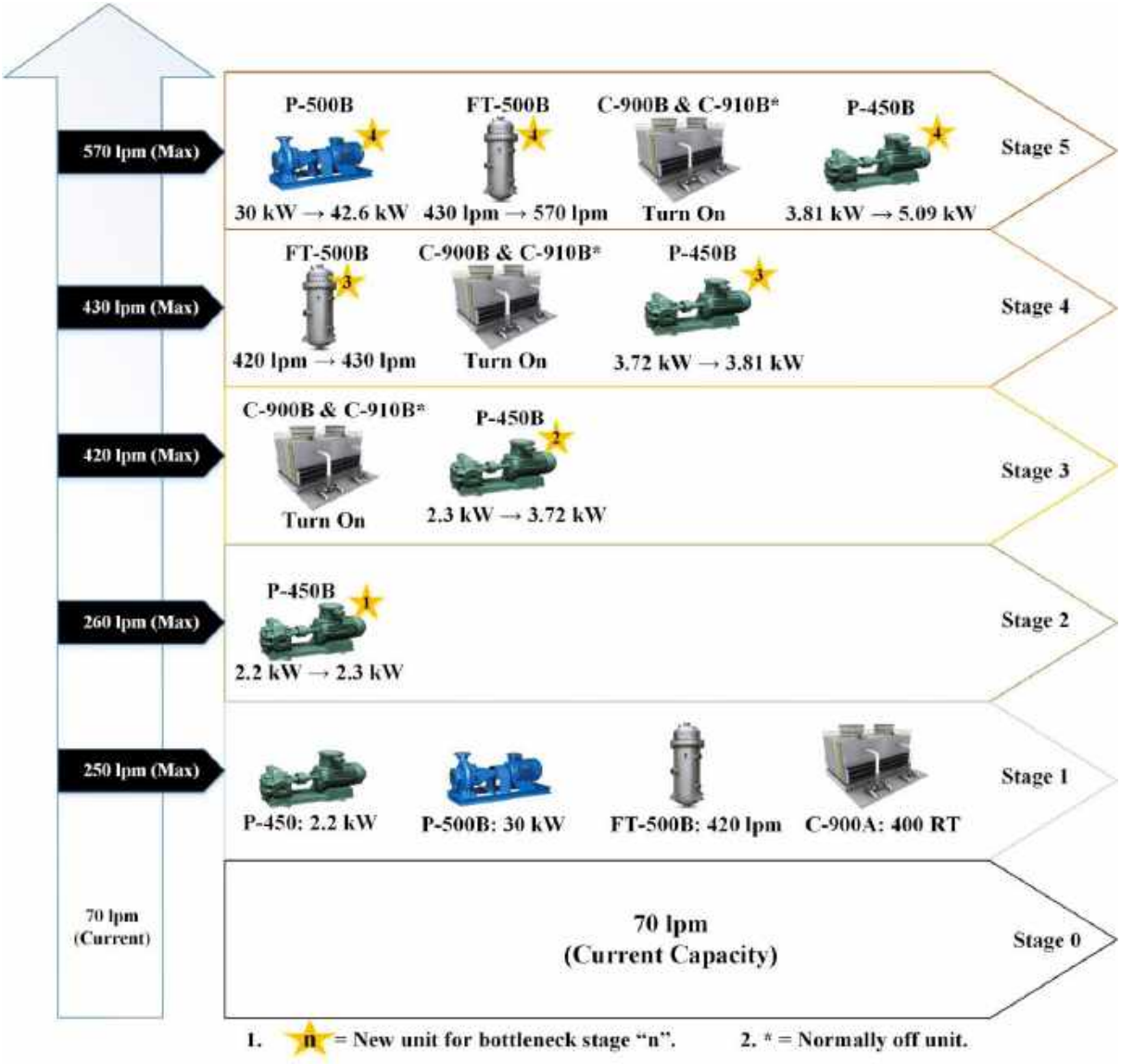


Figure 8.11: Bottleneck tree diagram for region 1

The process debottlenecking tree diagram for the second region of the process is shown in Figure 8.12. The process is debottlenecked by one of the major units, which is the wiped film evaporator (WFE-700B). Under the current conditions, the process can increase the capacity to 60 lpm. Next, to achieve 325 lpm capacity, wiped film evaporator WFE-700B has to be increased to 325 lpm and Short Path Evaporator (SPE-800) capacity to 111.7 lpm.

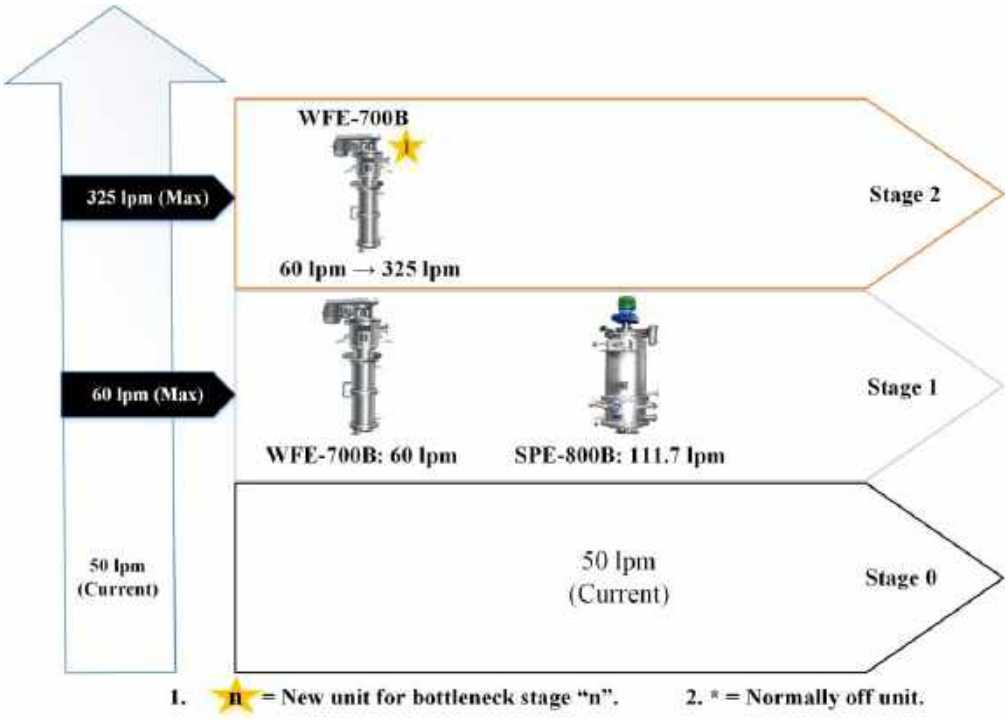


Figure 8.12: Bottleneck tree diagram for region 2

8.5.2 Green and Lean Index (GLI) Evaluation and Bottleneck Pathway

Figure 8.13 shows all the feasible solutions which integrate the design configuration for both regions. In total, there are 15 feasible solutions (i =1, 2, 3, …, 15) for this case study. Note that, based on the process flow diagram shown in Figure 8.9, since the Region 2 is the successive stage of Region 1, the flow capacity in Region 1 has to be greater than that in Region 2 (otherwise, it will not comply with the laws of conservation of mass). Thus, all the infeasible solutions (i.e., flow capacity in Region 1 is lesser than that in Region 2) are excluded from the analysis ("-" sign in Figure 8.13 indicates that this configuration is infeasible). The performance of each solution (in terms of GWP, AP, energy, ROI, product yield and quality) is simulated through the Aspen Hysys model. The obtained results are then normalized and tabulated in Table 8.3. The results are then used to identify the GLI via four different MCDM methods (using the Equations formulated in Chapter 8.3.2) in Visual Basic (Excel). The

determined GLI (i.e., the final scores obtained from each method) for each solution is listed in Table 8.4. Take note that the capacities for the regions are not the responding variables, but the condition on the bottleneck tree matrix (Figure 8.13).

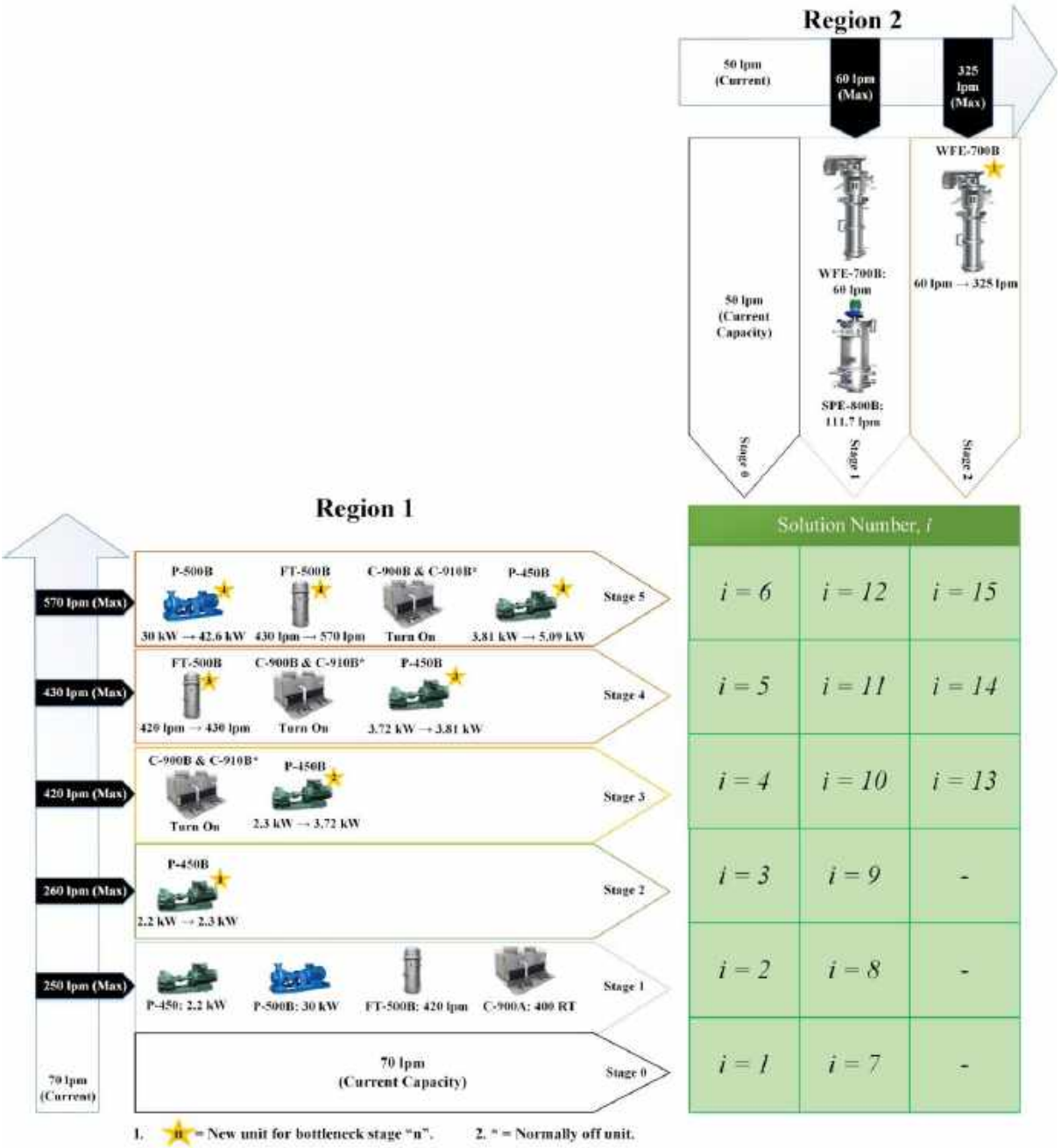


Figure 8.13: Feasible solutions and its configuration

Table 8.3: Decision matrix for region capacities and their responding variables

Solution Number	Region 1 Max Capacity (lpm)	Region 2 Max Capacity (lpm)	Norm. Global Warming Potential (%)	Norm. Acidification Potential (%)	Norm. Energy (%)	Norm. Return on Investment (%)	Norm. Product Yield (%)	Norm. Product Quality (%)
1	70	50	99.99	28.23	98.87	0.00	59.41	99.82
2	250	50	19.24	90.80	19.12	33.00	16.82	100.00
3	260	50	16.70	90.75	16.58	32.47	15.63	100.00
4	420	50	5.56	65.85	5.59	29.68	4.42	99.46
5	430	50	5.10	65.83	5.13	29.51	3.98	99.87
6	570	50	0.00	60.14	0.00	21.94	0.00	96.14
7	70	60	100.00	68.60	100.00	39.63	100.00	99.82
8	250	60	19.24	99.78	19.54	58.36	24.06	100.00
9	260	60	16.71	100.00	17.00	57.28	22.33	100.00
10	420	60	5.56	65.85	5.78	54.84	8.73	99.46
11	430	60	5.10	65.83	5.31	51.45	7.66	99.87
12	570	60	8.04	60.39	0.16	67.00	5.48	98.19
13	420	325	14.04	0.00	5.33	93.17	16.09	41.67
14	430	325	13.43	0.03	4.80	100.00	16.60	41.67
15	570	325	7.85	1.26	0.00	80.68	8.24	0.00

Table 8.4: Scores of different MCDM methods

Solution Number	Final Score for Ranking			
	TOPSIS	VIKOR*	GRA-TOPSIS	GRA-VIKOR*
1	0.4568	0.1759	0.4329	0.1679
2	0.5475	0.5618	0.5225	0.1535
3	0.5507	0.5718	0.5258	0.1720
4	0.5850	0.6610	0.5568	0.7931
5	0.5854	0.6573	0.5577	0.4206
6	0.5836	0.6067	0.5574	0.3401
7	0.4827	0.2430	0.4796	0.1291
8	0.5660	0.4859	0.5350	0.1959
9	0.5676	0.4887	0.5370	0.2117
10	0.6152	0.7876	0.5814	0.8366
11	0.6115	0.7664	0.5792	0.4604
12	0.6296	0.7830	0.5921	0.5344

Table 8.4 (Cont.): Scores of different MCDM methods

Solution Number	Final Score for Ranking TOPSIS	Solution Number VIKOR*	Final Score for Ranking TOPSIS	Solution Number GRA-VIKOR*
13	0.6436	0.9732	0.6199	0.4490
14	0.6490	1.0000	0.6292	0.5000
15	0.5591	0.3345	0.5656	0.3969

*Value obtained from 1-R_i to synchronize the trend (higher value indicates higher viability)

All the MCDM methods exhibit similar properties during debottlenecking stages. Firstly, solutions with no major changes to the process units are poorly ranked (ranked 13th and below; see blue dotted circles in Figure 8.14) by the MCDM algorithm. The process shown in these solutions is still being underutilized. This type of solution corresponds to the outermost layer of the reversed onion model. Next, by subsequently moving up into higher stages of bottleneck tree, the rank is gradually improved (i.e., achieve higher ranking). This situation sustains until a rank reversal point is reached (see green dotted circles in Figure 8.14). The reversal points serve as "stopping points" for the debottlenecking work, as further debottlenecking will no longer beneficial in terms of lean and green sustainability. This can be observed after the "stopping points", which the ranks of the solutions gradually worsen (see black dotted circles in Figure 8.14). This situation corresponds to the inner "separation and recycle" layer of the reversed onion model, where heuristically, debottlenecking in this layer (i.e., replacement of separation unit, recycle system and reactors) is not preferable.

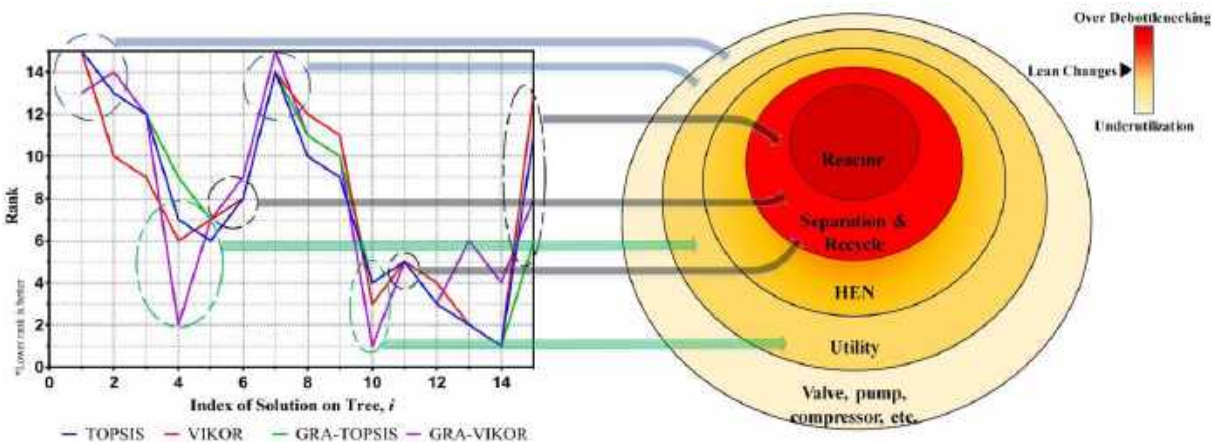


Figure 8.14: The correspondence of MCDM ranks towards reversed onion model

Table 8.5: Spearman's correlation coefficient of MCDM methods

Methods	Spearman's Correlation Coefficient			
	TOPSIS	VIKOR	GRA-TOPSIS	GRA-VIKOR
TOPSIS	1.0000	0.9393	0.9429	0.8607
VIKOR	0.9393	1.0000	0.8571	0.8143
GRA-TOPSIS	0.9429	0.8571	1.0000	0.8321
GRA-VIKOR	0.8607	0.8143	0.8321	1.0000

As seen in Table 8.5, the Spearman's correlations for all methods are above 0.8. This converts to a high 99.99% certainty (P-value<0.0001) for all MCDM methods using *t*-distribution (McDonald, 2014). Thus, this shows that regardless of which method is applied, consistent and reliable trends of solution ranks can be obtained. This further validates the concept of the suggested reversed onion hierarchy for debottlenecking. Subsequently, the resulting Green and Lean Index (GLI) can be formulated from all the MCDM methods in Table 8.5.

To demonstrate how the obtained GLI helps in debottlenecking decisions, the results generated from GRA-based VIKOR is used as an illustrative example. The obtained GLI for each feasible solution shown in Figure 8.15, is used to demonstrate the viability of each debottlenecking stage. Note that the GRI in this example is adapted from the value of 1-R_i (higher value indicates higher viability) determined from GRA-based VIKOR method. Concisely, the ranking scores for GRA-VIKOR in Table 8.4 were filled into the bottleneck tree matrix (see Figure 8.13) as shown in Figure 8.15.

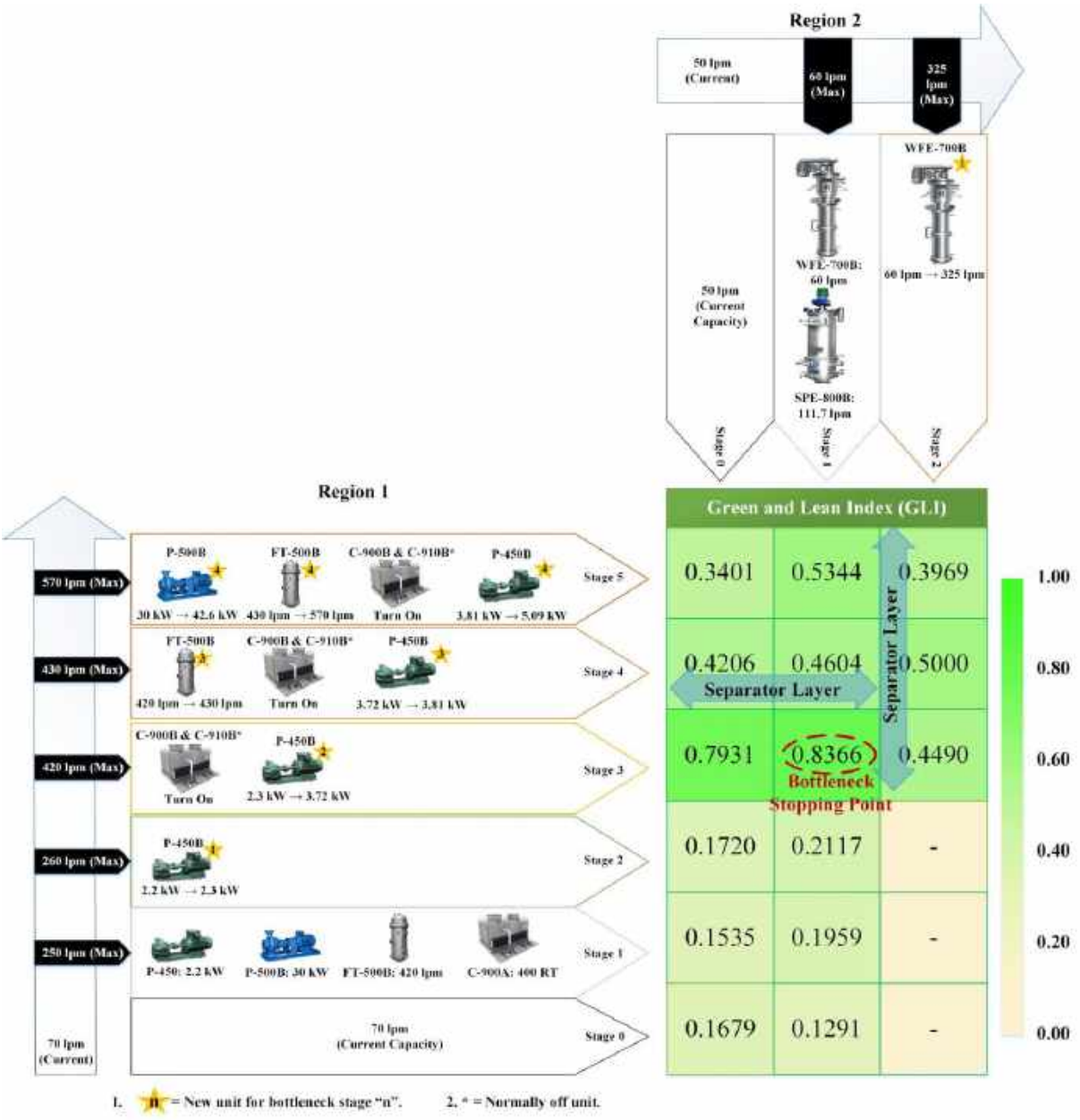


Figure 8.15: Bottleneck tree matrix for Green and Lean Index (GLI)

By analysing the first column in Figure 8.15 (compare GLI vertically), the GLI increased gradually with debottlenecking stages in Region 1; and finally achieved maximum point at stage 2 (i.e., 0.7931). Any further debottlenecking which involves changes in the core layers of the onion model (i.e., separation, recycling, and reactor systems) will lead to the dropping of GLI value. This agrees with the concept of reversed onion hierarchy which prioritizes the debottlenecking at outer layer compared to those in the inner core. A similar trend can be observed for the debottlenecking in region 2. For instance, assumed region 1 had been debottlenecked up to the second stage, the overall green and lean performance is improved to 0.8366 when higher flow capacity is allowed in Region 2 (compare GLI horizontally).

Similarly, GLI has declined when the second stage of debottlenecking which involves the replacement of the flash tank, is implemented.

Aside from this, for the first column of region 1, when the capacity increases from the current capacity to stage 0, the GLI drops from 0.1679 to 0.1535. This situation, although the capacity in region 1 increases, the capacity in region 2 limits the product yield. Hence, more resources are being used to produce the same amount of product. As a counterexample, in the second column of region 1, from the current capacity to stage 0, the GLI improves from 0.1291 to 0.1959. This shows that simultaneous increasing the capacity in both region 1 and region 2 to stage 0 is beneficial for the process, as the product flow rate is not limited by the capacity of region 2. An interesting evaluation from GLI arises in the second row of region 2 when the capacity of region 2 increases from 0.4604 to 0.5000. In solution i = 11 (with GLI of 0.4604), an expensive flash tank was already purchased, however, the process is once again limited by the capacity of region 2. Thus, the GLI recommends investing more capital to increase the capacity of Wiped Film Evaporator (WFE-700B) in order to enable a higher product yield. Nevertheless, following this framework, the debottlenecking work will never proceed to this stage because the cost of the flash tank is not justified for the sustainability of the process plant.

Based on the 5 rules for constructing bottleneck path in Chapter 8.3.1 (rules are also stated in Figure 8.16), the bottleneck pathway is constructed and shown in Figure 8.16. Authors would like to point out some other pathways that are eliminated by the 5 rules, however, may be not obvious to the reader. Firstly, cells *i=13* to *15* are labelled with "-" since the process violates the conservation of mass at these states (refer to Figure 9 for understanding of process), making it impossible. The path from the original operation cell (*i=1*) cannot move to cell with *i=2* or *7* due to reversal in the GLI values. The pathway of i = [1,4,10] is valid, however, it is not the longest path, as only three steps were taken. This leaves the possible paths of *i = [1,3,9,10]* and *i = [1,8,9,10]* with both having 4 steps to reach the BSP (cell *i=10*). By considering rule 5: If multiple pathways exist, take the path that passes through the cell with stage 1 in all regions (cell *i=8* in this case). Therefore, only one bottleneck path remains with *i = [1,8,9,10]* and is highlighted with yellow in Figure 8.16.

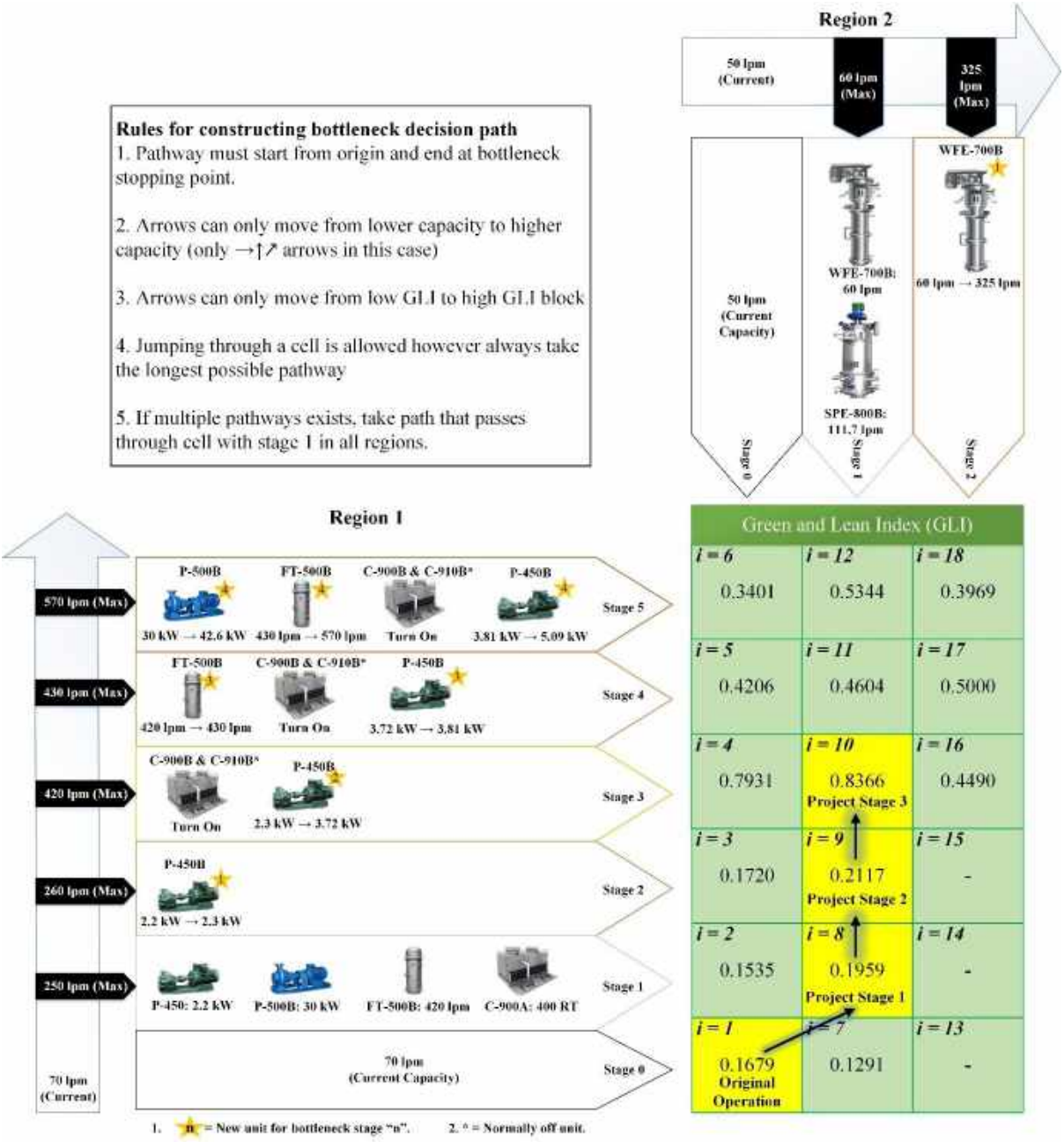


Figure 8.16: Bottleneck decision path based on GLI

8.5.3 Project Implementation by Scheduling

Another advantage of utilizing the BOTA and GLI approach for debottlenecking is the implementation project can be scheduled into multiple stages. This is shown in Figure 8.16, where the bottleneck decision path is constructed. Thus, the project can be scheduled into 3 stages, where each stage significantly improves the capacity and sustainability of the plant. Using the rules for constructing the bottleneck decision path, the debottlenecking schedule is guaranteed to not create contradicting process changes, thus maximizing economic returns.

However, the bottleneck decision pathway only serves as a guide for decision-makers in developing debottlenecking strategies. In other words, the decision-maker is recommended to perform a complete debottlenecking project by investing the full amount of capital investment as soon as possible to collect maximum returns. However, if there is capital investment shortage, the decision-makers can schedule the debottlenecking project with the aid of the bottleneck decision path. Thus, the decision-maker can gradually invest fractional sums of total capital cost with time, to gain returns at a slower rate. This is a critical benefit as the initial investment costs can be decreased dramatically.

As an example, the processing plant is assumed to have financial problems in investing any of the capital costs upfronts. Thus, the process plant company only pays for the debottlenecking project from its future net profit. By using the BOTA-GLI approach for project scheduling, the debottlenecking project is economically feasible and turn-over period can be shortened. Table 8.6 tabulates all four possible debottlenecking strategies (sub-pathways of full bottleneck path) which aim to improve the process GLI from 0.1679 to 0.8366, while the corresponding net present value performance is illustrated in Figure 8.17.

Table 8.6: Debottlenecking project scheduling strategies

Strategies	Description	Project Stage 1	Project Stage 2	Project Stage 3
S0	Without debottlenecking	-	-	-
S1	One stage	$i=1 \rightarrow i=10$	-	-
S2(a)	Two stage (a)	$i=1 \rightarrow i=8$	$i=8 \rightarrow i=10$	-
S2(b)	Two Stage (b)	$i=1 \rightarrow i=9$	$i=9 \rightarrow i=10$	-
S3	Three stage	$i=1 \rightarrow i=8$	$i=8 \rightarrow i=9$	$i=9 \rightarrow i=10$

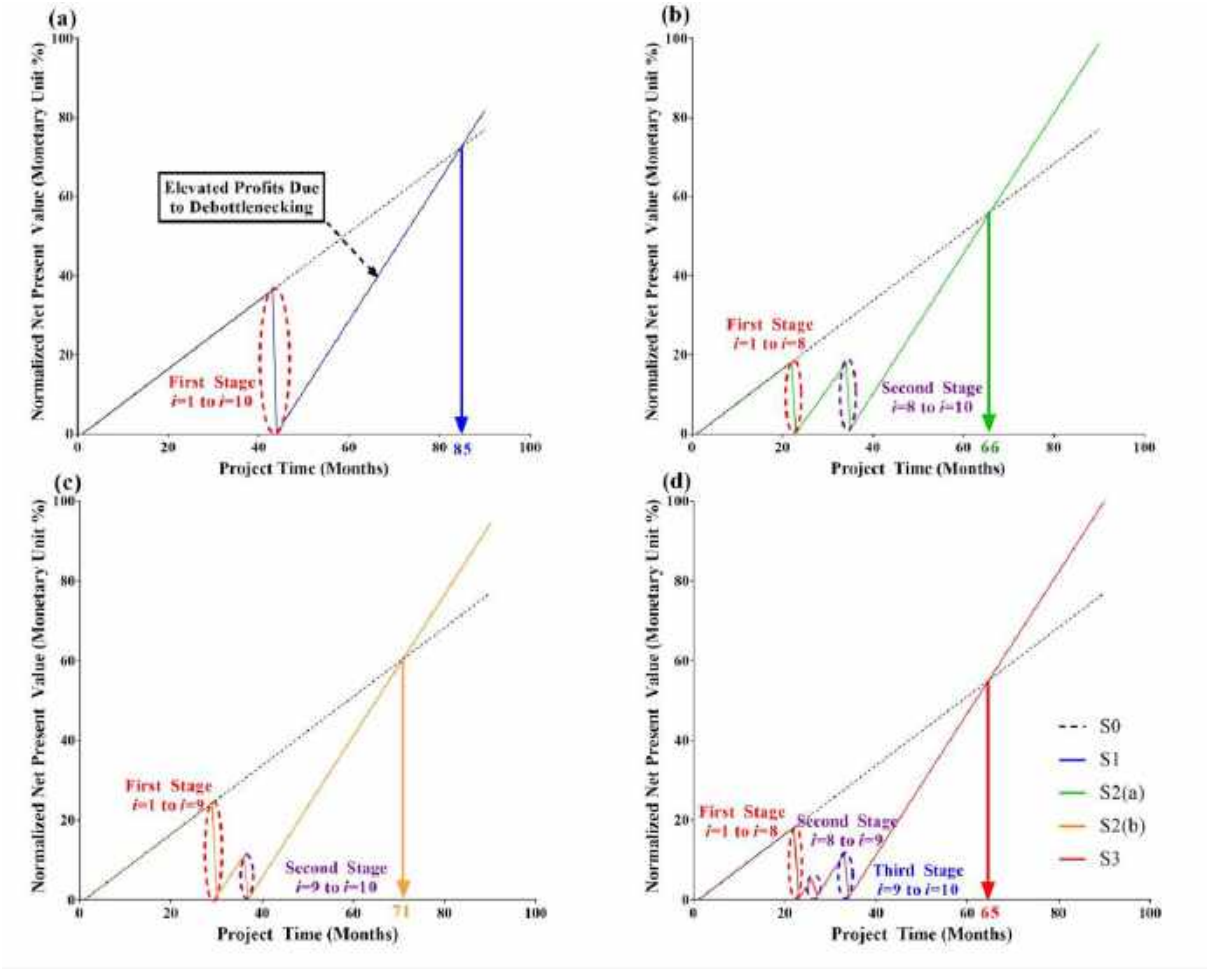


Figure 8.17: Normalized net present value over project time of strategy (a) S1; (b) S2a; (c) S2b; (d) S3 (for zero initial investment problems)

From Figure 8.17, it is shown that in such cases, the elevated net profit that is resulting from the three-step scheduling of the project is the best and will result in 65 months of turnover period. Two-step scheduling (a) and (b) have reduced the turnover period from 85 months to 66 and 71 months, respectively. Scheduling with the BOTA-GLI framework enables a faster turnover period for debottlenecking projects, which will result in a higher net present value after the project is completed. In effect, smaller fractionated capital costs can be invested to gain monetary and sustainability benefits from the process before investing the full sum.

8.6 Conclusion

Onion model has been widely used in the process design hierarchy. Contrarily, this work suggests that the direction of the hierarchy may be reversed during the debottlenecking process. Furthermore, this work presents a novel debottlenecking technique, called Bottleneck Tree

Analysis (BOTA) to identify debottlenecking opportunities according to the utilization factors of each unit.

The lean and green indicators (i.e., product yield, product quality, global warming potential (GWP), acidification potential (AP), energy and return on investment (ROI)) of each debottlenecking alternative are evaluated. Using this information, with the aid of multiple criteria decision-making (MCDM) methods (including TOPSIS, VIKOR, GRA-based TOPSIS and GRA-based VIKOR), the overall green and lean index (GLI) are formulated. From the illustrative case study, all methods are found to give certainty of above 99.99% (p-value <0.0001). This further validates the concept of the reversed onion model for debottlenecking processes which aims to avoid the involvement of inner layers (separation, recycle and reactor) in debottlenecking works.

The identified "stopping point" for the oil re-refining process debottlenecking provides GLI of 0.8366, which is at 420 lpm and 60 lpm of max capacity at region 1 and region 2 respectively (i=10 in Figure 16). For this solution, the gear oil pump capacity must be increased while backup cooling towers must be turned on. GLI is also found to be indicative of debottlenecking paths while reflecting on the green and lean sustainability factor of the plant. This is shown to be particularly useful in scheduling for debottlenecking projects, which has reduced the payback period from 85 months to 65 months. Future extension on formulating optimum scheduling strategies under different financial situations can be carried out. Although this debottlenecking framework guarantees to systematically debottleneck a process, the fluctuation in product demands within the product supply chain can potentially cause a management decision to slow down production.

Nomenclature

Abbreviation	Definition
AP	Acidification Potential
BOTA	Bottleneck Tree Analysis
FT	Flash Tank
GLI	Green and Lean Index
GM	Green Manufacturing
GRA	Grey Relational Analysis

GWP	Global Warming Potential
LM	Lean Manufacturing
MCDM	Multiple Criteria Decision-Making
MINLP	Mixed Integer Non-linear Programming
PSE	Process System Engineering
ROI	Return on Investment
SPE	Short Path Evaporator
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VIKOR	Višekriterijumsko Kompromisno Rangiranje
WFE	Wiped Film Evaporator

Reference

Ahmad, A., Goransson, M., Shahzad, A., 2010. Limitations of the Analytic Hierarchy Process Technique with Respect to Geographically Distributed Stakeholders. In Proceedings of World Academy of Science, Engineering and Technology, 4(9), 97-102.

Ahmad, S., Hui, D., 1991. Heat recovery between areas of integrity. Computers & Chemical Engineering, 15(12), 809-832.

Alshekhli, O., Foo, D.C.Y., Hii, C.L., Law, C.L., 2011. Process simulation and debottlenecking for an industrial cocoa manufacturing process. Food and Bioproducts Processing, 89(4), 528-536.

Andiappan, V., Ng, D.K., Tan, R.R., 2017. Design Operability and Retrofit Analysis (DORA) framework for energy systems. Energy, 134, 1038-1052.

Bagajewicz, M., Rodera, H., 2002. Multiple plant heat integration in a total site. AIChE Journal, 48(10), 2255-2270.

Bey, N., Hauschild, M.Z., McAloone, T.C., 2013. Drivers and barriers for implementation of environmental strategies in manufacturing companies. *CIRP Annals*, *62*(1), pp.43-46.

Carr, D., Shearer, J., 2007. Nonlinear Control and Decision Making Using Fuzzy Logic in Logix. Rockwell Automation, Inc.

Carvalho, H., Machado, V.H., Barroso, A.P., de Almeida, D. and Cruz-Machado, V., 2019. Using Lean and Green Indexes to Measure Companies' Performance. In *Lean Engineering for Global Development* (pp. 293-318). Springer, Cham.

Čuček, L., Hjaila, K., Klemes, J., Kravanja, Z., 2017. Onion Diagram Implementation to the Synthesis of a Biogas Production Network. Chemical Engineering Transactions, 61, 1687-1692.

D'Errico, F., Perricone, G., Oppio, R., 2009. A new integrated lean manufacturing model for magnesium products. JOM, 61(4), 14-18.

de Haan, J., Yamamoto, M., Lovink, G., 2001. Production planning in Japan: rediscovering lost experiences or new insights?. *International Journal of Production Economics*, *71*(1-3), pp.101-109.

Deng, J.L., 1982. "Control problems of grey systems", Systems and Control Letters, 5, 288-94.

Deng, J.L., 1989. "Introduction to grey system theory", The Journal of Grey System, 1, 1-24.

Diaz, S., Serrani, A., Beistegui, R., Brignole, E., 1995. A MINLP strategy for the debottlenecking problem in an ethane extraction plant. Computers & Chemical Engineering, 19, 175-180.

Dieste, M., Panizzolo, R., Garza-Reyes, J.A., Anosike, A., 2019. The relationship between lean and environmental performance: Practices and measures. *Journal of Cleaner Production*, *224*, pp.120-131.

Douglas, J., 1985. A hierarchical decision procedure for process synthesis. AIChE Journal, 31(3), 353-362.

Dües, C.M., Tan, K.H., Lim, M., 2013. Green as the new Lean: how to use Lean practices as a catalyst to greening your supply chain. *Journal of cleaner production*, *40*, pp.93-100.

Dunn, R., El-Halwagi, M., 2003. Process integration technology review: background and applications in the chemical process industry. Journal of Chemical Technology & Biotechnology, 78(9), 1011-1021.

Gai, H., Jiang, Y., Qian, Y., Zhang, L., Yang, C., 2018. Analysis and debottlenecking retrofits for the coal-gasification wastewater treatment process. Huaxue Gongcheng, [online] 35(8), pp.57-60. Available at: jglobal.jst.go.jp/en/public/20090422/201102202172882341 [Accessed 6 Jul. 2018].

Hasani, H., Akhavan Tabatabaei, S., Amiri, G., 2012. Grey Relational Analysis to Determine the Optimum process parameters for Open-end Spinning Yarns. Journal of Engineered Fibers and Fabrics, [online] 7(2), 81-86. Available at: www.jeffjournal.org/papers/Volume7/7.2.11H.Hasani.pdf [Accessed 21 Jul. 2018].

Hegedic, M., Gudlin, M., Stefanic, N., 2018. Interrelation of Lean and Green Management in Croatian Manufacturing Companies. Interdisciplinary Description of Complex Systems, 16(1), 21-39.

Heitmann, M., Schembecker, G., Bramsiepe, C., 2017. Framework to decide for a volume flexible chemical plant during early phases of plant design. Chemical Engineering Research and Design, 128, 85-95.

How, B.S., Yeoh, T.T., Tan, T.K., Chong, K.H., Ganga, D.P., Lam, H.L., 2018a. Debottlenecking of sustainability performance for integrated biomass supply chain: P-graph approach. Journal of Cleaner Production, 193,720-733.

How, B.S., Lam, H.L., 2018b. PCA Method for Debottlenecking of Sustainability Performance in Integrated Biomass Supply Chain. Process Integration and Optimization for Sustainability. doi.org/10.1007/s41660-018-0036-3

Hsu, H., Cheng, A., Chao, S., Chang, J., Teng, L., Chen, S., 2015. The Grey Relational Analysis of Quality Investigation of Concrete Containing Solar PV Cells. MATEC Web of Conferences, 27, p.01006.

Hwang, C.L., Yoon, K.P., 1981. Multiple attributes decision making methods and applications. Springer-Verlag, Berlin.

Joint Research Centre-European Commission, 2008). *Handbook on constructing composite indicators: methodology and user guide*. OECD publishing.

Kasivisvanathan H., Tan R.R., Ng, D.K.S., Abdul Aziz, M.K., 2014. Heuristic framework for the debottlenecking of a palm oil-based integrated biorefinery. Chemical Engineering Research and Design, 92(11), 2071-2082.

Kemp I.C., 2007. Pinch analysis and process integration: A user guide on process integration for the efficient use of energy, vol. 2. Butterworth-Heinemann, Oxford, UK.

King, A.A., Lenox, M.J., 2001. Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production and operations management*, *10*(3), 244-256.

Koulouris, A., Calandranis, J., Petrides, D., 2000. Throughput analysis and debottlenecking of integrated batch chemical processes. Computers & Chemical Engineering, 24(2-7), 1387-1394.

Kuo, M., Liang, G., 2011. Combining VIKOR with GRA techniques to evaluate service quality of airports under fuzzy environment. Expert Systems with Applications, 38(3), 1304-1312.

Kuo, Y., Yang, T., Huang, G., 2008. The use of grey relational analysis in solving multiple attribute decision-making problems. Computers & Industrial Engineering, 55(1), 80-93.

Lee, J., Khan, R., Phelps, D., 2008. Debottlenecking and CFD Studies of High-and Low-Pressure Production Separators. SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA.

Leong, W.D., Lam, H.L., Ng, W.P.Q., Lim, C.H., Tan, C.P., Ponnambalam, S.G., 2019a. Lean and Green Manufacturing—a Review on its Applications and Impacts. *Process Integration and Optimization for Sustainability*, *3*(1), 5-23.

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P., Ponnambalam, S.G., 2019b. Adaptive Analytical Approach to Lean and Green Operations. *Journal of Cleaner Production*, 235, 190-209.

Lever, J., Krzywinski, M., Altman, N., 2017. Points of Significance: Principal component analysis. Nature Methods, 14(7), 641-642.

Li, X., Wang, K., Liu, L., Xin, J., Yang, H., Gao, C., 2011. Application of the Entropy Weight and TOPSIS Method in Safety Evaluation of Coal Mines. Procedia Engineering, 26, 2085-2091.

Liang, X., Kang, L., Liu, Y., 2016. The flexible design for optimization and debottlenecking of multiperiod hydrogen networks. Industrial & Engineering Chemistry Research, 55(9), 2574-2583.

Liliana, L., 2016. November. A new model of Ishikawa diagram for quality assessment. In *IOP Conference Series: Materials Science and Engineering* (Vol. 161, No. 1, p. 012099). IOP Publishing.

Lin, Y., Liu, S., 2006. Solving Problems with Incomplete Information: A Grey Systems Approach. Advances in Imaging and Electron Physics, 77-174.

Linnhoff B., Townsend D.W., Boland P., Hewitt G.F., Thomas B.E.A., Guy A.R., Marsland R.H., 1982. User Guide on Process Integration for the Efficient Use of Energy. IChemE, Rugby, UK.

Litzen, D.B., Bravo, J.L., 1999. Uncover low-cost debottlenecking opportunities. Chemical Engineering Progress, 95(3), 25-32.

Mardani, A., Jusoh, A., Nor, K., Khalifah, Z., Zakwan, N., Valipour, A., 2015. Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja*, *28*(1), 516-571.

Matsuda, K., Hirochi, Y., Tatsumi, H., Shire, T., 2009. Applying heat integration total site based pinch technology to a large industrial area in Japan to further improve performance of highly efficient process plants. Energy, 34(10), 1687-1692.

McDonald, J. H., 2014. Handbook of Biological Statistics. 3rd ed.Sparky House Publishing, Baltimore, Maryland, 210-214.

Metternich, J., Bechtloff, S., Seifermann, S., 2013. Efficiency and economic evaluation of cellular manufacturing to enable lean machining. *Procedia CIRP*, *7*, 592-597.

Mizsey, P., Fonyo, Z., 1990. Toward a more realistic overall process synthesis—the combined approach. Computers & Chemical Engineering, 14(11), 1213-1236.

Offord, P., 2011. RPR: A Problem Diagnosis Method for IT Professionals. Advance Seven Limited, Essex, U.K.

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559-572.

Pechmann, W., 1948. Plant organization. Plant Management. Industrial & Engineering Chemistry, 40(7), 77A-92A.

Ran, R., Wang, B., 2015. Combining grey relational analysis and TOPSIS concepts for evaluating the technical innovation capability of high technology enterprises with fuzzy information. Journal of Intelligent & Fuzzy Systems, 29(4), 1301-1309.

Ray, S.C., 2015. Nonparametric measures of scale economies and capacity utilization: An application to US manufacturing. *European Journal of Operational Research*, *245*(2), 602-611.

Saaty, R., 1987. The analytic hierarchy process—what it is and how it is used. Mathematical Modelling, 9(3-5), 161-176.

Sullivan, W.G., McDonald, T.N., Van Aken, E.M., 2002. Equipment replacement decisions and lean manufacturing. *Robotics and Computer-Integrated Manufacturing*, *18*(3-4), 255-265.

Opricovic, S., 1990. Programski paket VIKOR za visekriterijumsko kompromisno rangiranje. SYMOPIS.

Opricovic, S., 1998. Multicriteria Optimization in Civil Engineering. PhD Thesis, Faculty of Civil Engineering, Belgrade.

Opricovic, S., Tzeng G.-H., 2007. Extended VIKOR Method in Comparison with Outranking Methods. European Journal of Operational Research, 178(2), 514–529.

Seifert, T., Schreider, H., Sievers, S., Schembecker, G., Bramsiepe, C., 2015. Real option framework for equipment wise expansion of modular plants applied to the design of a continuous multiproduct plant. Chemical Engineering Research and Design, 93, 511-521.

Shannon, C.E., 1948. A mathematical theory of communications. Bell Systems Technical Journal, 27(3), 379-423.

Sufiyan, M., Haleem, A., Khan, S., Khan, M.I., 2019. Evaluating food supply chain performance using hybrid fuzzy MCDM technique. *Sustainable Production and Consumption*, *20*, 40-57.

Sun J., Peng, X., Xu, Y., Sun, B., Ma, N., Lv, P., 2017. VIKOR-GRA based Intuitionistic Fuzzy Multi-Attribute WEB Service Selection. Boletín Técnico, 55(7), 982-996.

Tan, J., Foo, D.C.Y., Kumaresan, S., Aziz, R.A., 2004. Modelling, optimisation, and debottlenecking of a pharmaceutical production process utilising a batch process simulator. The Malaysian Society for Molecular Biology and Biotechnology Scientific Meeting (MSMBB 2004), Melaka.

Tan, J., Foo, D.C.Y., Kumaresan, S., Aziz, R.A., 2006. Debottlenecking of a Batch Pharmaceutical Cream Production. Pharmaceutical Engineering, July/August 2006, 72-84.

Tan, J., Low, K.Y., Sulaiman, N.M.N., Tan, R.R., Promentilla, M.A.B., 2016. Fuzzy analytic hierarchy process (FAHP) for multi-criteria selection of microalgae harvesting and drying processes. Clean Technologies and Environmental Policy, 18(7), 2049-2063.

Tan, M., Andiappan, V., Wan, Y., 2018. Process Debottlenecking and Retrofit of Palm Oil Milling Process via Inoperability Input-Output Modelling. MATEC Web of Conferences, 152, 01012.

Tan, R., Aviso, K., Foo, D.C.Y., 2017. P-graph and Monte Carlo simulation approach to planning carbon management networks. Computers & Chemical Engineering, 106,872-882.

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Cheah, A.C.S., Motavasel, Z., Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. *Journal of Cleaner Production*, *225*, 359-375.

Van Duc Long, N., Lee, M., 2011. Improved energy efficiency in debottlenecking using a fully thermally coupled distillation column. Asia-Pacific Journal of Chemical Engineering, 6(3), 338-348.

Van Duc Long, N., Lee, S., Lee, M., 2010. Design and optimization of a dividing wall column for debottlenecking of the acetic acid purification process. Chemical Engineering and Processing: Process Intensification, 49(8), 825-835.

Walker, W.R., 2011. Environmental regulation and labor reallocation: Evidence from the Clean Air Act. *American Economic Review*, *101*(3), 442-47.

Wang, Q., Peng, A., 2010. Developing MCDM Approach Based on GRA and TOPSIS. Applied Mechanics and Materials, 34-35, 1931-1935.

WAR GUI, 2011. Waste Reduction Algorithm: Chemical Process Simulation for Waste Reduction. Available at:www.epa.gov/chemical-research/waste-reduction-algorithm-chemical-process-simulation-waste-reduction [Accessed 1 May 2019].

Weng, Z.K., 1996. Manufacturing lead times, system utilization rates and lead-time-related demand. *European Journal of Operational Research*, *89*(2), 259-268.

Yang, S., Wang, P., Li, G., Huang, K., 2015. Group Decision Making Model for Weapon Selection Using Extended VIKOR Method under Intuitionistic Fuzzy Environment. Journal of System Simulation, 27(9), 2169-2273.

Yang, Y., Xue, D., 2017. Grey differential system and control problems based on the fractional calculus. 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, China.

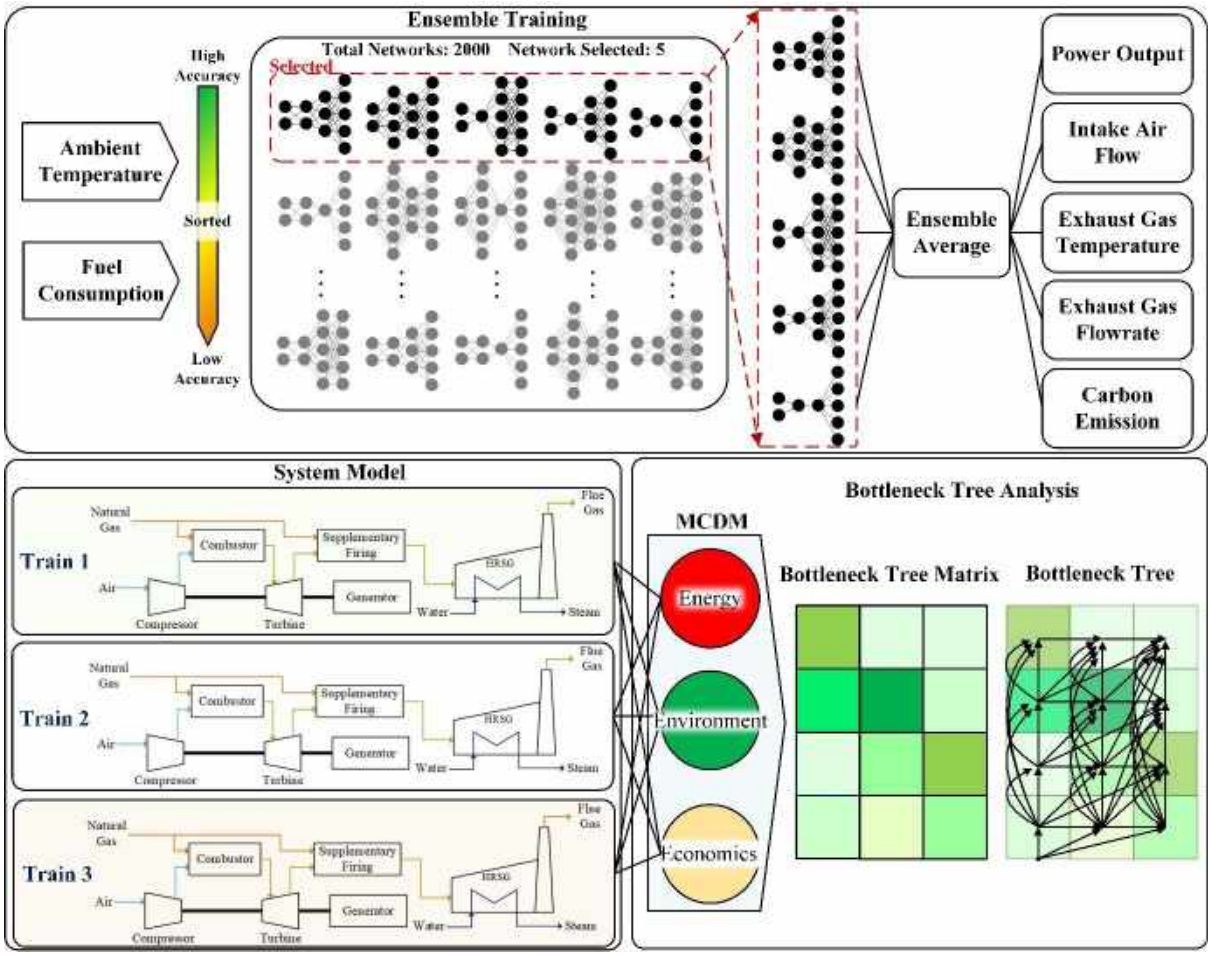
Zhang, L.J., Li, Z.J., 2006. Gene Selection for Classifying Microarray Data Using Grey Relation Analysis. In: Todorovski, L., Lavrac, N., Jantke, K.P., Discovery Science. Springer, Berlin, Germany.

Zhang, J., Zhu, X., Towler, G., 2001. A Level-by-Level Debottlenecking Approach in Refinery Operation. Industrial & Engineering Chemistry Research, 40(6), 1528-1540.

Zou, Z.-H., Yun, Y., Sun, J.-N., 2006. Entropy method for determination of weight of evaluating in fuzzy synthetic evaluation for water quality assessment. Journal of environmental Sciences, 18(5), 1020-1023.

CHAPTER 9 DATA-DRIVEN MULTI-DIMENSIONAL BOTTLENECK TREES

This chapter is currently under review in Energy journal.



Abstract

Cogeneration systems and technologies have recently grown towards maturity within many countries and regions. With an increasing amount of policy support for combined heat and power (CHP) plants and feed-in tariff (FiT), cogeneration systems have blossomed in many existing industries and became their backbone technology for energy generation. Nevertheless, with ever-increasing energy demands, the required capacity of cogeneration gradually grows yearly. This situation unveils a crawling problem in the back of the picture where many existing cogeneration systems require more energy output than their allocated design capacity. This work analyses capacity debottlenecking for multi-train cogeneration plants from the perspective of data-driven and multi-criteria planning approaches. First, cogeneration systems were modelled using an ensemble neural network model to obtain information that can be translated into energy, environment, and economic indicators. These indicators are then

evaluated using a multi-criteria decision making (MCDM) method to perform a bottleneck tree analysis (BOTA) to identify optimal pathways to plan for debottlenecking projects in a multi-train cogeneration plant. This study resulted in an improvement of 54.2 % carbon emission reduction per power generation and 59 % improvement in total energy production which is achievable within payback period of 93.9 weeks.

Keywords: Combined Heat and Power (CHP), Bottleneck Tree Analysis (BOTA), Artificial Neural Network, Multi-Criteria Decision-Making (MCDM), Grey Relational Analysis, TOPSIS.

9.1 Introduction

The global initiative of Sustainable Development Goal (SDG) by the United Nations (UN) has laid out a blueprint for peace and prosperity for people and the planet (SDG, 2019). The agenda of SDG is to preserve the sustainable lifestyle and living of the global population. One of the main reasons that made sustainable production so important is the contribution of manufacturing in a country's economy. Additionally, The World Bank (2020) has shown a steady increase in manufacturing demand since 2000. In the European Union (EU), the industries are generating 24 % of the EU gross domestic product (GDP) (Behun et al., 2018). In response to the SDG initiative, the industrialists have sought for alternative solutions for green technology such as the cogeneration process. Cogeneration, also known as combined heat and power (CHP), burns natural gas or diesel fuel to generate both electric and thermal energy (Leong et al., 2019b). The CHP has been proved to be able to reduce the greenhouse gas (GHG) emission by 30-60 % (IRP, 2017). According to the Environmental Protection Agency (EPA, 2020), cogeneration can achieve an overall efficiency of 75 % as compared to a combination of boiler and power plant which has an overall efficiency of 51 %.

The CHP system is driven by the prime mover unit including gas turbine, internal and external combustion engine, steam turbine, microturbine and fuel cell (Al-Sulaiman et al., 2011). Under specific conditions, CHP can yield up to 80% efficiency (EPA, 2020). The flexibility in cogeneration setup allows the industrialist from different sectors to take advantage of the technology. The CHP system can be branch out into many configurations as illustrated in Figure 9.1.

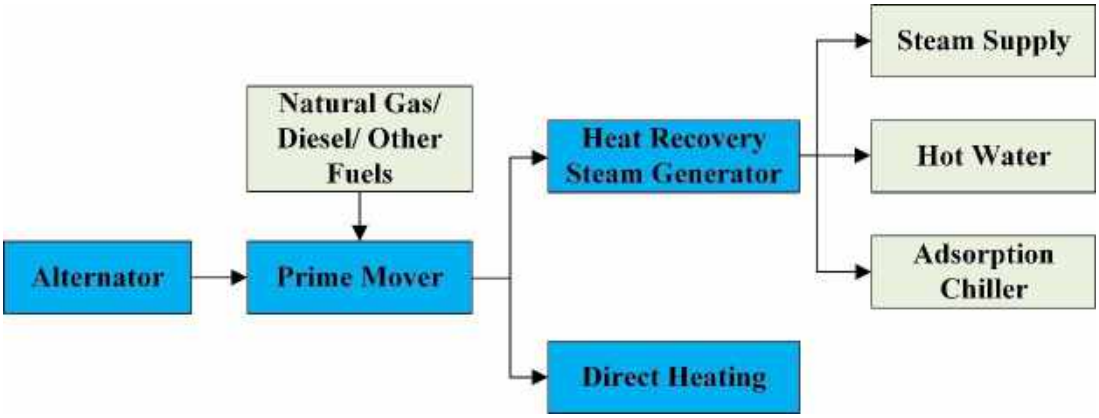


Figure 9.1: Possible configuration for cogeneration system

The key CHP technology providers are Kawasaki Heavy Industries, Siemens AG, Mitsubishi Heavy Industries, Aegis Energy Services Inc, Rolls Royce Plc., etc (TMR, 2020). The proprietary technology and method in the prime mover's design are unique from each technology provider. For instance, Mitsubishi has developed its proprietary burner to burn hydrogen fuel (PEI, 2018). Kawasaki Heavy Industries has a proprietary remote monitoring system to assist their client in monitoring the performance of their gas turbine (KHI, 2018). The prime mover of the CHP system is designed based on the International Organisation for Standardization (ISO) rating where the prime mover's performance is measured at 101.3 kPa, 15 ^{o}C and 60 % relative humidity. Thus, the ambient temperature will determine the final performance efficiency of the system. Nevertheless, the prime mover's equipment design is proprietary technology, industrialists are required to adhere to the recommended schedule maintenance. This resulted in the difference between the recommended maintenance schedule with the exact prime mover condition which reflects on energy, environmental and economic performances of the system. With such sub-optimal process management, the corporation might face unnecessary loss of profit.

The application of CHP system has been implemented in many areas, such as palm oil industry (Abd Majid et al., 2014), pulp and paper factory (Hashim et al., 2010), petrochemical industry (Ahmadi et al., 2019), independent power plant (Leong et al., 2019b), etc. Due to confidentiality of the proprietary technology, industrialists often relied on the technology provider to provide information and technology support for the CHP system. On the existing CHP system, the energy demand has increased with the expansion in energy consumption. Many research works have been done on optimising the design of CHP in the application of district cooling and underground thermal energy storage (Reverberi et al., 2011), district

heating network (Franco and Bellina, 2018), integration in housing complex (Fuentes-Cortés et al., 2015) and application in edible oil plant (Azhdari et al., 2009). The consideration of various parameters needs to be included in the optimisation process including inlet air intake, fuel consumption, ambient temperature, exhaust temperature, exhaust gas flowrate, heat recovery efficiency, manpower cost and environmental emission (Leong et al., 2019b). Aklilu and Gilani (2010) stated that complex mathematical models should be developed to simulate the actual condition of the CHP system. For instance, Gimelli and Muccillo (2013) developed a multi-objective approach to determine the optimal configuration of a CHP system by analysing the energy and economic benefit for hospital application. Kanbur et al. (2019) also focus on multi-objective optimisation procedure to develop a new sustainable index for liquified natural gas (LNG) micro-cogeneration system in identifying the optimal operating condition. Furthermore, Zare et al. (2019) implemented multi-objective genetic algorithms to locate the optimum fuel mixing ratio for biogas and natural gas. At a 50 % mixing ratio, the outcome resulted in higher exergy efficiency and lower cost. Braun et al. (2016) has optimised the CHP system performance with a neuro-genetic strategy to minimise fuel consumption while maximising both electric generation and heat recovery. Braun et al. (2016) added that optimisation of the CHP system was done on-line during the operation with the aid of a neural network. Li et al. (2018) reduced the CHP heating energy consumption by 47.7 % through optimising the heat recovery system. Based on the literature review, the optimisation of the CHP system can be challenging due to fluctuation of energy price (Gu et al., 2015) and accuracy of the model in real time (Rist et al., 2017). Thus, to cope with increased energy demand, bottlenecks of the CHP system shall be identified. In conventional practices, CHP owners consider investing in new systems to cope with increased energy demand. This can be capital intensive and time consuming. Therefore, identifying the bottleneck of the system can be critical to improve productivity with minimum capital investment.

Process debottlenecking methods are derived from the root of process system design (Teng et al., 2020). Generally, the heuristic approach to process synthesis is developed from the Onion model (Linnhoff and Boland, 1982) and hierarchy of decision (Douglas, 1985). Based on literature, the bottleneck associated with the CHP system focuses on the macro perspective including CHP as a distributed power plant for the local community (Pehnt et al., 2006), reducing transmission and energy loss (Danny Harvey, 2012) and sufficient supply of energy for city consumption (DFIC, 2016). Dr. Fromme International Consulting (DFIC) (DFIC, 2016) added that CHP systems can be a debottleneck solution to many energy solutions. However,

the debottlenecking of the CHP system on micro scale has not received attention. As the prime mover of the CHP system is designed with optimal conditions, the debottlenecking strategy in the CHP system can increase the total energy output by targeting the small component in the system.

To establish the debottlenecking model, multi-criteria decision making (MCDM) is used to determine the relationship between the operating parameters. MCDM method is a complex decision-making tool that converts quantitative and qualitative data input for decision making (Mardani et al., 2015). The output of MCDM is mainly expressed in weight to represent the importance of the parameter. Teng et al. (2020) highlighted that the weight of the parameters is normally expressed in the range of 0 – 1. With that, a parameter with higher weight reflects a higher priority during decision making. The common MCDM method being used in the field such as analytic hierarchical process (AHP), analytic network process (ANP), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Grey relational analysis (GRA), Višekriterijumsko Kompromisno Rangiranje (VIKOR), etc. Besides, there is also a combination of various integrations between MCDM methods such as GRA-TOPSIS and fuzzy-ANP. The combined MCDM approaches are normally used when an individual MCDM method is unable to determine the accuracy of the performance of the alternative solution and ideal solution (Wang and Peng, 2010). For example, GRA-TOPSIS method demonstrated that the grey relation coefficient is used to measure the closeness between feasible and ideal solutions without the need of weight input. Teng et al. (2020) illustrated the use of GRA-TOPSIS in bottlenecking strategy in industrial refineries to identify the optimum bottleneck pathway. Kirubakaran and Ilangkumaran (2016) have also applied GRA-TOPSIS to identify the optimum maintenance strategy for the paper pulp industry. As for fuzzy-ANP, the fuzzy set theory accommodates uncertainty factors associated with the interaction between each element included in the supermatrix (Asan et al., 2012). The selection of MCDM method is highly dependent on the requirement of the case study.

For industrial analysis, the paradigm shift in Industry Revolution 4.0 has introduced the machine learning (ML) model as a powerful tool for industrial analytics. One of the most popular ML method is by using neural networks to extract complex Big Data representation at high-level of abstraction (Najafabadi et al., 2015). The neural network has been applied in many areas including pattern classification, language processing, complex system modelling, optimisation, and prediction (Abiodun et al., 2018). Having said that, a huge amount of data is

required to obtain higher accuracy in the data-analytical process. Lenz et al. (2018) applied a data-driven method in a smart manufacturing system for predictive maintenance for cost optimisation. Vondra et al., (2019) stated that data-driven methods can identify the techno-economic study of biogas plants more effectively. Teng et al. (2019e) has extended the data-driven application to optimise thermal degradation of microalgae. Fast and Palmé (2010) stated that the application of neural network models in CHP plants produced highly accurate predictions for condition monitoring and diagnosis.

By incorporating data-driven methods in the debottlenecking strategy, the best debottlenecking pathway for the complex CHP system can be identified more effectively. This work aims to debottleneck the CHP system with the implementation of neural network models to increase the system capacity with higher efficiency at a lower cost.

9.2 Methods

Process systems are modelled using an ensemble of neural networks with mass and energy balance. Firstly, the neural network ensemble is modelled with a supervised strategy with uncertainty and variable data as input and critical process data as output. Uncertainty and variable data are normally provided from the technology providing company, performance data sheet and historical operational data. Critical process data are processing variables that exhibit highly non-linear behaviour and are important toward assessment of the full process model in terms of sustainability (elements such as energy, environmental impacts, and economics). This allows for a data-driven insight into the uncertainty and variable data on the full process (see Figure 9.2).

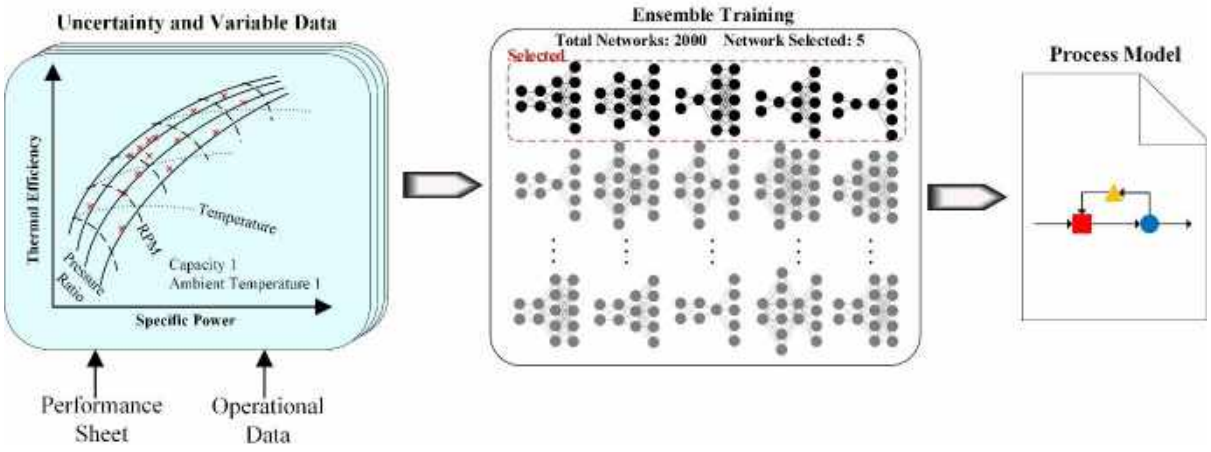


Figure 9.2: Overall method for data-driven modelling with neural network ensemble to consider uncertainty and variable data.

Neural network ensembles are sets of multiple individual neural networks being trained to provide a combined prediction power based on either accuracy or diversity (Hansen and Salamon, 1990). (Brown, 2005) provided mathematical evidence that prediction accuracy can be improved in ensemble models by distributing the errors of the full input space into multiple individual base models. In this work, this work proposes an ensemble training algorithm by utilising random sampling with elitist strategy to select 5 neural networks with top validation accuracy. Each neural network is individually trained using the popular gradient descent algorithm, Adam (Kingma and Ba, 2015) and cross-validated using separate datasets. The computational cost is bounded with the maximum number of evaluated neural networks (set to 2000 in this case). Moreover, the activation functions for the neural networks within the ensemble can be selected from multiple activation functions which include sigmoid, exponential (Exp), logarithmic (Log), linear, identity, radial basis function (RBF). The number of neurons within each layer can be selected from between integers of 1 to 200 which allows for many combinations of neural architectures. The ensemble prediction is provided by the ensemble averaging method (Naftaly et al., 1997). The pseudocode for neural network ensemble training and inference can be found in Figure 9.3.

Algorithm 1: Selection and Training Neural Networks for Ensemble

```
Act_func ← ['activation 1', 'activation 2', ...]      ▶Define all activation function available
Total_net ← 2000      ▶Set total number of neural networks to be evaluated
Network_select ← 5      ▶Set number of neural networks to be selected for ensemble network
Best_network_error ← [+inf]*Network_select  ▶Initialize selected error vector with correct size
while (i ≤ Total_net):
    HL_num ← random_uniform(1, Max_HL)  ▶Sample number of hidden neurons randomly
    Input_act ← Act_func[random_int(1, len(Act_func)] ▶Sample activation function of input randomly
    HL_act ← Act_func[random_int(1, len(Act_func)]  ▶Sample activation function of hidden layer
    Output_act ← Act_func[random_int(1, len(Act_func)] ▶Sample activation function of output layer
    Network ← Train_Network(act=[Input_act, HL_act, Output_act],
                                                num_neuron=HL_num)  ▶Train neural network
    Val_error ← Network.get_validation_loss()  ▶Obtain the loss/error from validation set
    if (Val_error < max(Best_network_error)):  ▶Check if the loss is better than previously recorded
        Best_network_error.append_sort(Val_error)  ▶Update the record
        Network.save_remove_worse()  ▶Save neural network model and remove the worse if required
        if (len(Best_network_error) > Network_select):  ▶Check if record is too big
            Best_network_error ← Best_network_error[0: Network_select]  ▶Remove excess record
    i = i+1      ▶Update number of networks evaluated
```

Algorithm 2: Ensemble Learning Inference

```
Predict = []      ▶Define a vector for prediction
for Network in Saved_Networks:      ▶Loop over all the selected top performing neural networks
    Network.load_network()      ▶Load the network
    Predict.append(Network.predict(x))  ▶Give the network an input, x and append the prediction
y = Predict.average()  ▶Average all the predictions from all the networks
```

Figure 9.3: Pseudocode for selection and training neural networks for ensemble (algorithm 1) and for ensemble learning inference (algorithm 2)

For capacity debottlenecking, the input capacity of the process is expected to increase until one of the units within the process has its operating capacity at the maximum design capacity (Litzen and Bravo, 1999). This is referred to as a "bottleneck" within the process. Project engineers can potentially remove this bottleneck by replacing the unit, modifying the design, installing additional units and other retrofit techniques (Van Duc Long and Lee, 2011). By further increasing the overall capacity, another unit then becomes the new bottleneck within the process and more improvement work is required. As a continuation, this process repeats itself until essentially all the equipment within the facility has been replaced. This is equivalent to building a new plant, which is highly illogical for debottlenecking studies since a new plant could be built separately with lower costs and lesser hassle. Therefore, in this field, it is crucial to plan when to stop carrying out debottlenecking works from a long-term sustainability point of view.

Bottleneck Tree Analysis (BOTA) was first proposed by Teng et al. (2020) for the purpose of process capacity debottlenecking. The concept of the holistic debottlenecking method considers of the long-term sustainability instead of short-term goals (such as purely profit) while decoupling the problem using a graphical matrix and tree approach to effective achieve 3 goals: (i) identification of multiple bottlenecks within multiple areas with varying operating capacities (ii) evaluate their benefits in-terms of long-term sustainability and find a pathway to achieve the most beneficial point (iii) schedule the most economically feasible way to achieve the pathway towards the most beneficial point. This most beneficial point has the maximum value (Green and Lean index) within the tree and is referred to as the bottleneck stopping point (BSP). Each possible bottleneck stage is evaluated in aspects related to environment, economics, energy, and other applicable criteria using a multi-criteria decision-making (MCDM) method. Effectively, the MCDM method assigns a score which is called the green and lean index (GLI), representing the benefits of long-term sustainability of the facility in each bottleneck stage (Leong et al., 2019b). Next, by aligning these GLI values in a capacity increasing order, the bottleneck tree matrix is formed. The global maximum value is the BSP, and the pathway is selected using a latent objective of maximising the number of steps required while maintaining logical constraints. This latent objective functions as a solution reduction technique to ensure only solutions which utilises all benefits from incremental improvements are selected. Finally, the economic feasibility of each possible schedule of the pathway is evaluated. This method is compact and allows for simplification of complex debottlenecking problems into illustrative graphical representations and solves them effectively.

For process system analysis, due to the complexity of most systems, there is a convention to decouple the full facility into areas of integrity for further analysis (Ahmad and Hui, 1991). Works such as Bagajewicz and Rodera (Bagajewicz and Rodera, 2002) decoupled the total site of processing plants into separate areas to carry out heat exchanger network optimisation. In the fields of process capacity debottlenecking, the original BOTA technology (Teng et al., 2020) showed the effectiveness of the analysis in debottlenecking an oil refinery by decoupling the full facility into two separate areas. The BOTA methodology demonstrated the advantages of being graphically explainable, computationally inexpensive, and easy to implement. Within the context of the BOTA debottlenecking study, the "number of areas" (also referred to as "dimension"), refers to the number of regions of process that have a nonsignificant effect of interrelated capacity towards its individual performance and can operate at different capacities separately. It is to be highlighted that this "area" related to debottlenecking studies is not equivalent to the number of units within the system or operation areas within the facility. Under conventional process design (Smith, 2005), the number of areas is significantly lesser than that of the units within the system and is rarely at a high value. The original work of BOTA (Teng et al., 2020) only demonstrated the debottlenecking methodology with a dimension of 2, while the usage of more than 2 dimensions was never formally demonstrated is non-straightforward for implementation. This is not an inherent problem with BOTA, but rather a general difficulty faced by graphical methods (Chambers et al., 2018). For example, the famous graphical optimisation tool, P-graph (Friedler et al., 1992) also faced this problem and works such as How et al. (2018) studied the challenges with multi-dimensional modelling within the graphical framework. Thus, this work formally demonstrates the implementation of BOTA technology for multidimensional problems of 3 and above.

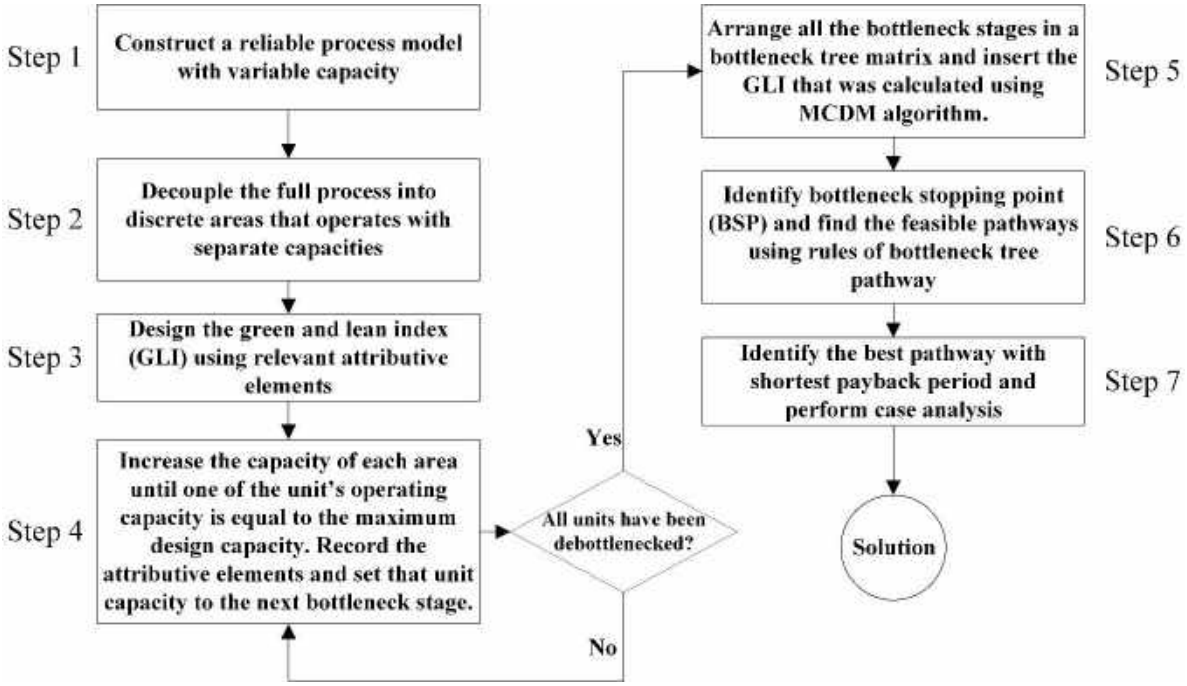


Figure 9.4: Flow diagram for the multi-dimensional bottleneck tree analysis (BOTA) technology

In general, the method for BOTA can be explained in 7 steps as shown in Figure 9.4. For construction of a reliable process model (Step 1), authors such as Koulouris et al. (Koulouris et al., 2000) deployed the use of a commercial process simulator, SuperPro Designer software. Alternative, Teng et al. (2019c) demonstrated the use of Aspen software for simulation of an oil refinery. However, for systems with higher variability, data-driven approaches are required to produce a reliable process model (Ge, 2017). This work used an ensemble neural network model to predict process variables with high variability and input-output models for equipment with static nature. In step 2, the full process should be decoupled into areas with different operational capacities while having minimal performance dependencies between each other. This can exist for areas that are in sequential or parallel configuration within the process. For the case study of this work, the area decoupling is for area in parallel configuration since the case study is for multiple trains of cogeneration units. For step 3, the design of green and lean index (GLI) generally depends on the company or facility of implementation. This work recommends a systematic approach of selecting relevant attributive elements for the specific case study (Leong et al., 2019b). Subsequently, the capacity of each area is increased until one of the unit operating capacity is at its maximum design capacity in Step 4, this capacity is a bottleneck capacity. This work records all the attributive elements at this stage, set the capacity of this unit to the next bottleneck capacity and continue to increase the area capacity. In practise,

the design of the bottlenecked equipment will be retrofitted or improved to give higher capacity. For computational simplification, this work proposes the use of a utilisation factor where utilisation factor is at 1 for when current operating capacity is equal to maximum design capacity.

$$Utilization\ factor = \frac{Current\ Operating\ Capacity}{Maximum\ Design\ Capacity} \tag{9.1}$$

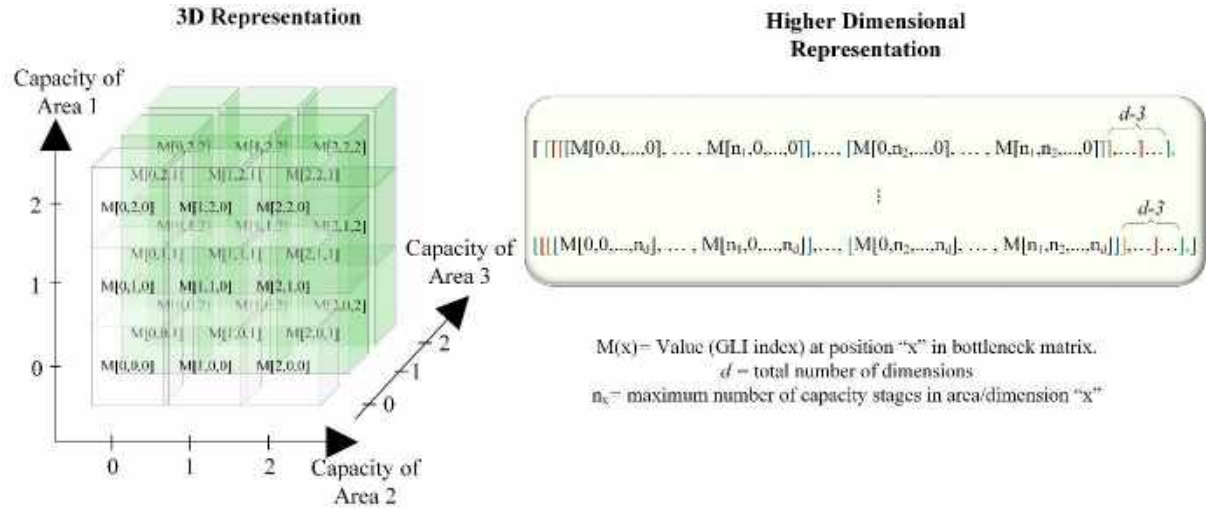


Figure 9.5: Representation for bottleneck matrix for three dimension (3D) and higher dimensions

In step 5, the bottleneck stages are arranged in a *d*-dimensional bottleneck tree matrix (see Figure 5) in an ascending manner (where *d* is the number of areas). The bottleneck tree matrix will contain a cell (state) for each combination of bottleneck capacity. From the attributive elements of each state, the *GLI* can be computed using MCDM algorithm (Refer to Chapter 2.1). The global maximum of *GLI* within the matrix is set as the *BSP*, where this is the state of the process which gives the highest benefit for long-term sustainability. To identify feasible pathways from the origin towards the *BSP*, this work deploys an improved version for bottleneck tree pathway rules:

Rule 1: The bottleneck pathway should start from origin and end at *BSP*.

Rule 2: The bottleneck pathway can only move from low capacity to high capacity.

Rule 3: The bottleneck pathway can only move from low *GLI* to high *GLI*.

Rule 4: Jumping through a cell is allowed, however always take the longest possible pathway (it is possible to have multiple longest pathways).

Rule 5: If there is more than one longest pathway, only accept pathways that pass through the cell highest *GLI* within 1 step of the origin.

Note that the original fifth rule from the previous Chapter (Teng et al., 2020) alternatively states to accept pathway with stage 1 in all regions as a heuristic. This work further improved this heuristic with a mathematical proof (see appendix) and argue that that original work is a conditional case of the improved fifth rule. In terms of formal mathematical formulation, the multi-dimensional bottleneck tree pathway rules can be presented as in Eqn. 9.1 to 9.5.

$$\max S \tag{9.1}$$

$$x_{n+1} = x_n + \left[\Delta x_{n,1}, \dots, \Delta x_{n,d}\right] \quad \forall n \in [0, s], \ \Delta x \in N = \{0,1,2, \dots\} \tag{9.2}$$

$$M(x_{n+1}) > M(x_n) \tag{9.3}$$

$$x_0 = [0, \dots, 0] \ \ and \ x_s = u^* \quad where \quad M(u^*) = BSP \tag{9.4}$$

$$if \ count\big(argmax(S)\big) > 1 \ then \ \ p^* = p_d \quad where \quad x_1 = \max\left(x_0 + q\right) \in p_d \tag{9.5}$$

where S is the number of steps taken in project stages, x_n is the position of the cell in the matrix in the "*n*" project stage, $M(x_n)$ is the *GLI* at the cell position of x_n and q is a d-dimensional vector with binary value in any dimension.

Since the pathway rules have possibility of providing multiple feasible pathways, the secondary objective is to compare the payback period (*PP*) of each pathway of the project. All the feasible pathways and their schedules are then compared and the pathway with the minimal payback period is selected. The calculation for the time and state dependant net present value (NPV) is presented as in Eqn. 9.6 and the pathway dependant payback period can be represented in Eqn. 9.7 or solved graphically.

$$NPV(x,t) = II + \textstyle\sum_t NP(x,t) - \sum_t IC(x,t) \tag{9.6}$$

$$PP(p) = T \ \ where \ \ NPV(x_s, T) = II + \textstyle\sum_t NP(0,t) \tag{9.7}$$

Where II is the initial investment, NP is the state and time dependant net profit, IC represents the state and time dependant investment cost.

As a consideration for process variance, this work proposes carrying out case analysis for worst-case and best-case scenarios by reapplying the BOTA technology to the reliable process

model under worse and best process variances. This ensures that the schedule is a reliable solution under process variance.

9.2.1 Attributive Elements and Multi-Criteria Composite Index

The BOTA technology functions by evaluating the long-term sustainability of the studied facility under different stages of bottleneck levels by using a composite index. This index is called the green and lean index (*GLI*) and is commonly constructed using multi-criteria decision-making (MCDM) methods. This index ranges from 0 to 1, where a larger value indicates better expected long-term sustainability of the facility. The BOTA technology hence de-bottlenecks the studied facility until the maximum point is researched, also known as the BSP. Efficient pathways are then planned and scheduled to achieve the BSP, which strikes a beneficial balance between long-term sustainability and feasibility. The consideration of this composite index (*GLI*) can depend on many factors such as data source, supply chain, geographical location, and other aspects (Zhou et al., 2012). Dow Jones (2007) reported that composite indexes representing sustainability are heavily affected by the nature of implemented industry. Therefore, the design of *GLI* can vary from company to company and there is a need for a large consideration of attributive elements. For process improvement purposes, Leong et al. (Leong et al., 2019b) proposed the consideration of aspects related to material, manpower, money, machine and environment (4M1E) and only choose attributive elements that are relevant to construct the index. Liliana (Liliana, 2016) has also discussed the possibility to consider aspects related to manpower, machine, material, management, maintenance, money, and mother nature (7M) for this purpose. For the optimisation of an oil refinery, Teng et al. (2019c) highlighted the importance of considering product quality, yield, energy, and environmental impacts. For some less technologically mature systems, Ngan et al. (Ngan et al., 2020) even proposed the usage of risk indicators into consideration. Nevertheless, the most important attributive elements for any system would include aspects of energy, environment, and economics (Omer, 2008). In this work, the case study specific attributive elements are selected to be energy, environment, and economics. Their specific calculation methods for attributive elements are demonstrated in the subsequent paragraph.

With consideration of the energy aspect, two forms of energy are calculated from the facility, being total electric energy generated and total heat energy generated. The corresponding equations can be found in Eqn. 9.8 and 9.9.

$$A_E = \sum_{i=1}^{n} E_i \quad (9.8)$$

$$A_H = \sum_{i=1}^{n} H_i \quad (9.9)$$

Where A_E is the attributive element for electrical power generated (MW), A_H is the attributive element for heat energy produced (MW), n is the number of sources (which in this case is equal to 3).

For the environmental attributive element, the carbon emission is considered for the studied cogeneration plant, as it is the only emission towards the environment. The element for carbon emission is calculated as following:

$$A_C = \sum_{i=1}^{n} F_{c,i} C_i \quad (9.10)$$

Where A_C is the attributive element for carbon emission (CO_2-eq tonnes/h), F_C is the flowrate of carbon dioxide containing stream (flue gas in this case), and C is mass fraction of carbon dioxide within the stream.

The economic attributive element is calculated with two economic elements, being capital expenditures (CapEx) and operating expenditures (OpEx). For debottlenecking purposes, the attributive element for CapEx, A_{Capex} is a direct quotation of equipment and installation that were procured from equipment providing contractors of the case study company. Alternatively, readers can use investment cost estimation methods from Peters et al. (2003) or Sinnott and Towler (2013) for front-end studies. For this case study, the income includes sales of power and heat energy while expenditure consists of natural gas and maintenance costs. The OpEx is calculated as the following:

$$A_{Opex} = \sum_{i=1}^{n} F_{E,i} \cdot P_E + \sum_{i=1}^{n} F_{H,i} \cdot P_H - \sum_{i=1}^{n} F_{N,i} \cdot P_N - M \sum_{i=1}^{n} F_{N,i} \quad (9.11)$$

where A_{Opex} is the attributive element for operating expenditures (m.u./year), F_E is the generation rate of electrical energy (MW), P_E is the price of electrical energy per year (m.u./MW year), F_H is the production rate of steam (ton/h), P_H is the price of heat energy per year (m.u. h/ton year), F_N is the flowrate of natural gas that is being consumed (MMSCFD), P_N is the price of natural gas (m.u./ MMSCFD year), M is the estimation of maintenance cost per year correlated to the feed capacity (m.u./MMSCFD year). Note that m.u. is the monetary unit that depends on the currency of the country (MYR for this case). These attributive elements will be combined into a composite index (the *GLI*) using the GRA-based TOPSIS multi-criteria decision-making algorithm.

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is one of the well-established multi-criteria decision-making (MCDM) approaches to select the best options, based on a given list of criteria set. The earliest application of TOPSIS can be backdated to the early 1980s (Hwang and Yoon, 1981). In general, TOPSIS will rank the alternatives based on its geometric distance from positive-ideal solution (PIS) and negative-ideal solution (NIS) (Hwang et al., 1993). In other words, the option that has the shortest geometric distance from PIS and the longest geometric distance to the NIS, is considered as the best alternative. Grey relational analysis (GRA), on the other hand, has been used to address complex problems that contain incomplete information and convoluted interrelationship between a given list of criteria (Eraslan and Çağman, 2017). Recently, integration between TOPSIS and GRA (or GRA-based TOPSIS) have been proposed to overcome the accuracy issues on the approximation of the geometric relationship between a given solution to the respective PIS and NIS (Wang and Peng, 2010). To date, this hybrid method has been proposed by researchers to address various complex process system engineering (PSE) problems such as supplier selection problems in supply chain management (Peng and Wang, 2011), equipment and process selections (P. Wang et al., 2016), process debottlenecking (Teng et al., 2020), etc. Figure 9.6 outlines the overall concept of the GRA-based TOPSIS, while the detailed formulations for GRA-based TOPSIS are presented as follows.

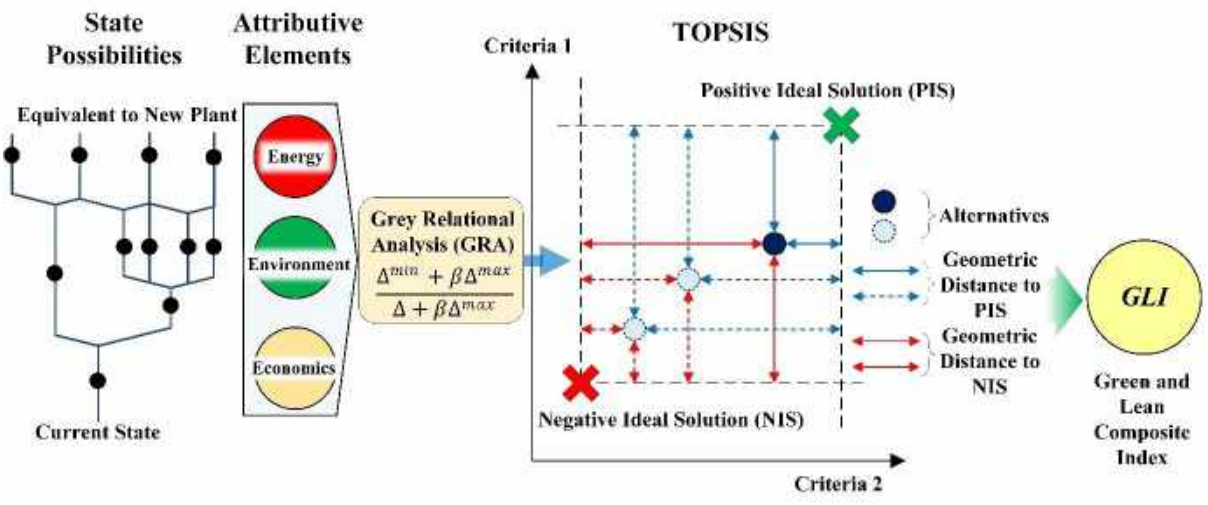


Figure 9.6: Illustration for the concept of GRA-based TOPSIS

Under GRA-based TOPSIS, the closeness between each alternative is defined using grey relational coefficient (GRC) instead of being expressed in the form of geometric distance

(Wang and Peng, 2010). The closeness between the given alternative *i* to the respective PIS and NIS in terms of criteria *j* (i.e., ξ_{ij}^{+} and ξ_{ij}^{-} respectively) can be mathematically expressed as:

$$\xi_{ij}^{+} = \frac{\Delta_{ij}^{+min} + \beta\, \Delta_{ij}^{+max}}{\Delta_{ij}^{+} + \beta\, \Delta_{ij}^{+max}} \qquad \forall i \in I, \forall j \in J \tag{9.12}$$

$$\xi_{ij}^{-} = \frac{\Delta_{ij}^{-min} + \beta\, \Delta_{ij}^{-max}}{\Delta_{ij}^{-} + \beta\, \Delta_{ij}^{-max}} \qquad \forall i \in I, \forall j \in J \tag{9.13}$$

where Δ_{ij}^{+max} and Δ_{ij}^{+min} denote the upper and lower limit of Δ_{ij}^{+}; the maximum and minimum value of Δ_{ij}^{-} are referred as Δ_{ij}^{-max} and Δ_{ij}^{-min} respectively; while β indicates the distinguish coefficient which used to control the resolution scale of the problem. In this work, this parameter is commonly set to be 0.5 to give a balanced result between PIS and NIS (Sylajakumari et al., 2018; Zhang and Li, 2006). Equation (9.14) and (9.15) is used to determine $\Delta_{ij}^{+}$and Δ_{ij}^{-}:

$$\Delta_{ij}^{+} = \left| x_j^{Best} - x_{ij} \right| \qquad \forall i \in I, \forall j \in J \tag{9.14}$$

$$\Delta_{ij}^{-} = \left| x_j^{Worst} - x_{ij} \right| \qquad \forall i \in I, \forall j \in J \tag{9.15}$$

Note that x_j^{Best} and x_j^{Worst} refer to the performance of PIS and NIS in terms of criteria *j* respectively. The relative closeness for each alternative *i* (C_i) is determined using Equation (9.16), where n denotes the total number of criteria, while γ_i^{+} and γ_i^{-} are computed using Equations (9.17) and (9.18). Please note that C_i is used as an indicator to rank all the alternatives (greater C_i is preferable). In this case, C_i is used as the *GLI*.

$$C_i = \frac{\gamma_i^{-}}{\gamma_i^{+} + \gamma_i^{-}} \qquad \forall i \in I \tag{9.16}$$

$$\gamma_i^{+} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}^{+} \qquad \forall i \in I \tag{9.17}$$

$$\gamma_i^{-} = \frac{1}{n} \sum_{j=1}^{n} \xi_{ij}^{-} \qquad \forall i \in I \tag{9.18}$$

9.3 Case Study

This case study demonstrates the proposed enhanced BOTA on CHP system. The case study adopted from an CHP facility in Malaysia has a series of CHP systems as illustrated in Figure 9.7. The multiple-train setup of the CHP system consists of 2 identical CHP systems in Train 1 and 2 (T1, T2) and a lower capacity CHP system in Train 3 (T3). Each CHP system operates

independently to cater for the final energy requirement. The main products that are produced from the CHP system are electricity and steam energy. The CHP system is classified into 4 main chapters including prime mover (i.e. gas turbine chapter), generator (i.e. electricity generation), supplementary firing and heat recovery steam generator (HRSG).

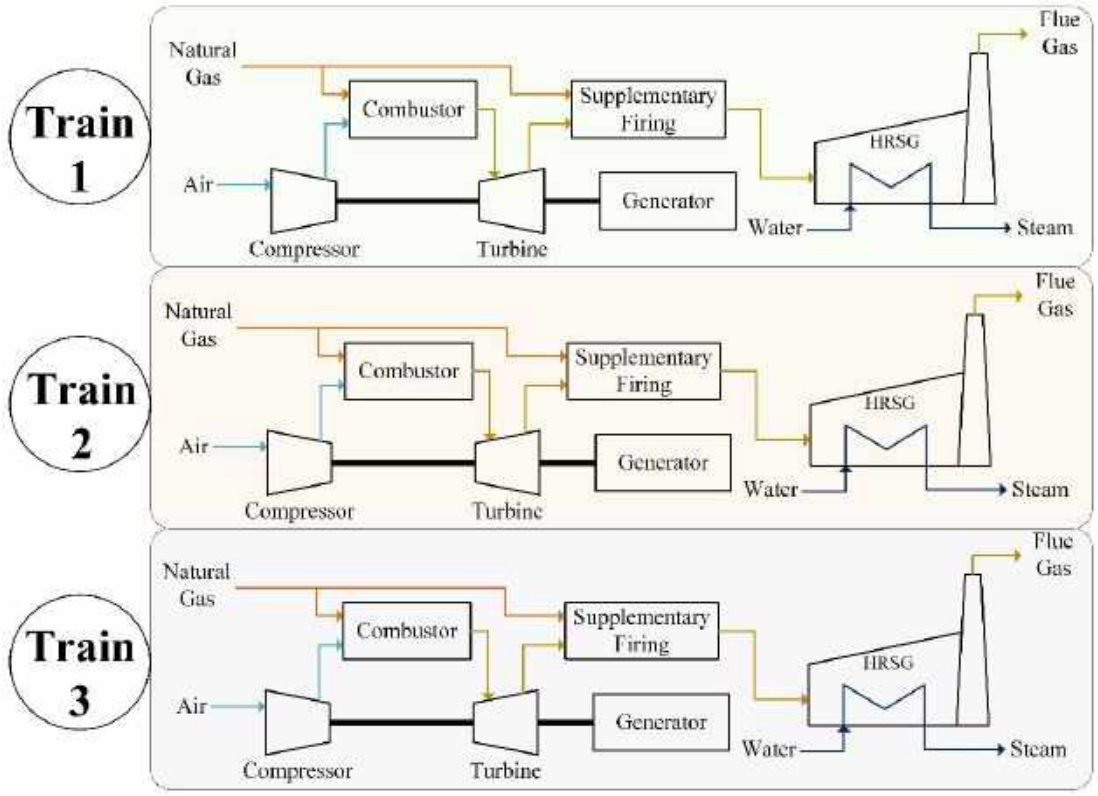


Figure 9.7: Simplified process flow diagram of a multi-train cogeneration plant located in Malaysia.

The lean and green (L&G) approach is considered in this study to contribute to SDG initiative for sustainability. The L&G approach is defined as elimination of non-value-added products from operation and environment (Leong et al., 2019a). As such, the performance parameters considered in the CHP system will be based on L&G approach. Fang and Lahdelma (2016) highlight the importance of ambient temperature in affecting the performance of prime mover especially for gas turbines. The inlet air flow rate is dependent on the ambient temperature and elevation that can influence the air density (Kohl and Nielsen, 1997). The average ambient temperature profile in Malaysia is illustrated in Figure 9.8(a). The maximum and minimum temperatures are 31.3 ^{o}C and 22.1 ^{o}C respectively (Figure 9.8(b)). This marks the temperature boundary to analyse the CHP performance. Sarkar (2015) added that the performance of gas turbines is degraded when the temperature is higher than the ISO condition. The fuel selection

and consumption are also important to determine the efficiency of the CHP system (Abbasi et al., 2015). On the CHP outlet, the total energy output including electricity and thermal energy, is the primary performance parameter for the CHP system (Bhatia, 2014). Monfort et al. (2014) highlighted that the exhaust gas temperature and flow determine the total recoverable waste heat from the prime mover. The typical exhaust gas profile ranged between 200 ^{o}C to 700 ^{o}C (Nolan, 2017). Siddiqui and Dincer (2019) added that considering the global warming potential (GWP) of the CHP system, carbon dioxide footprint is required to be measured.

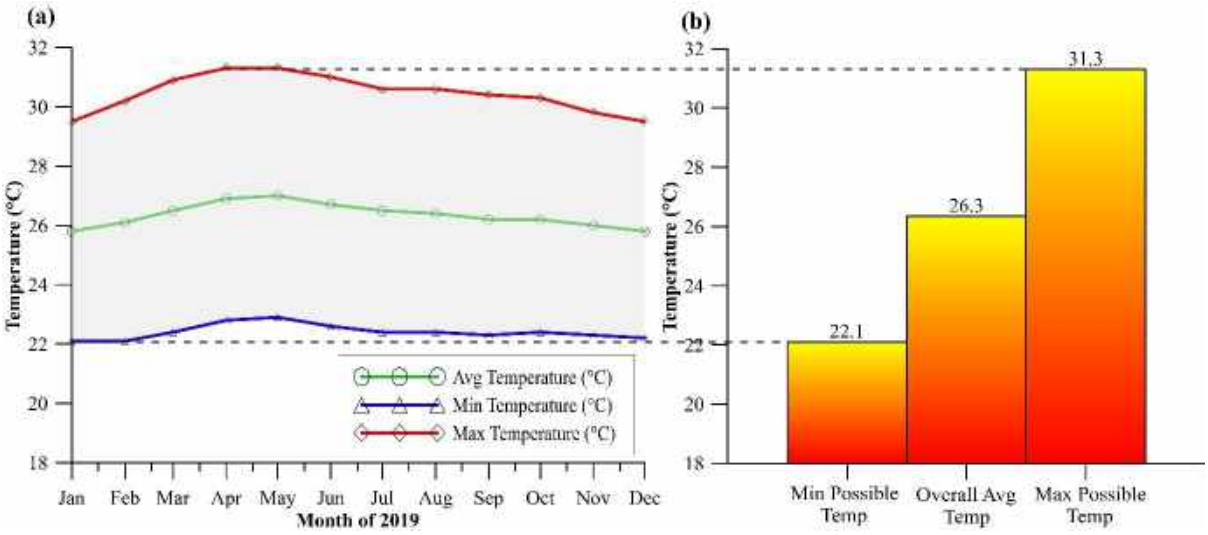


Figure 9.8: Temperature in case study location, Malaysia: (a) monthly temperature trend (data source: OpenWeatherMap Database (OpenWeatherMap, 2020)) (b) Minimum possible temperature, overall average temperature and maximum possible temperature.

9.4 Results and Discussion

The cogeneration performance data (from the case study company) was modelled using an ensemble of neural networks. The input for the neural network is ambient temperature and fuel capacity while power output, intake airflow, exhaust gas temperature, exhaust gas flowrate and carbon emission were predicted. A total of 241 datasets were used from the technology provider and operational data historian with the split of 70 % training data, 15 % testing data and 15 % validation data. This allows for a data-driven model for process performances with variation in ambient temperature and fuel capacity. Using the ensemble training method as described in Chapter 9.2, top five of the best selected neural networks have hidden layers of [112, 5], [107, 5], [105,5], [113,5], [46,5] with neuron of [Exp, Exp], [Sigmoid, Linear], [Sigmoid, Exp], [Exp,

Exp], [Sigmoid, Exp] respectively. Interestingly, the second hidden layer was searched and found to be optimal at 5 hidden neurons for this application. The prediction values from these best-performing neural networks were pooled to give an average combined prediction which gave a test result (Figure 9.9(a)) of 0.0154 RMSE (R^2=0.9978). This work further validated the ensemble prediction model with a separate dataset (Figure 9.9(b)) and obtained a similar result of 0.0170 RMSE (R^2=0.9970), confirming the validity of the training procedure. These predicted values were used as the basis to perform mass and energy balance analysis to ultimately compute the attributive elements representing the process (in terms of energy, environment, and economics).

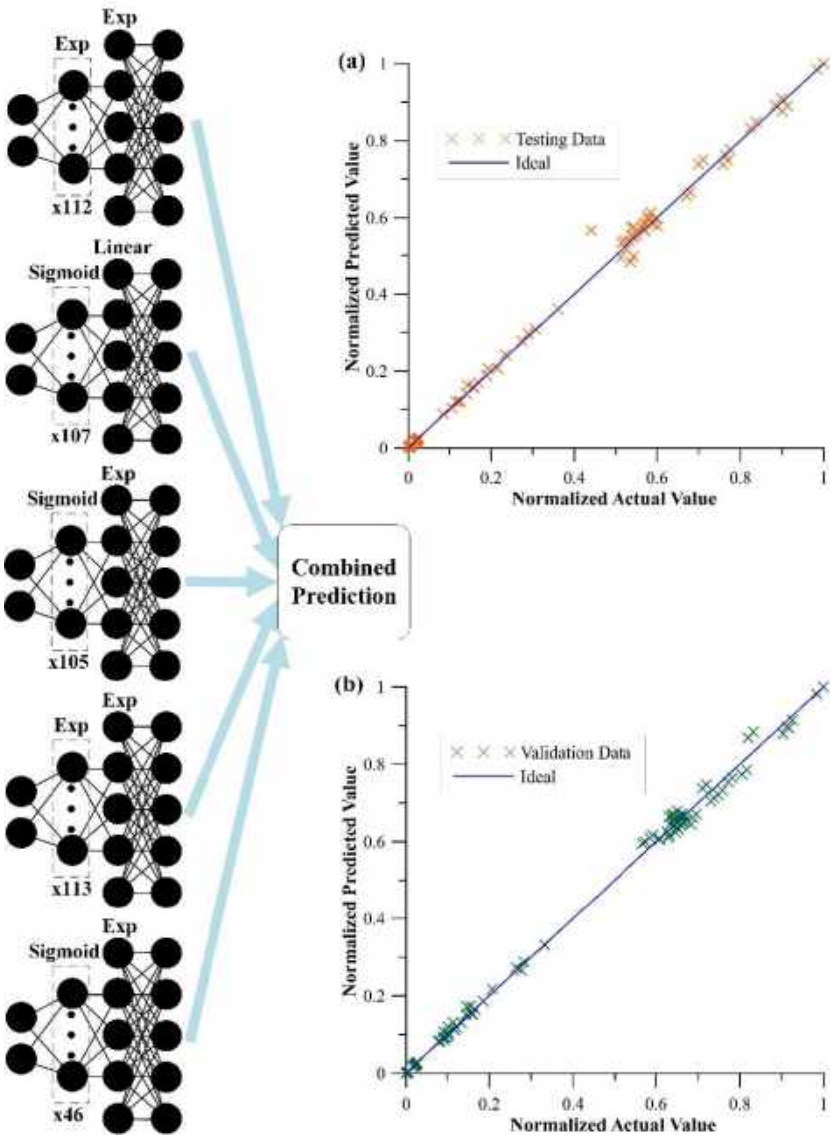


Figure 9.9: Architecture of neural network ensemble and performance for (a) testing dataset (b) validation dataset

Figure 9.10 shows the 3-dimensional bottleneck tree matrix generated for this case study, where the three axes show the capacities of the three cogeneration trains. There are 125 cells in the bottleneck tree matrix based on the five capacity levels of each cogeneration train (5 x 5 x 5). Using the capacity level of cogeneration train 3 as basis, the tree matrix is decomposed into 5 layers (see Figure 9.10). Each cell is assigned with a *GLI* value which is computed by using the GRA-based TOPSIS approach shown in Eqn. 9.1 to 9.7. Based on Figure 9.10, the bottom leftmost cell in the first layer (*GLI* = 0.056) is the origin state where the debottlenecking process is not carried out. Its adjacent cells, on the other hand, refer to the state when the corresponding cogeneration unit is operated under its maximum capacity (e.g., the cell located above the origin cell showed a *GLI* value of 0.302, this is the *GLI* can be obtained when cogeneration unit 1 is operated at its maximum capacity). Thereafter, further along the axis, each cell indicates the replacement of a larger-capacity cogeneration unit.

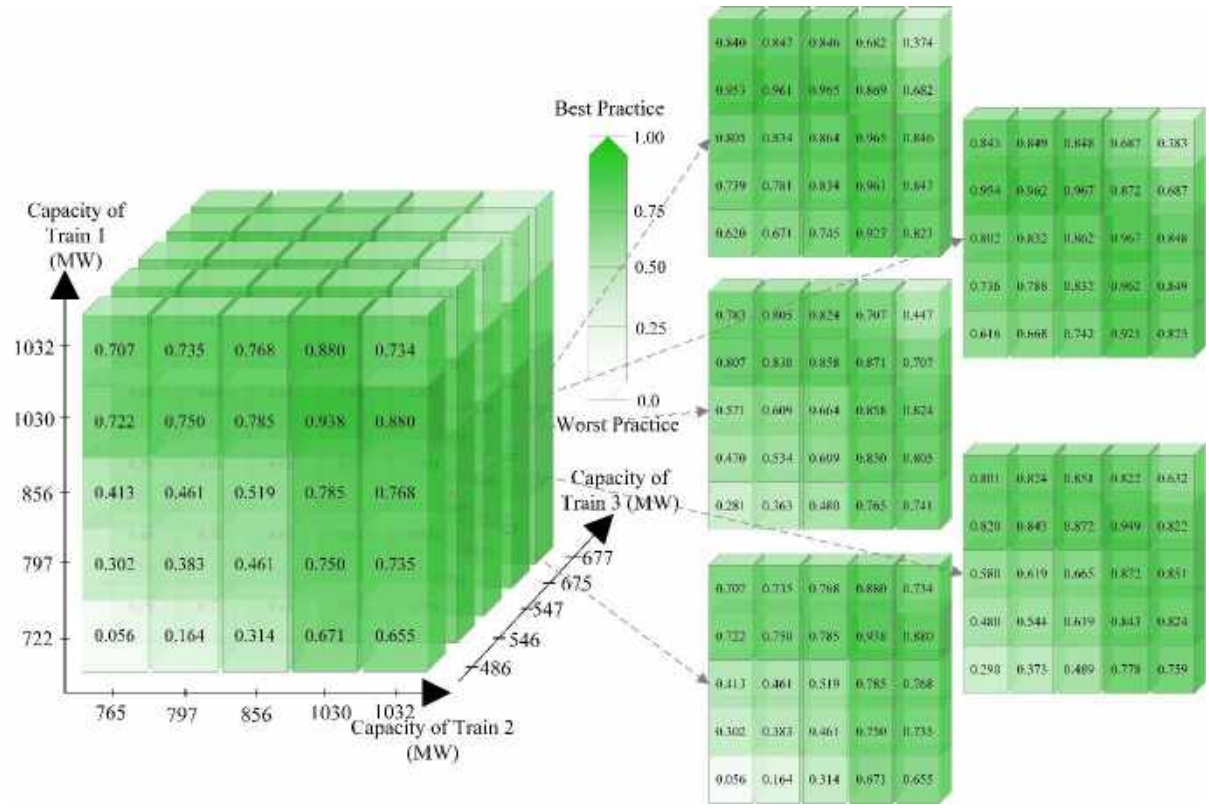


Figure 9.10: Three-dimensional bottleneck tree matrix for the presented case study

It can be clearly seen that the greatest achievable *GLI* appears at the fourth layer, i.e., 0.967 (red circles in Figure 9.11). These cells are considered as the BSPs for this case study. In other words, the debottlenecking process should start from the origin cell, progress gradually to another cell which provides higher *GLI* (i.e., higher sustainability), until it reaches one of the

BSP. It is worth noting that the existence of two BSPs in this case study is mainly due to the use of two identical sets of cogeneration units. This creates two superimposable options that provide the similar *GLI* value (e.g., the performance of using a 1030 MW overall input capacity of first cogeneration train and a 856 MW of that in the second cogeneration train, is the same as the performance of using a 856 MW overall input capacity of first cogeneration train and a 1030 MW of that in second cogeneration train). By following the five rules presented in Chapter 2, the feasible pathways can then be outlined in Figure 9.11. The first bottleneck stage shows a leap from the origin cell to a layer-two cell which provides a *GLI* of 0.544 (i.e., the sustainability performance when all three cogeneration units are operated under the existing maximum capacities). This is the common first debottlenecking action that should be conducted without the need of changing the existing process design (Hodge, 2018). Further on, the debottleneck process will be carried out by replacing the cogeneration unit with another with larger scale. Based on the second and third rule, one should avoid (i) replacing the existing unit with another unit with smaller capacity; and (ii) making changes that lead to lower *GLI*. For instance, further increase the capacity scale of cogeneration T2 from 1030 MW input capacity to 1032 MW is not preferable as a lower *GLI* will be obtained. The drop of *GLI* is mainly due to the lower cost effectiveness of these changes (the increase in energy generation is relatively insignificant, thus not worth the high investment). With these rules, all combinatorial feasible bottleneck paths can be formed. However, only those with the longest pathway are considered as this work assumes the decision-makers are “aggressive players” who will opt to improve the system whenever there is opportunity (i.e., not saving up money, but to invest for improvement whenever it is possible). In this work, with the compliance of the aforementioned rules (see Chapter 2), all paths that take five bottleneck stages are considered as the longest paths. For the ease of the reader, all these feasible paths are extracted and presented in the tree representation diagram in Figure 9.11.

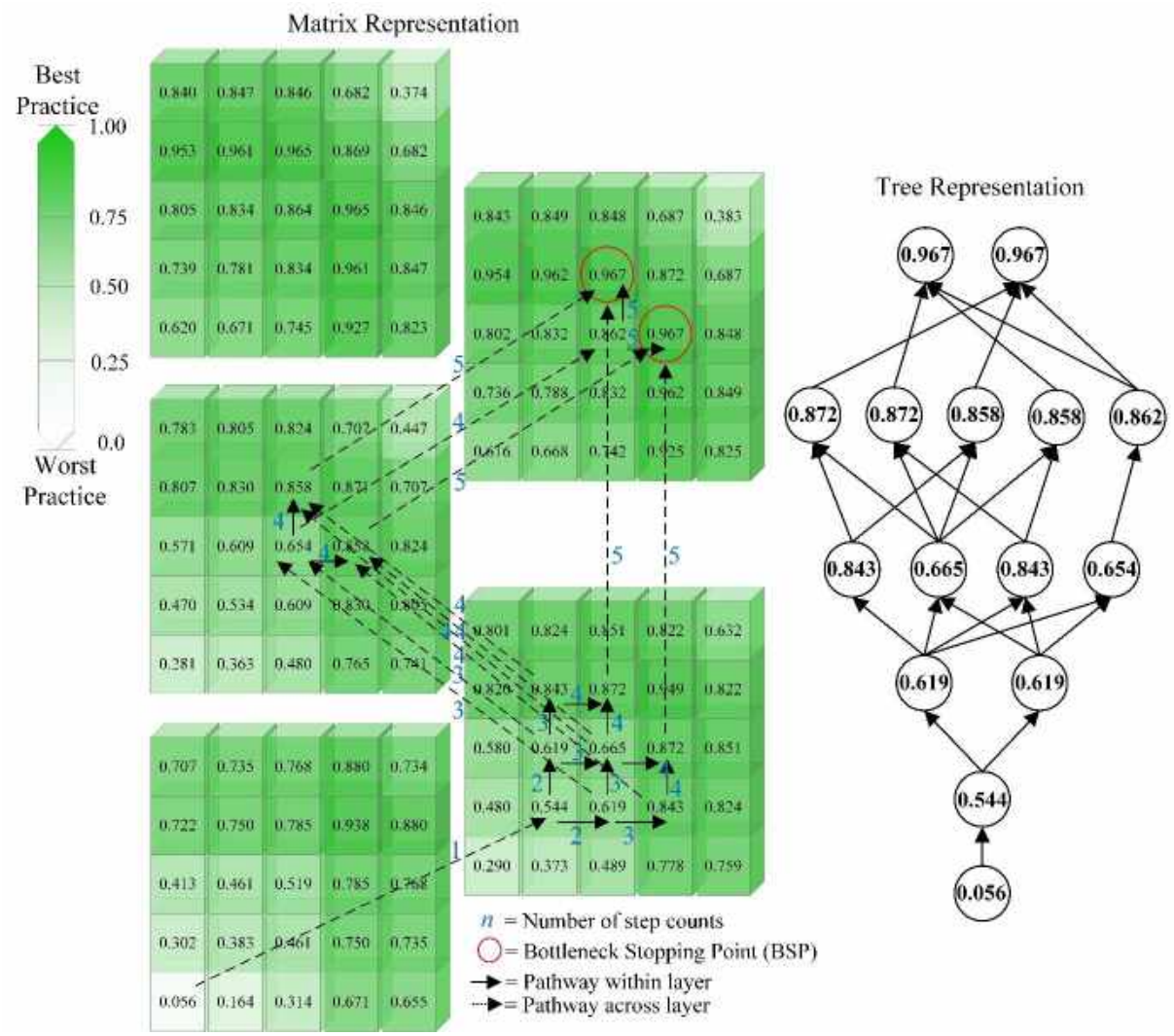


Figure 9.11: Identification of feasible pathways using the bottleneck tree matrix by obeying the pathway rules.

As mentioned, since the cogeneration train 1 and cogeneration train 2 are identical units, there are pathways which are equivalent to each other (e.g., fully utilising the cogeneration train 1 and fully utilising the cogeneration train 2 are two different strategies, but both pathways impose the same *GLI* improvement). Thus, before analysing the feasible paths, these equivalent paths which provide similar *GLI* improvement (i.e., the "faded" paths in Figure 9.12), can be removed. To note, the remaining paths are named as the "effective feasible pathways". They can be further decoupled into five individual pathway strategies (i.e., the "possible pathway solution" listed in Figure 9.12). Despite all five strategies lead to the same BSP state (i.e., *GLI* = 0.967), the different sequences of the bottleneck path will lead to different economic performance in terms of PP (see Chapter 9.2 and Figure 9.14). In terms of attributive elements, the BSP has a 54.2 % improvement in carbon emission per unit power production, 46.3 %

improvement in OpEX, 59.0 % improvement in heat energy production, 58.9 % improvement of power production with an additional investment of 48.7 % of the original capital costs. The industrialists can determine their improvement strategy based on the possible pathway solutions to maximise operation, environment, and economic performances. As debottlenecking of the system takes time, the possible pathway offers the industrialist an insight on potential system upgrade with financial consideration.

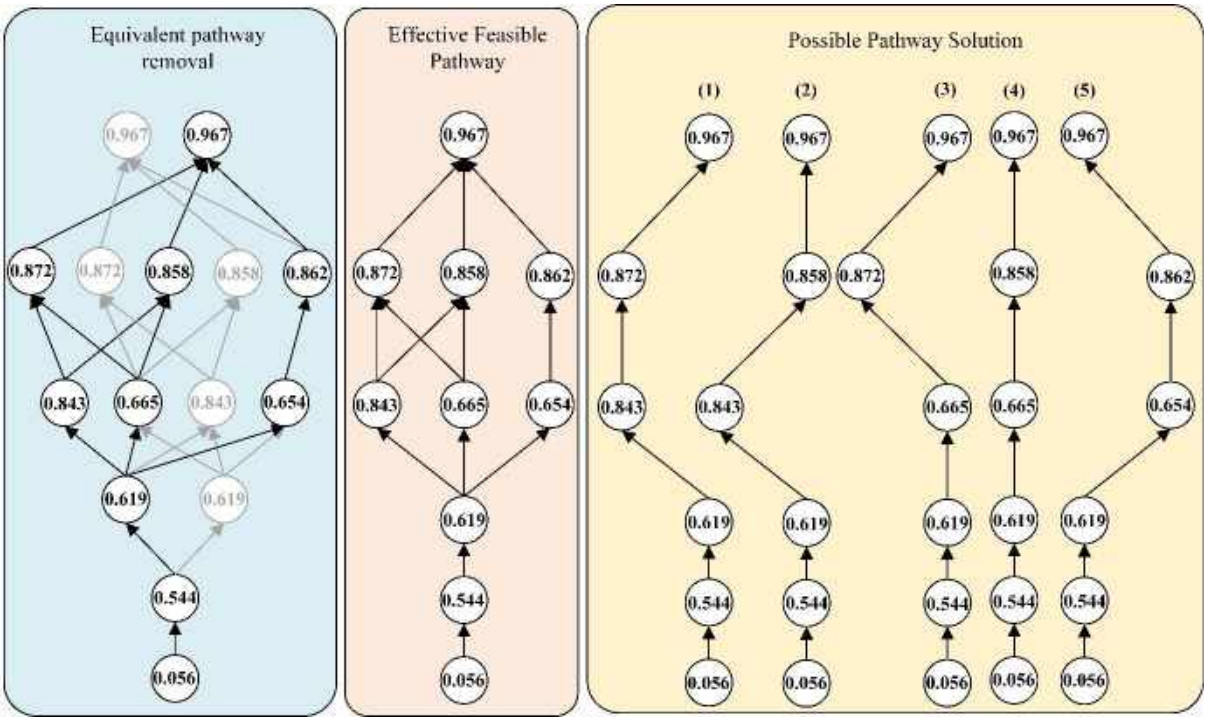


Figure 9.12: Equivalent pathway removal and decomposition of individual pathway solution

As the five strategies lead to a final identical outcome for all possible pathways, the most effective and optimum pathway is preferable. Figure 9.13 summarised a detailed illustration of the de-bottleneck pathway for the case study. The current state of the case study has the *GLI* at 0.056 and best BSP is at 0.967. Again, it is reminded that the aim of the bottlenecking analysis strategy focuses on maximising the operation, commercial and environmental potential. As the configuration of the case study has 2 identical CHP system (i.e. T1 and T2) and smaller capacity CHP system (i.e. T3), the debottleneck of T1 and T2 are identical and improvement on either unit will produce identical results.

The recommended de-bottleneck actions for all debottlenecking pathways are the same for the first 3 steps. The current state (Step 1) reflects a *GLI* of 0.056 can be improved to 0.544 by maximising the capacity of all the CHP systems (Step 2). The pathway has also indicated that

upgrading the generator capacity from maximum capacity to 321 MW for T1 (both are identical units) can improve the *GLI* to 0.619 (Step 3).

The pathway (1) recommended the turbine set of T1 to be upgraded in Step 4 with fuel consumption of 1030 MW with further enhancement of the generator to 397 MW. Such improvement can elevate the *GLI* from 0.619 to 0.843. In Step 5, the generator of T2 is improved to accommodate 321MW electric generation that leads to a *GLI* of 0.872. Lastly, the enhancement of T3 CHP system in Step 6 to improve the capacity of the turbine set with 676 MW fuel consumption, up to 249 MW electricity generation, supplementary firing improved to 391MW with HRSG accommodating 403 MW. This leads the final *GLI* to 0.967.

In pathway (2), Step 4 recommended the same as pathway (1) to improve the turbine set and generator of T1. Step 5 is similar to pathway (1) as well but with an additional upgrade of supplementary firing system for T3 to 340 MW. The *GLI* for Step 5 is improved from 0.843 to 0.858. In Step 6, the upgrades are identical to pathway (1) to focus on upgrading the T3 system.

Moving on to pathway (3), T2 turbine set and generator set are recommended to be upgraded to 856 MW and 321 MW respectively with a *GLI* of 0.665. In Step 5, either T1 or T2 turbine set and generator set need to be further upgraded to 1030 MW and 397 MW respectively. By doing so, the *GLI* will be improved to 0.872. Finally, Step 6 is similar to pathway (1) that yields a *GLI* of 0.967.

Pathway (4) exercises a similar debottlenecking strategy as pathway (3) in Step 4. In the next step, T1 turbine set and generator are suggested to be upgraded with 1030 MW and 397 MW respectively. On top of that, the supplementary firing system of T3 is also upgraded concurrently. This will improve the *GLI* from 0.665 to 0.858. The debottlenecking strategy in Step 6 is identical to pathway (1).

As for pathway (5), Step 4 suggested to upgrade T2 turbine set and generator as Step 4 in pathway (3) with additional enhancement on supplementary firing system of T3. This improvement reflects a *GLI* of 0.654. Moving on, Step 5 focuses on increasing the capacity of the T3 system including upgrading the compressor set, generator, supplementary firing system and HRSG capacity to 676 MW, 249 MW, 391 MW and 403 MW respectively. The upgrading of T1 or T2 turbine and generator to 1030 MW and 397 MW is recommended in Step 6 that leads to *GLI* of 0.967.

Based on the possible debottlenecking pathways, there are several factors that need to be highlighted. In this case study, the T1 and T2 system are identical. The industrialist can select which unit to debottleneck first and define the first debottlenecking unit as T1. The possible pathway is generated to cater to financial capability of the organisation. For instance, pathway 3 and 4 can have lower investment cost in Step 4 as the equipment required upgrade is smaller capacity compared to other possible pathways which also indirectly reflect on the *GLI*. Similarly, for Step 4 in pathway 5, the additional upgrade in the supplementary firing system as compared to pathway 3 and 4 yields lower *GLI*. This is mainly due to the consideration of operational and environmental factors. The system (i.e. turbine and generator as a set) upgrade is considered in the debottlenecking strategy with the understanding of the industry practice. Besides, the industrialist can select the possible pathway according to their financial capability and operation strategy.

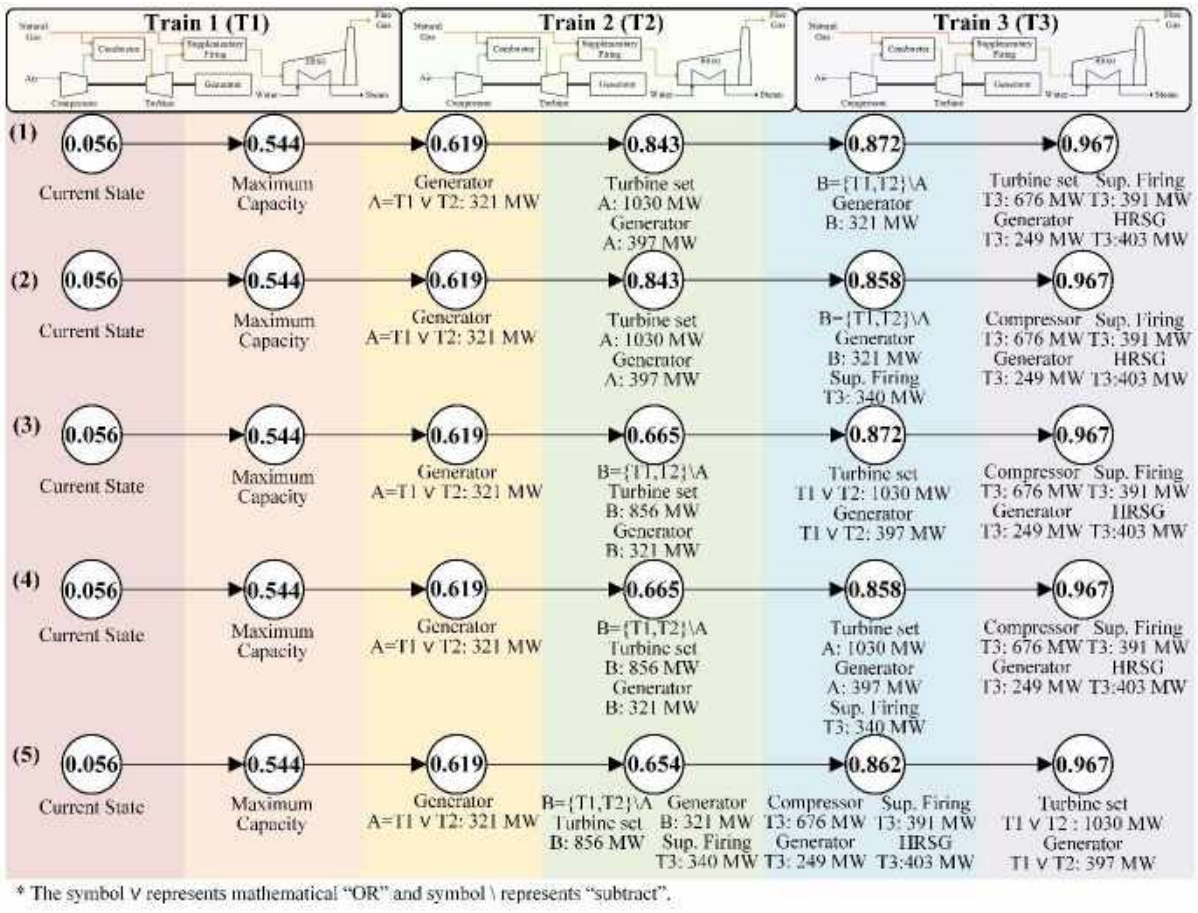


Figure 9.13: Process changes in specific areas with regards to states in pathways.

Figure 9.14 shows the cash flow diagram of all five strategies, where the red dotted lines refer to the cash flow for the base case (without debottlenecking), while the blue solid lines refer to

the estimated cash flow attributed to each strategy. The intercept point of these lines indicates the breakeven point for the investment made. Therefore, the time required to achieve this breakeven point is notified as the PP. Please note that the NPV calculations are made by assuming that all five cases are “zero initial investment cases”, where the elevated profit attributed from the first stage of debottlenecking will be fully reinvested into the second stage of debottlenecking. This causes the “zigzag” pattern on the NPV flow (see Figure 9.14(a) to (e)). Theoretically, the decision maker can directly invest the full amount of capital to achieve BSP in a single step instead of having multiple stages of debottlenecking sequence. However, the “zero initial investment scheme” is a relatively realistic and lucrative strategy for most small and medium enterprises (SMEs) and even for large corporates, especially when budget constraints are significant.

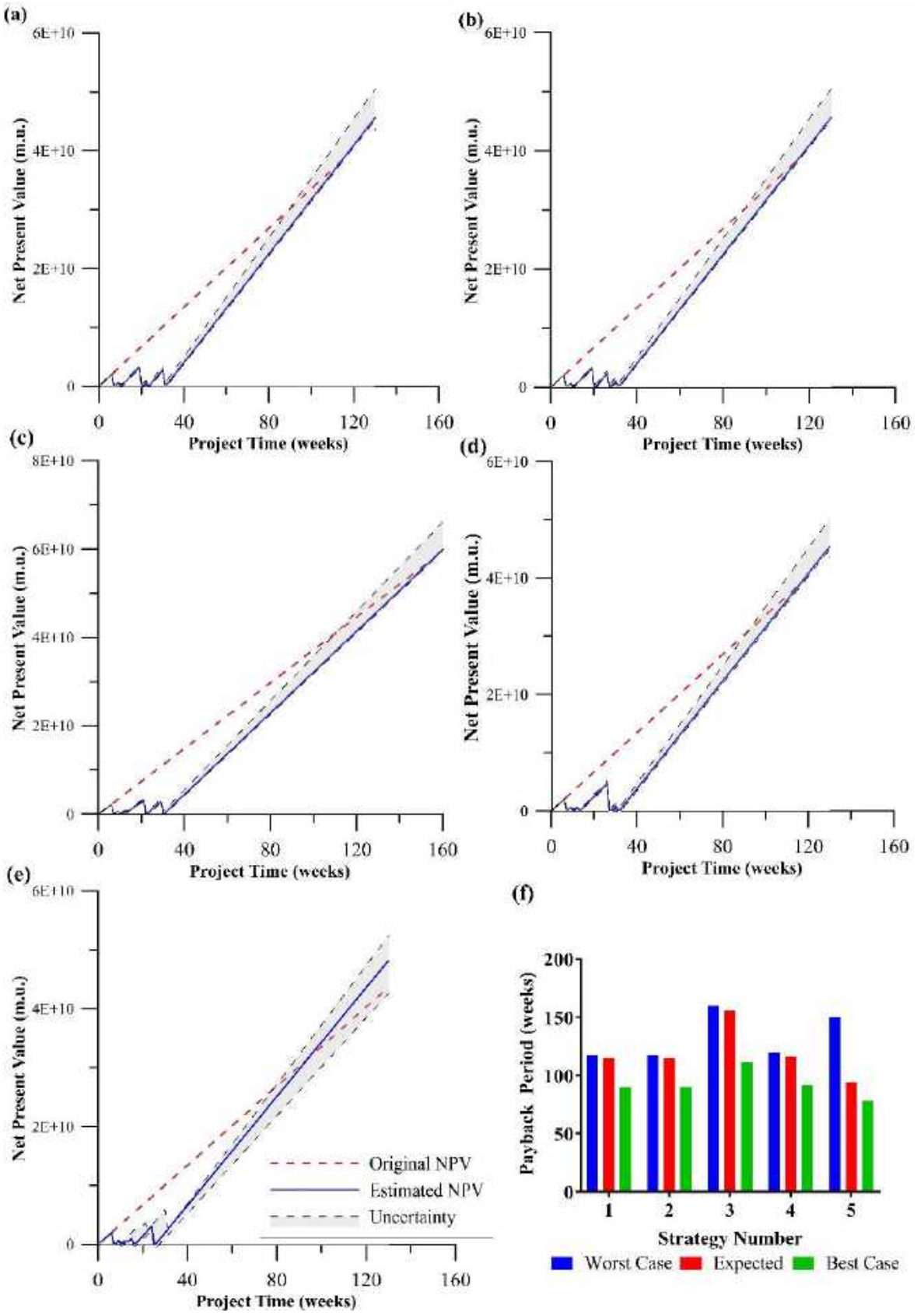


Figure 9.14: Graphical plot for (a) Net present value of strategy 1 (b) Net present value of strategy 2 (c) Net present value of strategy 3 (d) Net present value of strategy 4 (e) Net present value of strategy 5 (f) Payback period of each strategies

Among the five strategies, strategies 1, 2 and 4 provide the most interesting payback periods (PP) with 114.5 weeks, 114.6 weeks, and 115.9 weeks, respectively. On top of the low PP, these strategies are also associated with lower risk (less deviations between the best- and worst-case scenarios). These strategies focus on upgrading the turbine generator set for electricity generation followed by the supplementary firing system at the later stage. Besides, upgrading of the turbine generator set also has an advantage with shorter upgrading downtime which contributes to faster payback. Strategy 5, on the other hand, provides the best expected PP with 93.9 weeks. This shows that the early instalment of larger supplemental firing units can lead to a better NPV growth (due to greater power generation) as compared to the late instalment (as opted in Strategies 1, 2 and 4). However, this strategy associates with a higher uncertainty (the deviations between the best and the worst-case scenarios is large). This higher uncertainty level is attributed to the effects of ambient temperature with regards to the more unreliable small turbine generator set in strategy 5. Moreover, the PP of strategy 5 is directly affected by the fuel efficiency for steam generation compared with steam boiler. Nevertheless, with the increase in fuel cost, the capacity in recovery of thermal energy can give financial advantages to the CHP plant. In contrast, Strategy 3 should never be selected due to the high expected PP with 155.6 weeks. As mentioned above, the timing to upgrade the supplementary firing unit has a significant impact on the PP. In this strategy, the supplementary firing unit is upgraded to the largest capacity at the final debottlenecking stage. The high value investment cost (i.e. turbine set 1030 MW) at the later stage has prolonged the PP of the project. This further restricts the NPV growth of the process and therefore, requires a longer time to payback. The PP for all strategies is highly related to the investment cost and sequence of debottlenecking. With the potential debottlenecking pathway, the industrialist can develop the debottlenecking strategy according to the production demand and financial condition.

9.5 Conclusion

This chapter formally extends the methodology for bottleneck tree analysis (BOTA) towards problems of multi-dimension. This work also demonstrate that process variations can be effectively modelled using a data-driven neural network ensemble method. This highlights that BOTA has a model-agnostic property and can be flexibly used with any type of process models. This improvement of the BOTA technology has been successfully implemented in a multi-train cogeneration plant that is situated in Malaysia, and provides 54.2 % improvement in carbon emission per unit power production, 46.3 % improvement in operating expenditure, 59.0 % improvement in heat energy production, 58.9 % improvement of power production with an

additional investment of 48.7 % of the original capital costs. By taking a zero initial investment case, this work shows that the average payback period (PP) can be between 114 to 116 weeks with low uncertainty (Strategy 1, 2, 4) when de-bottleneck activities are primarily focused on the turbine generator set. Nevertheless, industrialists can also choose to focus de-bottlenecking activities on the supplementary firing unit to give an average PP of 93.9 weeks, however with high uncertainty (worst-case is 150.1 weeks). A pitfall for this approach is that, focusing too much efforts on de-bottlenecking in the supplementary firing unit will lead to detrimental average PP of 155.6 weeks as shown in Strategy 3. The benefit of elevating process capacities with a zero initial investment scheme is very lucrative for most small and medium enterprises (SME) and even for larger corporates with difficulties with funding allocation. This is systematically achievable in multi-train cogeneration plants with BOTA technology. A contribution of this work is that data-driven analytics has been allowed within the application of debottlenecking process systems.

Nomenclature

Abbreviation	Definition
4M1E	Material, Manpower, Money, Machine and Environment
7M	Manpower, Machine, Material, Management, Maintenance, Money, and Mother Nature
AHP	Analytical Hierarchical Process
ANP	Analytic Network Process
BOTA	Bottleneck Tree Analysis
BSP	Bottleneck Stopping Point
CapEx	Capital Expenditures
CHP	Combined Heat and Power
DFIC	Dr. Fromme International Consulting
EU	European Union
Exp	Exponential
FiT	Feed-in Tariff
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GLI	Green and Lean Index (Dimensionless, no unit)
GRA	Grey Relational Analysis

HRSG	Heat Recovery Steam Generator
ISO	International Organisation for Standardization
L&G	Lean and Green
LNG	Liquified Natural Gas
Log	Logarithmic
m.u.	Monetary Unit (depending on currency of country, MYR for this case study)
MCDM	Multi-Criteria Decision-Making
ML	Machine Learning
MMSCFD	Million Standard Cubic Feet Per Day, ISO standard unit for natural gas.
MW	Mega-watt, unit for power, heat energy or potential energy in natural gas.
MYR	Malaysian Ringgit
NIS	Negative Ideal Solution
NPV	Net Present Value
OpEx	Operating Expenditures
P-graph	Process Graph
PIS	Positive Ideal Solution
PP	Payback Period
PSE	Process System Engineering
R^2	Pearson's Correlation Coefficient
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SDG	Sustainable Development Goal
SME	Small and Medium Enterprise
Tn	Refers to cogeneration train number "n". Symbol "n" represents 1,2,3 in this work.
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
UN	United Nations
VIKOR	Višekriterijumsko Kompromisno Rangiranje

Appendix

Definition. Supposed the *d*-dimensional bottleneck tree matrix is expressed as $M(x_0, \dots, x_d)$, each element within M is called the green and lean index (*GLI*) and is a function of the *m*-dimensional attributive elements $M(x) = GLI(x) = f\{e_0(x), \dots, e_m(x)\}$ where f is a multi-criteria decision-making function. The bottleneck stopping point (*BSP*) is defined as $global\ \max(M)$. The *d*-dimensional optimal capacity parameters for the plant is represented as $u^* = [u_0, \dots, u_d] = argmax\ (M)$ and the evolution of *GLI* from origin to *BSP* can be generally expressed as a sequence $GLI = [M(x_0), \dots, M(x_s)]$. The corresponding pathway is $p = [x_0, \dots, x_s]$ where $x_0 = [0, \dots, 0]$ and $x_s = u^*$. The logical constraints of the pathway must follow the following:

$$x_{n+1} = x_n + \left[\Delta x_{n,0}, \dots, \Delta x_{n,s}\right] \quad \forall n \in [0, s]\,,\ \Delta x \in N = \{0,1,2, \dots\} \tag{A9.1}$$

$$M(x_{n+1}) > M(x_n) \tag{A9.2}$$

The first constraint ensures that no reversible pathway exists in any dimension, while second constraint restricts decrease in GLI. We also define the net present value in state *x* and time *t* as $NPV(x,t) = II + \sum NP(x,t) - \sum IC(x,t)$, where *II* is the initial investment, *NP* is the net profit and *IC* is the investment costs. The payback period for pathway *p* is defined as $PP(p) = T$ where $NPV(x_s, T) = II + \sum NP(0,t)$.

Proposition. If the optimisation problem has a single objective of $\max(M)$, the optimal pathway $p^* = [x_0, K, x_s]$ and K is any sequence that follows the logical constraints from Eqn. A9.1 and A9.2. We state that if $x_s \neq x_1 + q$, where q is a *d*-dimensional vector with binary value in any dimension, K is not singular.

Lemma. From proposition, if the problem with $x_s = x_0 + q$ has a single pathway solution, the number of steps, s is at maximum. Otherwise, there are multiple pathway solutions and we formulate an approximated two-step optimisation approach to minimise payback period (*PP*) as the lower objective. From definition, $PP(p) \propto \frac{1}{\sum NP(x,t) - \sum IC(x,t)}$ and since $\sum NP(x,t) - \sum IC(x,t) \propto \sum_n M(x_n)t_n$ and $\forall n \in [0, s]$, $PP(p) \propto \frac{1}{\sum_n M(x_n)t_n}$. For a minimal *PP(p)* and since Eqn. A2 holds, *s* must be maximal.

Proof. Consider a collection of pathways with similar length, $P_L = \{p_j\}$ where $p_j = [x_{0,j}, x_{1,j} \dots, x_{L,j}]$. By design of bottleneck tree matrix, at $x_{2,j}$, $\sum IC(x,t) = 0\ and\ PP(p) \propto \frac{1}{\sum NP(x,t)}$. Consider up to the first two elements of a pathway ($p_{f,2}$) with $x_{1,j} = \max(x_{0,j} + q)$ and that of a pathway otherwise ($p_{s,2}$), $PP(p_{f,2}) < PP(p_{s,2})$. Consider other elements of the

two pathways $p_{f,L}$ and $p_{s,L}$ respectively, if no dimension within the d-dimensional M is redundant, $p_{s,L} \in p_{f,L}$. Thus, a pathway $p = [x_0, \dots, x_s]$ with $x_1 = \max(x_0 + q)$ is the secondary condition for $PP(p)$ to be minimal. With this and from lemma, both conditions hold and thus completes the proof.

References

Abbasi, M., Deymi-Dashtebayaz, M., Farzaneh-Gord, M., Abbasi, S., 2015. Assessment of a CHP system based on economical, fuel consumption and environmental considerations. Int. J. Glob. Warm. https://doi.org/10.1504/IJGW.2015.067757

Abd Majid, M.A., Ghazali, Z., Shin Min, N.T., 2014. Techno-economic Evaluation on Enhancing Cogeneration Plant Capacity: Case Study of Palm Oil Mill Cogeneration Plant. J. Appl. Sci. https://doi.org/10.3923/jas.2014.285.290

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Ahmad, S., Hui, D.C.W., 1991. Heat recovery between areas of integrity. Comput. Chem. Eng. https://doi.org/10.1016/0098-1354(91)80027-S

Ahmadi, G., Toghraie, D., Akbari, O., 2019. Energy, exergy and environmental (3E) analysis of the existing CHP system in a petrochemical plant. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2018.10.009

Aklilu, B.T., Gilani, S.I., 2010. Mathematical modeling and simulation of a cogeneration plant. Appl. Therm. Eng. https://doi.org/10.1016/j.applthermaleng.2010.07.005

Al-Sulaiman, F.A., Hamdullahpur, F., Dincer, I., 2011. Trigeneration: A comprehensive review based on prime movers. Int. J. Energy Res. https://doi.org/10.1002/er.1687

Asan, U., Soyer, A., Serdarasan, S., 2012. A Fuzzy Analytic Network Process Approach, in: Computational Intelligence Systems. Atlantis Press, Paris, pp. 155–179.

Azhdari, A., Ghadamian, H., Ataei, A., Yoo, C.K., 2009. A new approach for optimization of combined heat and power generation in edible oil plants. J. Appl. Sci. https://doi.org/10.3923/jas.2009.3813.3820

Bagajewicz, M., Rodera, H., 2002. Multiple plant heat integration in a total site. AIChE J. https://doi.org/10.1002/aic.690481016

Behun, M., Gavurova, B., Tkacova, A., Kotaskova, A., 2018. THE IMPACT OF THE MANUFACTURING INDUSTRY ON THE ECONOMIC CYCLE OF EUROPEAN UNION COUNTRIES. J. Compet. https://doi.org/10.7441/joc.2018.01.02

Bhatia, S.C., 2014. Cogeneration, in: Advanced Renewable Energy Systems. Elsevier, pp. 490–508. https://doi.org/10.1016/B978-1-78242-269-3.50019-X

Braun, M.A., Seijo, S., Echanobe, J., Shukla, P.K., del Campo, I., Garcia-Sedano, J., Schmeck, H., 2016. A neuro-genetic approach for modeling and optimizing a complex cogeneration process. Appl. Soft Comput. J. https://doi.org/10.1016/j.asoc.2016.07.026

Brown, G., Wyatt, J.L., Tiňo, P., 2005. Managing diversity in regression ensembles. J. Mach. Learn. Res.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 2018. Graphical methods for data analysis, Graphical Methods for Data Analysis. https://doi.org/10.1201/9781351072304

Danny Harvey, L.D., 2012. A handbook on low-energy buildings and district-energy systems fundamentals, techniques and examples, A Handbook on Low-Energy Buildings and District-Energy Systems: Fundamentals, Techniques and Examples. https://doi.org/10.4324/9781849770293

DFIC, 2016. How to Produce Energy Efficiently: A practical guide for experts in emerging and developing economies.

Douglas, J.M., 1985. A hierarchical decision procedure for process synthesis. AIChE J. https://doi.org/10.1002/aic.690310302

Dow Jones, 2007. Sustainability indexes.

EPA, 2020. CHP Benefits:Combined Heat and Power (CHP) Partnership.

Eraslan, S., Çağman, N., 2017. A Decision Making Method by Combining Topsis And Grey Relation Method Under Fuzzy Soft Sets. Sigma J. Eng. Nat. Sci. 8, 53–64.

Fang, T., Lahdelma, R., 2016. Optimization of combined heat and power production with heat storage based on sliding time window method. Appl. Energy. https://doi.org/10.1016/j.apenergy.2015.10.135

Fast, M., Palmé, T., 2010. Application of artificial neural networks to the condition monitoring and diagnosis of a combined heat and power plant. Energy. https://doi.org/10.1016/j.energy.2009.06.005

Franco, A., Bellina, F., 2018. Methods for optimized design and management of CHP systems for district heating networks (DHN). Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2018.07.009

Friedler, F., Tarján, K., Huang, Y.W., Fan, L.T., 1992. Graph-theoretic approach to process synthesis: axioms and theorems. Chem. Eng. Sci. 47, 1973–1988. https://doi.org/10.1016/0009-2509(92)80315-4

Fuentes-Cortés, L.F., Ponce-Ortega, J.M., Nápoles-Rivera, F., Serna-González, M., El-Halwagi, M.M., 2015. Optimal design of integrated CHP systems for housing complexes. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2015.04.036

Ge, Z., 2017. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemom. Intell. Lab. Syst. https://doi.org/10.1016/j.chemolab.2017.09.021

Gimelli, A., Muccillo, M., 2013. Optimization criteria for cogeneration systems: Multi-objective approach and application in an hospital facility. Appl. Energy. https://doi.org/10.1016/j.apenergy.2012.11.076

Gu, C., Xie, D., Sun, J., Wang, X., Ai, Q., 2015. Optimal operation of combined heat and power system based on forecasted energy prices in real-time markets. Energies. https://doi.org/10.3390/en81212427

Hansen, L.K., Salamon, P., 1990. Neural Network Ensembles. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/34.58871

Hashim, Z.B., Nabil, M., Muhtazaruddin, B., Binti, M., Majid, A., 2010. Application of Combined Heat and Power in Malaysia Industrial Sector. Univ. Teknol. Malaysia.

Hodge, G., 2018. Points to Consider in Manufacturing Operations, in: Development, Design, and Implementation of Manufacturing Processes. Elsevier, Amsterdam, The Netherlands, pp. 987–998.

How, B.S., Yeoh, T.T., Tan, T.K., Chong, K.H., Ganga, D., Lam, H.L., 2018. Debottlenecking of sustainability performance for integrated biomass supply chain: P-graph approach. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.04.240

Hwang, C., Yoon, K., 1981. Multiple Attribute Decision Making: Methods and Applications, A State of the Art Survey, Springer-Verlag. https://doi.org/10.1007/978-3-642-48318-9

Hwang, C.L., Lai, Y.J., Liu, T.Y., 1993. A new approach for multiple objective decision making. Comput. Oper. Res. https://doi.org/10.1016/0305-0548(93)90109-V

IRP, 2017. A report of the International Resources Panel, in: Green Technology Choices: The Environmental and Resources Implications of Low-Carbon Technologies. Nairobi, Kenya, p. 76.

Kanbur, B.B., Xiang, L., Dubey, S., Choo, F.H., Duan, F., 2019. Sustainability and thermoenvironmental indicators on the multiobjective optimization of the liquefied natural gas fired micro-cogeneration systems. Chem. Eng. Sci. https://doi.org/10.1016/j.ces.2019.03.024

KHI, 2018. Technologies for Supporting the Operation and Maintenance of Power Generation Facilities: Kawasaki Heavy Industries.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.

Kirubakaran, B., Ilangkumaran, M., 2016. Selection of optimum maintenance strategy based on FAHP integrated with GRA–TOPSIS. Ann. Oper. Res. https://doi.org/10.1007/s10479-014-1775-3

Kohl, A.L., Nielsen, R.B., 1997. Thermal and Catalytic Conversion of Gas Impurities, in: Gas Purification. https://doi.org/10.1016/b978-088415220-0/50013-6

Koulouris, A., Calandranis, J., Petrides, D.P., 2000. Throughput analysis and debottlenecking of integrated batch chemical processes. Comput. Chem. Eng. 24, 1387–1394. https://doi.org/10.1016/S0098-1354(00)00382-3

Lenz, J., Wuest, T., Westkämper, E., 2018. Holistic approach to machine tool data analytics. J. Manuf. Syst. https://doi.org/10.1016/j.jmsy.2018.03.003

Leong, W.D., Lam, H.L., Ng, W.P.Q., Lim, C.H., Tan, C.P., Ponnambalam, S.G., 2019a. Lean and Green Manufacturing—a Review on its Applications and Impacts. Process Integr. Optim. Sustain. https://doi.org/10.1007/s41660-019-00082-x

Leong, W.D., Teng, S.Y., How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P., Ponnambalam, S.G., 2019b. Adaptive analytical approach to lean and green operations. J. Clean. Prod. 235, 190–209. https://doi.org/10.1016/j.jclepro.2019.06.143

Li, W., Tian, X., Li, Y., Ma, Y., Fu, L., 2018. Combined heating operation optimization of the novel cogeneration system with multi turbine units. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2018.06.015

Liliana, L., 2016. A new model of Ishikawa diagram for quality assessment. IOP Conf. Ser. Mater. Sci. Eng. 161. https://doi.org/10.1088/1757-899X/161/1/012099

Linnhoff, B., Boland, D., 1982. A user guide on process integration for the efficient use of energy. https://doi.org/10.1016/0300-9467(83)80027-6

Litzen, D.B., Bravo, J.L., 1999. Uncover Low-Cost Debottlenecking Opportunities. Chem. Eng. Prog. 95, 25–32.

Mardani, A., Jusoh, A., Nor, K.M.D., Khalifah, Z., Zakwan, N., Valipour, A., 2015. Multiple criteria decision-making techniques and their applications - A review of the literature from 2000 to 2014. Econ. Res. Istraz. . https://doi.org/10.1080/1331677X.2015.1075139

Monfort, E., Mezquita, A., Vaquer, E., Celades, I., Sanfelix, V., Escrig, A., 2014. Ceramic Manufacturing Processes: Energy, Environmental, and Occupational Health Issues, in: Comprehensive Materials Processing. https://doi.org/10.1016/B978-0-08-096532-1.00809-8

Naftaly, U., Intrator, N., Horn, D., 1997. Optimal ensemble averaging of neural networks. Netw. Comput. Neural Syst. https://doi.org/10.1088/0954-898x_8_3_004

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. J. Big Data. https://doi.org/10.1186/s40537-014-0007-7

Ngan, S.L., How, B.S., Teng, S.Y., Leong, W.D., Loy, A.C.M., Yatim, P., Promentilla, M.A.B., Lam, H.L., 2020. A hybrid approach to prioritize risk mitigation strategies for biomass polygeneration systems. Renew. Sustain. Energy Rev. 121, 109679. https://doi.org/10.1016/j.rser.2019.109679

Nolan, D.P., 2017. Classified Area Pump Installations, in: Fire Pump Arrangements at Industrial Facilities. Elsevier, pp. 161–167.

Omer, A.M., 2008. Energy, environment and sustainable development. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2007.05.001

OpenWeatherMap, 2020. Current Weather And Forecast - Openweathermap [WWW Document].

Pehnt, M., Praetorius, B., Schumacher, K., Fischer, C., Schneider, L., Cames, M., Voß, J.P., 2006. Micro cogeneration: Towards decentralized energy systems, Micro Cogeneration: Towards Decentralized Energy Systems. https://doi.org/10.1007/3-540-30821-0

PEI, 2018. MHPS tests gas turbine using hydrogen fuel mix.

Peng, A., Wang, Z., 2011. GRA-based TOPSIS decision-making approach to supplier selection with interval number, in: Proceedings of the 2011 Chinese Control and Decision Conference, CCDC 2011. https://doi.org/10.1109/CCDC.2011.5968478

Peters, M.S., Timmerhaus, K.D., West, R.E., 2003. Plant Design and Economics for Chemical Engineers 5th edition, McGraw-Hill.

Reverberi, A., Del Borghi, A., Dovì, V., 2011. Optimal design of cogeneration systems in industrial plants combined with district heating/cooling and underground thermal energy storage. Energies. https://doi.org/10.3390/en4122151

Rist, J.F., Dias, M.F., Palman, M., Zelazo, D., Cukurel, B., 2017. Economic dispatch of a single micro-gas turbine under CHP operation. Appl. Energy. https://doi.org/10.1016/j.apenergy.2017.05.064

Sarkar, D.K., 2015. Gas Turbine and Heat Recovery Steam Generator, in: Thermal Power Plant. pp. 239–283.

SDG, 2019. Sustainable Development Goals: Sustainable Development Knowledge Platform [WWW Document]. Sustain. Dev. Goals. URL https://sustainabledevelopment.un.org/?menu=1300 (accessed 5.2.20).

Siddiqui, O., Dincer, I., 2019. A Comparative Life-Cycle Assessment of Two Cogeneration Plants. Energy Technol. https://doi.org/10.1002/ente.201900425

Sinnott, R.K., Towler, G., 2013. Chemical Engineering Design, Chemical Engineering Design. https://doi.org/10.1016/C2009-0-61216-2

Smith, R., 2005. Chemical Process Design and Integration, John Wiley & Sons, Ltd. https://doi.org/10.1529/biophysj.107.124164

Sylajakumari, P.A., Ramakrishnasamy, R., Palaniappan, G., 2018. Taguchi grey relational analysis for multi-response optimization of wear in co-continuous composite. Materials (Basel). https://doi.org/10.3390/ma11091743

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Lam, H.L., 2020. Bottleneck Tree Analysis (BOTA) with green and lean index for process capacity debottlenecking in industrial refineries. Chem. Eng. Sci. 214. https://doi.org/10.1016/j.ces.2019.115429

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Siang Cheah, A.C., Motavasel, Z., Lam, H.L., 2019a. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. J. Clean. Prod. 225, 359–375. https://doi.org/10.1016/j.jclepro.2019.03.272

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F., Máša, V., 2019b. Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresour. Technol. 292, 121971. https://doi.org/10.1016/j.biortech.2019.121971

The World Bank, 2020. Manufacturing, value added data.

TMR, 2020. World-Class OEE – Set OEE Targets To Drive Improvement.

Van Duc Long, N., Lee, M., 2011. Improved energy efficiency in debottlenecking using a fully thermally coupled distillation column, in: Asia-Pacific Journal of Chemical Engineering. https://doi.org/10.1002/apj.577

Vondra, M., Touš, M., Teng, S.Y., 2019. Digestate Evaporation Treatment in Biogas Plants: A Techno-economic Assessment by Monte Carlo, Neural Networks and Decision Trees. J. Clean. Prod. 117870. https://doi.org/10.1016/j.jclepro.2019.117870

Wang, P., Zhu, Z., Wang, Y., 2016. A novel hybrid MCDM model combining the SAW, TOPSIS and GRA methods based on experimental design. Inf. Sci. (Ny). https://doi.org/10.1016/j.ins.2016.01.076

Wang, Q., Peng, A., 2010. Developing MCDM approach based on GRA and TOPSIS. Appl. Mech. Mater. 34–35, 1931–1935. https://doi.org/10.4028/www.scientific.net/AMM.34-35.1931

Zare, A.D., Saray, R.K., Mirmasoumi, S., Bahlouli, K., 2019. Optimization strategies for mixing ratio of biogas and natural gas co-firing in a cogeneration of heat and power cycle. Energy. https://doi.org/10.1016/j.energy.2019.05.182

Zhang, L.-J., Li, Z.-J., 2006. Gene Selection for Classifying Microarray Data Using Grey Relation Analysis, in: Discovery Science. Springer-Verlag, Berlin, Germany, pp. 378–382.

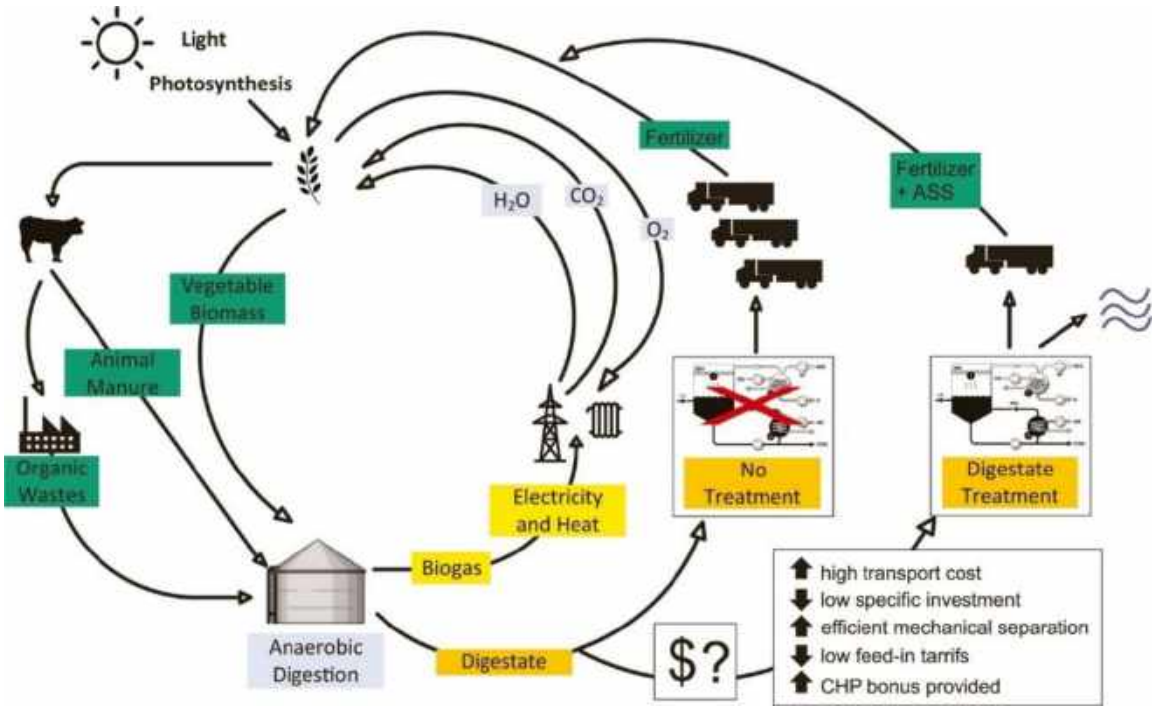
Zhou, L., Tokos, H., Krajnc, D., Yang, Y., 2012. Sustainability performance evaluation in industry by composite sustainability index. Clean Technol. Environ. Policy. https://doi.org/10.1007/s10098-012-0454-9

CHAPTER 10 DECISION-MAKING FOR IMPLEMENTING NEW TECHNOLOGIES IN EXISTING PLANTS: TECHNO-ECONOMIC ANALYSIS

This paper has been peer-reviewed and published in Journal of Cleaner Production.

Vondra, M., Touš, M. and Teng, S.Y., 2019. Digestate evaporation treatment in biogas plants: A techno-economic assessment by Monte Carlo, neural networks and decision trees. Journal of Cleaner Production, 238, 117870.

Graphical Abstract



Abstract

Biogas production is one of the most promising pathways toward fully utilizing green energy within a circular economy. The anaerobic digestion process is the industry standard technology for biogas production due to its lowered energy consumption and its reliance on microbiology. Even in such an environmental-friendly process, liquid digestate is still produced from the remains of digested bio-feedstock and will require treatment. With unsuitable treatment procedure for liquid digestate, the mass of bio-feedstock can potentially escape the circular supply chain within the economy. This work recommends the implementation of new evaporator systems to provide a sustainable liquid digestate treating mechanism within the economy. Studied evaporator systems are represented by vacuum evaporation in combination with ammonia scrubber, stripping, and reverse osmosis. Nevertheless, complex multi-dimensional decisions should be made by stakeholders before implementing such systems. This

work utilizes a novel techno-economics model to study the techno-economics robustness in implementing recent state-of-art vacuum evaporation systems with exploitation of waste heat from combined heat and power (CHP) units in biogas plants (BGP). To take into the account the stochasticity of the real world and robustness of the analysis, this work used the Monte-Carlo simulation technique to generate more than 20,000 of different possibilities for the implementation of the evaporation system. Favourable decision pathways are then selected using a novel methodology which utilizes the artificial neural network and a hyper-optimized decision tree classifier. Two pathways that give the highest probability of providing a fast payback period are identified. Descriptive statistics are also used to analyse the distributions of decision parameters that lead to success in implementing the evaporator system. The results highlighted that integration of evaporation system are favourable when transport costs and incentives for CHP units are large and while feed-in tariffs for electricity production and specific investment costs are low. The result of this work is expected to pave the way for BGP stakeholders and decision makers in implementing liquid digestate treating technologies within the currently existing infrastructure.

Keywords: Anaerobic Digestion; Machine Learning; Vacuum Evaporation; Liquid Digestate; Biogas Plant; Energy Consumption; Nutrient Recovery; Circular economy; Ammonium sulphate solution

10.1 Introduction

Biological processes have the inherent characteristics of preserving resources at a low cost and eco-friendly manner (Liguori and Faraco, 2016). One of the most attractive sources of energy from biological processes is the anaerobic digestion process (Ragazzi et al., 2017). It can effectively convert organic waste to biogas and can serve as a sustainable decentralised energy source even in developing countries (Azouma et al., 2018). The current boom of biogas plants (BGP) in Europe has led the Union to produce 61 TWh of electricity and 26.6 PJ of heat in the year 2015 (Scarlat et al., 2018). The ecosystem is causing biogas processing facilities to rapidly shift towards a circular economy (Fagerström et al., 2018). Biogas processing plants produce a large amount of leftover of digestate by-products from the bioreactor that must be properly processed (Al Seadi et al., 2013). This is a critical issue for sustainability development, as bio-resources can potential "escape" from the circular economy, giving net waste generation. Venkata Mohan et al. (2016) discussed that the utilisation of bio-waste as renewable feedstock

can provide potential in developments towards the circular economy by enhancing bio-refinery competitiveness and social acceptance.

Anaerobic digestion is the industrial standard biogas production method. Waste biomass such as animal manure and crops are collected and fed to the anaerobic bioreactor. The slurry of biomass feedstock is digested by anaerobic microorganisms to produce biogas. In order to maintain a stable and high-quality biogas production, the feedstock's C/N ratio is ideally between 20-30 (Gómez et al., 2006). It was also reported that an increasing amount of long-chain fatty acid will also cause the flocculation of biomass and act as inhibition (Dasa et al., 2016). The sensitive nature of the anaerobic digestion process has encouraged many innovations in the research field, leading to the development of multiple types of bioreactors. Uçkun Kiran et al. (2016) stated that bioreactor configuration should depend on the solid content of feedstocks, where low solid content bioreactor includes anaerobic lagoons, completely mixed reactors, anaerobic filter reactors, fluidized bed reactors, upflow anaerobic sludge blanket reactors, anaerobic baffled reactors and membrane reactors. The work also recommended plug flow reactors, completely mixed reactors, and contact reactors for medium solid content; while plug flow, completely mixed and leach-bed reactors can be used for high solid content feedstock. A review from Zhang et al. (2016) has shown that the research for anaerobic digestion has blossomed beautifully, covering detailed research on substrate analysis, pre-treatment, bioreactor configuration, operation, and system analysis.

On the other hand, the fast-growing anaerobic digestion industry has opened new opportunities for digestate utilisation (Monlau et al., 2015). The work studied the possibility of digestate utilisation by algae cultivation, energy production using biological or thermal process and pyrolysis. Works of Fuchs and Drosg (2013) even refers to digestate utilisation as the bottleneck of the biogas industry (see Figure 10.1). They concluded that digestate processing plant of European countries such as Austria, Germany, Switzerland, and Italy require an appropriate method for disposal of residue digestate. The industrial development of digestate to produce highly enriched nutrient was also recommended (Fuchs and Drosg, 2013). Life-cycle assessment of biogas production carried out by Berglund and Börjesson (2006) concluded with a set of transportation distance limits for different raw material to prevent negative energy balance. This indicates that dewatering (to reduce transportation load) is crucial in maintaining the net energy gain within the circular economy.

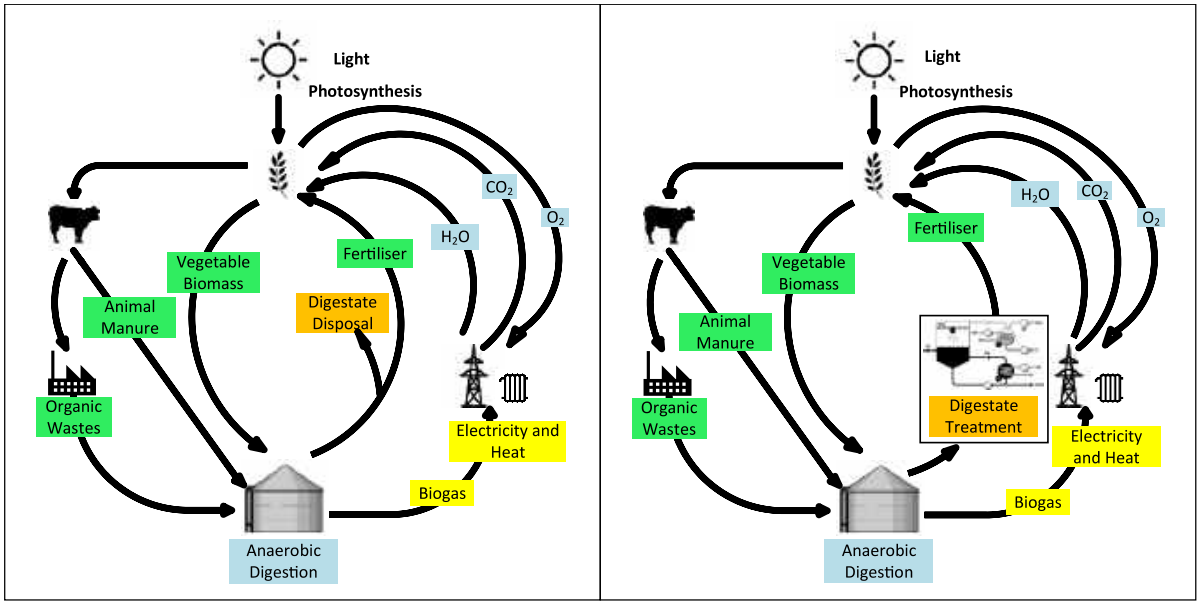


Figure 10.1: Digestate Treatment and the Circular Economy of Biogas Industry

Kirchherr et al. (2017) conceptualized the circular economy with core principles of the 4R framework (Reduce, Reuse, Recycle and Recover), waste hierarchy and system perceptive. Key strategies to achieve circular economy from the linear economy are proposed to include creating useful applications of materials, extending the lifespan of products and parts, innovating smarter product use and manufacturing (Potting et al., 2017). In this context, the digestate by-products can be treated and transformed into bio-resources that are able to contribute to the soil, horticulture, renewable energy and bio-products within the circular economy (van Haeff, 2015). To accelerate this progress, encouraging efforts have been carried out by the European Commission in terms of funding for environmental and circular economy research (European Commission, 2019a). In the European Union, biogas factories are actively transitioning to complete the circular economy by improving the digestate treatment process (Fagerström et al., 2018). From a techno-economical viewpoint, integration of digestate treatment is not always feasible for all BGPs (Vondra et al., 2018b), more industrial incentives are required to make the project favourable by industry players and stakeholders.

The design of digestate treatment system is technically challenging and requires multiple dimensions of considerations. A recent review (Monfet et al., 2018) concluded that there are 14 possible categories of processes for digestate treatment. Those 14 processes are: thickening, dewatering, membrane filtration, struvite precipitation, ammonia stripping, ammonia oxidation, microwave radiation, pH adjustment, chemical hydrolysis, thermal drying,

combustion, pyrolysis, gasification, and ultrasound. Vaneeckhaute et al. (2017) shown that matured nutrient recovery technologies for digestates include phosphorus crystallization or extraction, ammonia stripping, acid air scrubber, membrane filtration, ammonia/phosphorus sorption, macrophytes or microalgae production and evaporation. The selection decision of these processes can be carried out with decision tools such as Biomethane-Calculator (Miltner et al., 2013) and CDPQ Decision Tools (2018). Works of Rehl and Müller (2011) also shown a comprehensive life cycle assessment comparison of biogas digestate processing technologies that can be used for process selection. Tampio et al. (2016) demonstrated that an approach of mass, nutrient and energy balance can be applied to digestate treatment systems. These methods can only select the category of which process to be used. Much more research and industrial experience are required to implement the correct system design. Significant works from various authors and their contribution to the current digestate treatment research can be found in Table 10.1:

Table 10.1: Significant works for the development and implementation of digestate treatment technologies

Work	Contribution
Chen et al. (2009)	Highlighted ecological and environmental protection type disposal of anaerobic digestate for large-scale BGPs.
Fuchs and Drosg (2013)	Conveyed the development in state-of-art industrial-scale technologies. Shown that industrial applicable technologies mainly consist of a combination of membrane treatment, evaporation and stripping technologies.
Al Seadi et al. (2013)	Proposed that digestate quality can be affected by feedstock of anaerobic digestion and shown that consideration of such variability is critical.
Golkowska et al. (2014)	Compared the treatment costs of four different digestate treatment technologies. Focused mainly on drying, pelletizing, reverse osmosis, and denitrification.
Guercini et al. (2014)	Exploited the excess heat within the combined heat and power (CHP) unit of a BGP to a single-effect vacuum evaporator.
Koszel and Lorencowicz (2015)	Highlighted the use of biogas digestate as a replacement for agricultural fertilizer.

Table 10.1 (Cont.): Significant works for the development and implementation of digestate treatment technologies

Work	Contribution
Dahlin et al. (2015)	**Discussed the importance and barriers of marketing digestate products from a qualitative perspective.**
Vondra et al. (2016)	Compared various designs of vacuum evaporators for the application of digestate treatment. Highlighted that multi-stage flash vacuum evaporator performs most efficiently.
Xia and Murphy (2016)	Demonstrated a front-end research on utilizing microalgae cultivation in treating liquid digestate from biogas systems.
Hung et al. (2017)	Front-end research demonstrating that treated digestate can be converted into valuable material such as biochar. Demonstrated the future possibilities of digestate-derived products.
Vondra et al. (2018b)	Proposed the integration of multi-stage flash vacuum evaporator in a BGP to exploit the excess heat within CHP unit.
Bolzonella et al. (2018)	Re-emphasized the importance of digestate treatment with estimated costs for non-vacuum dryers, stripping and membrane technologies.

For liquid digestate (LD) treatment using dewatering methodologies, previous work (Vondra et al., 2016) has demonstrated the technical potential of LD thickening with the considerations of multiple industrial-standard evaporator systems. Chiumenti et al. (2013) tested the effectiveness of a two-stage configuration vacuum evaporator pilot plant in a 1 MWe anaerobic plant. The work studied the temperature, pH, total solid, volatile solids, total Kjeldhal nitrogen and acid consumption of the evaporation process. Full-scale evaporator plant is estimated to have energy consumption at 5-8 kWhe/m^3 of digestate and 350 kWht/m^3 of evaporated water (Chiumenti et al., 2013). A case study in Croatia has shown the high energy potential of treated digestate (Đurđević et al., 2018), which can be achieved using evaporation methodologies. This solution was also implemented in France where a 23,000 ton/year digestate evaporation plant was commissioned (Bamelis et al., 2015), showing that evaporation methodologies have become the industrial standards for digestate treatment. In a recent work, the techno-economic assessment of integrating a novel vacuum evaporator into a conventional BGP was studied

(Vondra et al., 2018b). From that work, it was concluded that the design decision and consideration of integrating the evaporator in a BGP can vary for each situation. As in any industrial processes, new technologies must be considered in terms of energy intensity, environmental impact, and financial aspects (Máša et al., 2013). Economic feasibility of integration depends on many factors. Besides the investment cost, it is digestate transport and application cost, dry matter content in liquid digestate, energy consumption of an evaporator system and many others. This work will guide the reader on the implementation of evaporator systems based on the criteria in the BGP.

The main objective of this work is to analyse the integration of an evaporation technology for digestate treatment into BGP regarding investment. The analysis provides important results subsequently used in a tool to support decision making on investment. The tool is developed for BGP owners to make decisions on investing in evaporator technology. For this purpose, a complex mathematical model is presented and rigorous simulations (possible scenarios) for large range of inputs are analysed. As a secondary objective, the study aims to investigate the significance of individual inputs in terms of the overall payback period of the project. The novelty of this work is that it analyses the implementation of current state-of-art industrial-scale digestate treatment technologies (evaporation systems). As the implementation of such technologies requires complex multi-dimensional considerations (Monfet et al, 2018), a novel techno-economic assessment methodology is introduced which utilizes Monte-Carlo sampling, artificial neural network (ANN) and hyper-optimized decision tree to search for the optimal policies and scenarios to implementing such technologies.

10.2 Methodology

This chapter described the methodology for the scope of the system, system model, Monte Carlo simulation, neural network variable selection, optimal policy search using decision tree, and descriptive statistics.

10.2.1. Introduction

The methodology and scientific procedure of this study are shown in Figure 10.2. Individual steps are described in the text below. At first, the BGP system was studied and clearly defined (Chapters 10.2.2, 10.2.3). Based on this definition, an analytical steady state techno-economic model was established (Chapter 10.2.4) and crucial input variables for further detailed research were identified. A thorough survey has revealed relevant ranges of values for a particular

operational variable (Table 10.2), which could occur in practice. Since there are no typical individual values which could be representative for a significant number of plants, this work has decided to generate combinations of variables within allowable ranges. In this way, it was possible to simulate and analyse a large number of operating states that could appear in real BGPs.

Two procedures for the generation of particular values of given variables were subsequently defined. During the "no-rules based" procedure, input variables could reach any value within the given range. The "rules-based" procedure introduces restrictions, which do not allow certain combinations of variables to enter the model since their appearance in a real BGP is improbable. Both procedures implement the Monte Carlo simulation method for the generation of datasets and are thoroughly described in Chapter 10.2.5. Each dataset represents a particular scenario (Base cases, "CHP bonus", "No ASS sale") which are also defined in Chapter 10.2.5. The aim of the subsequent data analysis was to provide support for decision making (decision analysis) and to identify trends for technology implementation (exploratory data analysis).

10.2.2 System Description

This study considers a typical agricultural BGP, which processes different biomass substrates for biogas production and subsequent heat and electricity generation. A major part of the electricity, not consumed by the BGP itself, is sold to the power grid and represents the main source of income. The produced heat is partly supplied to digesters for the anaerobic fermentation process, eventually to other thermal appliances (e.g. drying). In most cases, the produced heat is not fully utilized (Guercini et al., 2014) and needs to be wasted by air-cooled chillers to ensure a sufficiently low temperature of the recirculating cooling water. Digestate serves as valuable fertilizer, but its storage and transportation are associated with significant costs for operators. To save storage capacities, digestate is often subjected to mechanical separation, which reduces the volume of liquid fractions that needs to be stored in specialized tanks. However, the overall volume of fertilizer (liquid + solid fraction) stays the same.

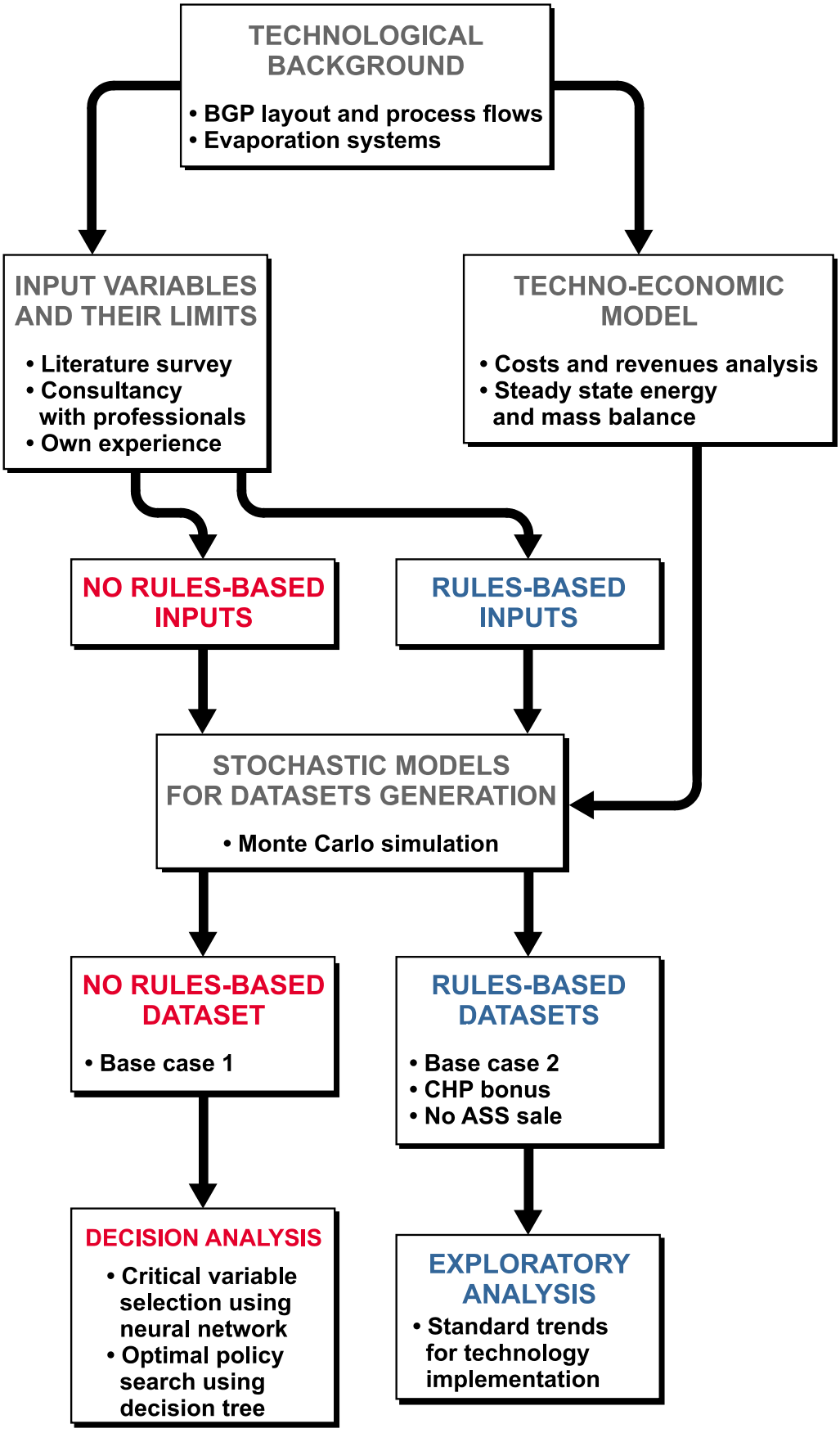


Figure 10.2: Scientific procedure of this chapter.

For further treatment of LD, which could take up to 90 % of the original digestate volume, this work assumes the integration of an evaporation system. For the purpose of this study, this work defines evaporation system (ES) as a processing unit that: (1) contains a vacuum evaporator (VE), (2) reduces LD volume and at the same time preserves its fertilizing potential, (3) can produce clean water suitable for discharge into the environment, i.e. is able to prevent ammonia from entering the outgoing freshwater stream. Considered ESs are visualized in Figure 10.3.

Each ES (A, B, C) produces two output streams during the vacuum evaporation – concentrate (thickened LD) and ammonia water. The degree of volume reduction of LD differs with the efficiency of VE and with input parameters of LD. Higher the input dry matter (DM) content, lower the potential for volume reduction. The concentrate contains most of the original nutrients. The only serious but major loss is dedicated to ammonia, which volatilizes during the evaporation step. This could be partially avoided by pH modification, using sulphuric or nitric acid. The dosage of acid depends on the concrete strategy for the ES operation and could be omitted. The operation of VE could also require the addition of a certain anti-foaming agent since LD is prone to foaming.

In the case of ES (A), presented by Maier (2018), ammonia and water are leaving the VE in gaseous form and are entering the scrubber where the ammonia is reacted with sulphuric acid, forming ammonium sulphate. Since the reaction is exothermic, the scrubber can prevent most of the water from condensing. The small portion of condensed water leaves process together with ammonium sulphate in the form of ammonium sulphate solution (ASS). This valuable by-product has the potential to be sold as a fertilizer and contribute as a possible secondary income for BGP owner. Alternatively, ASS could be used for standardization of nitrogen content in the concentrated LD. Condensation of the rest of the water vapour takes place right after the scrubber. Produced water should reach the required quality for discharge.

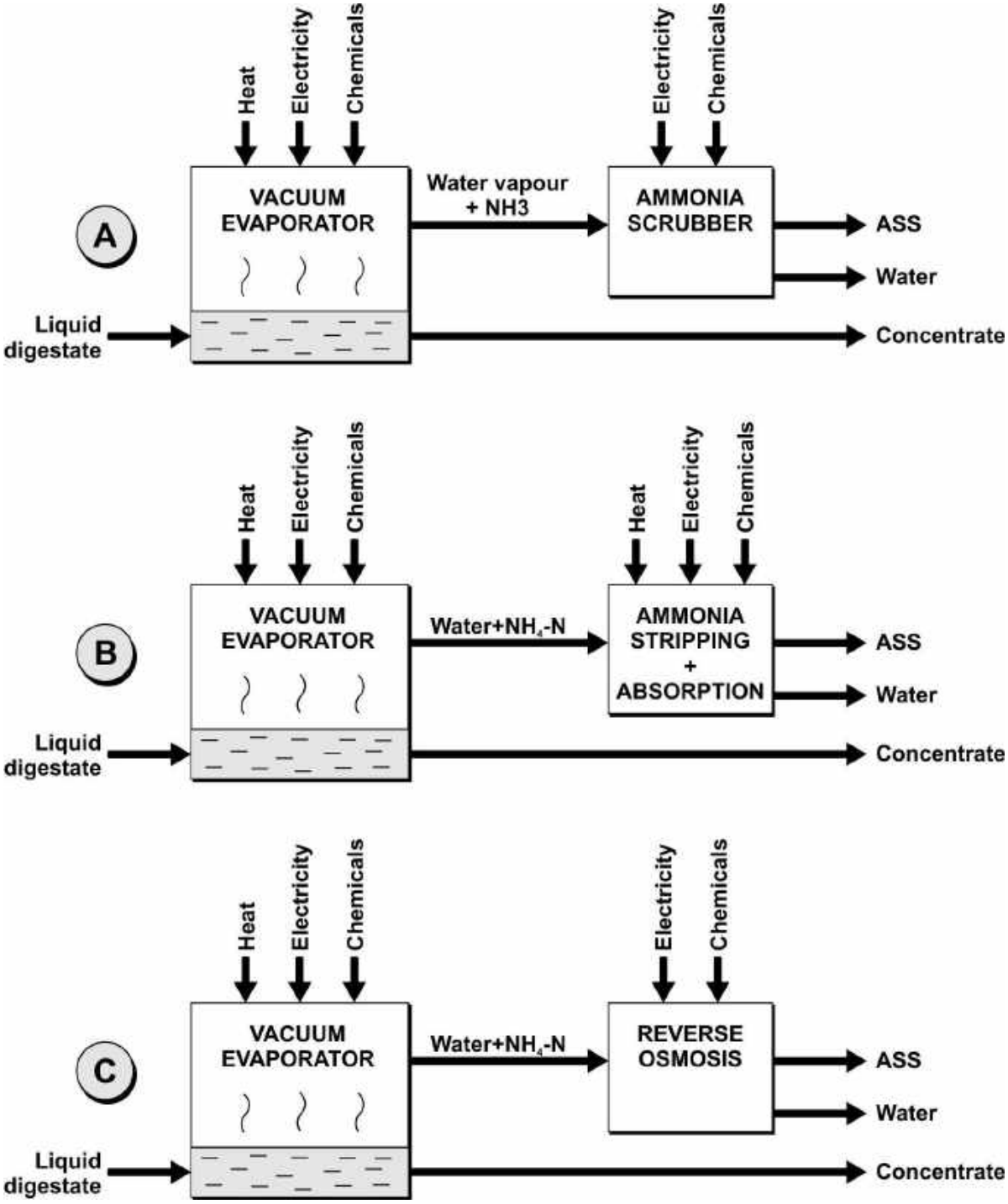


Figure 10.3: Considered evaporation systems for the liquid digestate treatment. ASS = Ammonium sulphate solution

Systems (B) and (C) need to deal with the liquid ammonia-water solution that is produced during condensation in the VE. In the ES (B), stripping with the subsequent absorption of ammonia in sulphuric acid is considered. This system was studied by (Melse and Verdoes, 2005). Increase in temperature and pH (e.g. with sodium hydroxide) of the solution are

favourable for ammonia release during the stripping process, hence the heat consumption of this system could be higher in comparison with systems (A) and (C). If there is enough waste heat available, this disadvantage does not apply.

A different approach for ammonia recovery is proposed in the system (C). The ammonia-water solution with decreased pH (using e.g. sulphuric acid) is treated with the reverse osmosis. Separated water could be eligible for discharge. The retentate (low concentrated ASS) could be subsequently thickened back in the VE, or directly stored for further use. This system was referred to by Bamelis et al. (2015).

10.2.3 Scope of Study

This work and its results are valid only for a given set of BGPs which comply with the following assumptions:

(a) Prior to entering ES, the digestate is treated by a suitable mechanical separator (e.g. screw press, decanter centrifuge), thus the LD is free of large solid particles and fibrous material.

(b) There is no extra heat available from external sources. If so, it is not considered.

(c) There are no legal restrictions on the proposed digestate processing and its subsequent utilization as a fertilizer.

(d) There is no further processing of products and associated extra revenues from these activities (e.g. drying and pelletizing of concentrated LD, advanced methods for nutrient recovery).

(e) The storage of concentrated LD, its transportation and application on a field are technically feasible with current machinery and equipment, there is no need for further investments in this regard.

(f) Costs associated with increased safety requirements for the storage of auxiliary chemicals are negligible.

(g) The technological layout and process flows of investigated BGPs are in compliance with Fig. 10.4.

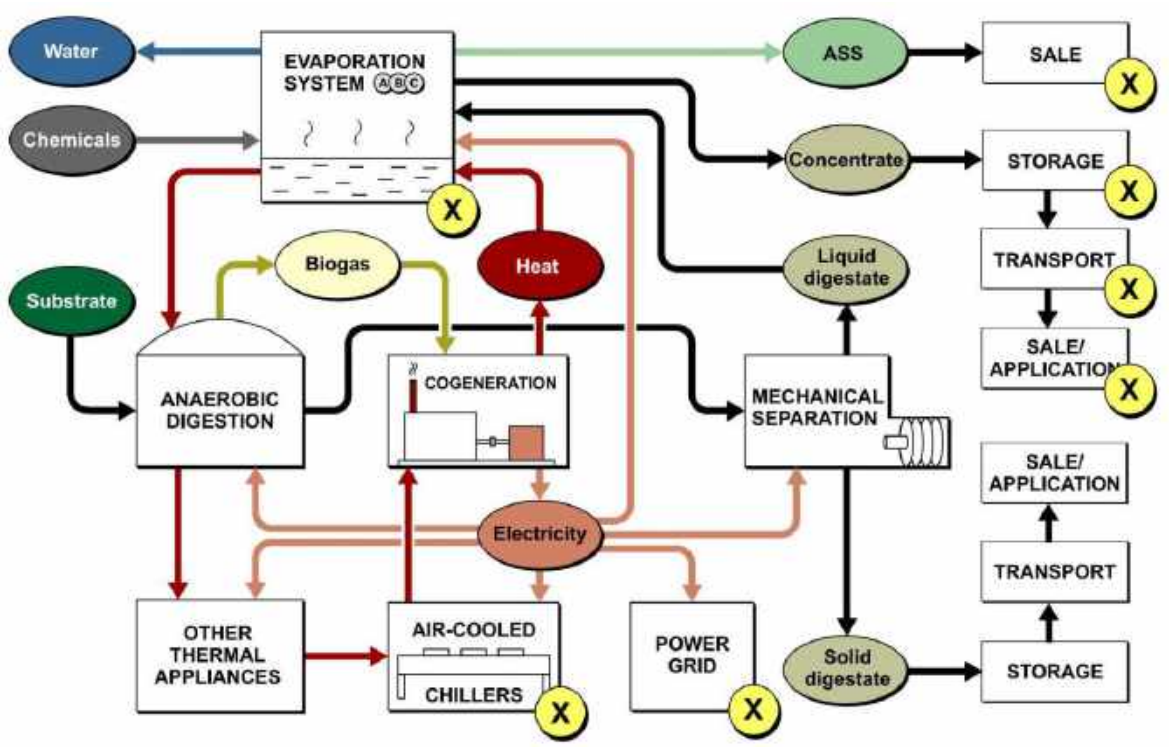


Figure 10.4: Technological layout and basic process flow diagram of the considered biogas plant. The "X" denominates process units and operations whose economy will be affected by the integration of suitable evaporation system.

10.2.4 Techno-economic Model

The techno-economic model considers only those process units and operations, whose economy will be affected by the integration of ES. These units and operations are marked in Figure 10.4 and their influence on costs and revenues is described in more detail in the following paragraphs. The techno-economic model was built on the basis of the presented mathematical relations, using MS Excel 2013 and VBA programming language. The model is steady state, analytical and is governed by energy and mass balance throughout the BGP.

10.2.4.1 Payback Period

Simple payback period (PP) was selected as the main criterion for the feasibility evaluation. The zeroth year of the payback period is defined as the exact time of digestate treatment technology being commissioned in the BGP. The payback period is dependent on investment costs and a change in the cash flow ($\Delta REV - \Delta CTS$), where the change is a direct result of the integration of an ES. Payback period was calculated in accordance with Eqn. 10.1.

$$PP = INV/(\Delta REV - \Delta CTS) \tag{10.1}$$

The overall change in revenues (Eqn. 10.2) is represented by the change in revenue from electricity sale (ΔREV_{el}^{sale}), change in revenue from the sale of concentrated LD (ΔREV_{conc}^{sale}) and by newly acquired revenue from ASS sale (REV_{ass}^{sale}). The overall change in costs (Eqn. 10.3) is represented by the change in costs for LD transport (ΔCTS_{ld}^{tran}) and by costs connected with the newly integrated ES, i.e. costs for chemicals consumption (CTS_{evap}^{chem}) and maintenance of the ES (CTS_{evap}^{mnt}).

$$\Delta REV = \Delta REV_{el}^{sale} + \Delta REV_{conc}^{sale} + REV_{ass}^{sale} \tag{10.2}$$

$$\Delta CTS = \Delta CTS_{ld}^{tran} + CTS_{evap}^{chem} + CTS_{evap}^{mnt} \tag{10.3}$$

10.2.4.2 Change in Revenues from Electricity Sales

Electric power is the most valuable output of BGPs. Since most feed-in tariffs (also including premium tariffs which are driven by market prices) are subsidized, BGPs are trying to maximize their electric output, thus minimizing their internal consumption. The change in revenues from the electricity sale comprises of two parts (Eqn. 10.4) – change in revenue from the net electricity production and optional extra income from the CHP generation.

$$\Delta REV_{el}^{sale} = sPE_{fit} \cdot SH \cdot \left(\Delta PC_{ac}^{el} + \Delta PC_{agit}^{el} - PC_{evap}^{el}\right) + sPE_{chp} \cdot SH \cdot PO_{cog}^{el}/1000 \tag{10.4}$$

Table 10.2: Considered variables, condition, and limits.

Input variable	**Notation**	**Units**	**Min**	**Max**	**Based on**
Electricity feed-in tariff	sPE_{fit}	EUR kWh^{-1}	0.06	0.25	Pablo-Romero et al. (2017)
Specific costs for LD application	$sCTS_{ld}^{apl,va}$	EUR m^{-3}	1	5	Auburger et al. (2015), Drosg et al. (2015), Author's databank*
Maintenance coefficient	k_{evap}^{mnt}	-	0.05	0.15	Author's databank*
Specific price of chemicals	sP_{chem}	EUR t^{-1}	0	4	Author's databank*
Specific costs for LD transport	$sCTS_{ld}^{tran,v}$	EUR m^{-3}	2	12	Auburger et al. (2015), Drosg et al. (2015), Author's databank*
Installed electrical power output	PO_{cog}^{el}	MW	0.05	3	Scarlat et al. (2018)
Specific annual production of digestate	$sPRA_{dig}$	t y^{-1} MW^{-1}	10000	30000	CZBA (2014), Author's databank*
DM concentration in digestate	DM_{dig}	wt%	4	12	Monlau et al. (2015), Drosg et al. (2015), Author's databank*
DM concentration in LD	DM_{ld}	wt%	2	8	Møller et al. (2002), Drosg et al. (2015), Author's databank*
DM concentration in concentrate	DM_{conc}	wt%	8	15	Guercini et al. (2014), Drosg et al (2015), Author's databank*
DM concentration in solid fraction	DM_{sf}	wt%	20	35	Møller et al. (2000), Drosg et al (2015)
Specific thermal power consumption of evaporation system	$sPCM_{evap}^{th}$	kWh t^{-1}	150	500	Cordes (2018), Vondra et al. (2018a), Author's databank*

Table 10.2 (Cont.): Considered variables, condition, and limits.

Input variable	Notation	Units	Min	Max	Based on
Specific electric power consumption of evaporation system	$sPCE_{evap}^{el}$	kWh t^{-1}	8	25	Cordes (2018), Chiumenti et al. (2013), Author's databank*
Coefficient of heat consumption	k_{bgp}^{th}	-	0.2	1	Daniel-Gromke et al. (2017), Author's databank
Specific power consumption of air-cooled chillers	$sPCE_{ac}^{el}$	kW MW^{-1}	5	15	Noel and Fourcroy (2017), Author's databank*
Change in specific power consumption of agitators	$\Delta sPCE_{agit}^{el}$	kW MW^{-1}	-3	6	Author's databank*
CHP bonus	sPE_{chp}	EUR kWh^{-1}	0	0.05[a)]	European Commission, (2019b)
Specific production of ASS	$s\dot{M}_{ass}$	wt%	0.025	0.035	Maier (2018), Cordes (2018)
Specific market price for ASS	sP_{ass}^{sale}	EUR t^{-1}	10[b)]	30[b)]	Maier (2018)
Specific investment costs of a project	$sINV$	EUR y t^{-1}	10	40	Author's databank*
Thermal efficiency of cogeneration unit	η_{cog}^{th}	%	42[c)]		Author's databank*
Electrical efficiency of cogeneration unit	η_{cog}^{el}	%	42[c)]		Author's databank*
Service hours	SH	h y^{-1}	8760[c)]		Author's databank*

Notes: *Author's databank: Own experience, consultancy with BGP operators and professionals from 2013 to 2019; (a) Only in the "CHP bonus" scenario, otherwise without CHP bonus; (b) Not valid in the "No ASS sale" scenario; (c) Constant variables

The net electricity production will be affected by the power consumption of air-cooled chillers, agitators in storage tanks and the ES itself. The most important parameter regarding the electricity sale is the feed-in tariff (sPE_{fit}) for electricity production from biogas, which could differ based on a particular policy or government strategy. Based on the work of Pablo-Romero et al. (2017) this work placed the value of sPE_{fit} in the range from 0.06 to 0.25 EUR kWh^{-1} (Table 10.2).

10.2.4.3 Changes in Power Consumption of Air-Cooled Chillers

The difference between the cooling performance of air-cooled chillers in the original BGP ($PC^{th}_{ac,orig}$) and after the ES integration ($PC^{th}_{ac,new}$) is equal to the thermal power consumption of the ES (PC^{th}_{evap}) (Eqn. 10.5). This is because the evaporation technology basically acts as a sink for the waste heat. PC^{th}_{evap} is calculated (Eqn. 10.5) from the specific thermal consumption of the ES ($sPCM^{th}_{evap}$) and LD flow rate ($\dot{M}_{ld}$). $sPCM^{th}_{evap}$ was estimated from Vondra et al. (2018a) to range between 150 and 500 kWh/t. The change in electricity consumption of air-cooled chillers (ΔPC^{el}_{ac}) (Eqn. 10.6) is also influenced by their specific electricity consumption ($sPCE^{el}_{ac}$), which could vary approx. between 5 to 15 W/kWh (Noel and Fourcroy, 2017).

$$PC^{th}_{evap} = PC^{th}_{ac,orig} - PC^{th}_{ac,new} = sPCM^{th}_{evap} \cdot \dot{M}_{ld} \quad (10.5)$$

$$\Delta PC^{el}_{ac} = PC^{el}_{ac,orig} - PC^{el}_{ac,new} = sPCE^{el}_{ac} \cdot \left(PC^{th}_{ac,orig} - PC^{th}_{ac,new}\right) \quad (10.6)$$

It must be added that in most cases, the evaporator itself requires a certain amount of cooling capacity, which could be satisfied in different ways (cooling towers, AC chillers, heat recovery etc.). For the purposes of this study, it is assumed that power and investment requirements for the cooling capacity are an implicit part of the overall technical specification and price of the ES.

10.2.4.4 Changes in Power Consumption of Agitators

Stored LD requires regular mixing to prevent solids from settling. If the digestate is currently stored in 2 or more tanks, it can be expected that after the thickening, fewer tanks will be used for storage and the total consumption of agitators will be reduced. If the digestate is stored in one tank only, the remainder after the thickening is stored again in this one tank, but consumption of agitators will increase as they will mix a thicker and more viscous suspension. Thus this work expects that the change in specific power consumption of agitators ($\Delta sPCE^{el}_{agit}$)

could be in a range from (-3,0) to 6 W per kW of electricity installed, where the negative value represents option with a single storage tank. The change in the power consumption of agitators (ΔPC_{agit}^{el}) is calculated according to Eqn. 10.7.

$$\Delta PC_{agit}^{el} = \Delta sPCE_{agit}^{el} \cdot PO_{cog}^{el} \quad (10.7)$$

10.2.4.5 Power Consumption of Evaporation Systems

The overall electricity balance will be negatively influenced by the consumption of ES (PC_{evap}^{el}), which is determined using Eqn. 10.8. Here the specific electricity consumption ($sPCE_{evap}^{el}$) and LD mass flow rate ($\dot{M}_{ld}$) are decisive factors. Based on works by Vondra et al. (2018a) and Chiumenti et al. (2013), this work expects $sPCE_{evap}^{el}$ to vary between 8 and 25 kWh per tonne of LD treated.

$$PC_{evap}^{el} = sPCE_{evap}^{el} \cdot \dot{M}_{ld} \quad (10.8)$$

10.2.4.6 Mass Balance

The production of digestate ($\dot{M}_{dig}$) could be estimated from installed power output of a BGP (PO_{cog}^{el}), number of service hours (SH) and specific annual digestate production ($sPRA_{dig}$), which is typically in the range from 10,000 m^3 to 30,000 m^3 per year and MW of electricity installed according to Eqn. 10.9.

$$\dot{M}_{dig} = (PO_{cog}^{el} \cdot sPRA_{dig})/SH \quad (10.9)$$

Mass flows during the mechanical separation are governed by Eqn. 10.10 and 10.11. The production of LD could differ based on the separator type and its setup. It is supposed that the dry DM content in digestate (DM_{dig}) from wet anaerobic digestion could vary between 4 and 12 %, which is within boundaries set by Monlau et al. (2015) and Drosg et al. (2015).

$$\dot{M}_{dig} = \dot{M}_{ld} \cdot \dot{M}_{sf} \quad (10.10)$$

$$\dot{M}_{dig} \cdot DM_{dig} = \dot{M}_{ld} \cdot DM_{ld} + \dot{M}_{sf} \cdot DM_{sf} \quad (10.11)$$

Production of the concentrated LD ($\dot{M}_{conc}$) in the ES is governed by Eqn. 10.12 under the assumption that all the DM stays in the concentrate. The final DM concentration (DM_{conc}) is an input provided by the technology manufacturer. This work supposes that DM_{conc} is between 8 and 15 % as more concentrated DM would require demanding types of evaporators with

stirrers. Production of fresh water ($\dot{M}_{fw}$) could be calculated using Eqn. 10.13. Mass flow rates of applied chemicals are neglected.

$$\dot{M}_{ld} \cdot DM_{ld} = \dot{M}_{conc} \cdot DM_{conc} \quad (10.12)$$

$$\dot{M}_{fw} = \dot{M}_{ld} - \dot{M}_{conc} - \dot{M}_{ass} \quad (10.13)$$

The density of digestate and its fractions were estimated using Eqn. 10.14, where the result is solely dependent on the DM content (DM) of a particular stream:

$$\rho_{dig} = 4.81 \cdot DM + 1026.2 \quad (10.14)$$

Hence the volumetric flow rate of digestate ($\dot{V}_{dig}$) and any of its fractions ($\dot{V}_{ld}, \dot{V}_{conc}$) can be calculated using Eqn. 10.15:

$$\dot{V}_{dig} = \dot{M}_{dig}/\rho_{dig} \quad (10.15)$$

10.2.4.7 Change in Revenues from the Sale of Concentrated LD

Thickening of the LD increases its price as the final product contains almost the same amount of nutrients but in a much smaller volume. It is assumed that the increase in the price of the LD is directly proportional to the change in its volume ($\dot{V}_{ld} - \dot{V}_{conc}$) as the change of volume is directly proportional to the saving of application (spreading) costs. This could be beneficial both for the external end user of the fertilizer or for the plant operator himself, if he is the one who utilizes the LD for the fertilization of soil. Change in revenues (ΔREV_{conc}^{sale}) is determined using Eqn. 10.16, where $sCTS_{ld}^{apl,var}$ stays for specific variable costs for the application of LD and are expected to range from 1 to 5 EUR/t.

$$\Delta REV_{conc}^{sale} = sCTS_{ld}^{apl,var} \cdot SH \cdot \left(\dot{V}_{ld} - \dot{V}_{conc}\right) \quad (10.16)$$

10.2.4.8 Revenues from the Ammonium Sulphate Solution (ASS)

The production of ASS will vary with the type of process unit (scrubber, stripper, reverse osmosis) responsible for the reaction between sulfuric acid and ammonia (NH_3) or ammonium (NH_4^+). Although the operation itself could be different, the overall mass flow rate of ASS produced will change only slightly, since the general requirement for ammonia removal is the same and is mainly driven by the terminal concentration of ammonia in the separated fresh

water. The specific mass flow rate of ASS ($s\dot{M}_{ass}$) will be provided by technology supplier and is estimated between 25 and 35 kg/t of LD. The revenue from ASS sale (REV_{ass}^{sale}) could be calculated using Eqn. 10.17. The market price for ASS (sP_{ass}^{sale}) may vary throughout the EU and will be influenced by demand and purchasing power of local companies.

$$REV_{ass}^{sale} = sP_{ass}^{sale} \cdot s\dot{M}_{ass} \cdot \dot{M}_{ld} \cdot SH/100 \quad (10.17)$$

10.2.4.9 Change in Costs for LD Transportation

The volume reduction of LD in comparison with the original situation will affect transportation costs. Overall transport costs reduction (ΔCTS_{conc}^{tran}) (see Eqn. 10.18) will be proportional to the difference between the volumetric flow rate of LD after the thickening process ($\dot{V}_{conc}$) and the original one ($\dot{V}_{ld}$). Total savings related to LD transport are influenced by specific variable transportation costs ($sCTS_{ld}^{tran,var}$), which are strongly related to average transport distance. Since the transport distance can exceed 100 km in certain areas (North-West Germany), it is expected that high fluctuation are present in $sCTS_{ld}^{tran,var}$ and therefore also in ΔCTS_{conc}^{tran}.

$$\Delta CTS_{conc}^{tran} = sCTS_{ld}^{tran,var} \cdot SH \cdot \left(\dot{V}_{conc} - \dot{V}_{ld}\right) \quad (10.18)$$

10.2.4.10 Costs for Chemicals Consumption

To reach a required quality of products using any type of proposed ES, utilization of certain chemicals is a must. Sulphuric acid is the cheapest solution (nitric acid as a more expensive option) for the ammonia recovery in ESs. Intensive foaming is typical for the evaporation of the LD, hence an antifoaming agent could be another chemical input consumed in the process. Specially modified evaporators are capable of zero AFA consumption, but higher investment or operational costs (e.g. for mechanical foam disruption) must be considered.

The overall consumption of chemicals could vary significantly based on the particular system configuration. To evaluate overall costs for chemicals consumption (CTS_{evap}^{chem}) according to Eqn. 10.19, a technology supplier must provide information about the expected specific price of chemicals (sP_{chem}) per tonne of treated LD. This study assumes sP_{chem} to reach up to 4 EUR/t.

$$CTS_{evap}^{chem} = sP_{chem} \cdot \dot{M}_{ld} \cdot SH \quad (10.19)$$

10.2.4.11 Costs for Technology Maintenance

As any other process unit, even an ES will require maintenance throughout its lifetime. It will mainly require regular washout of its components, especially heat transfer surfaces and pumps, which could suffer from clogging and fouling. Occasional professional service or replacement of damaged components should also be expected. LD is an organic solution with high viscosity and significant content of DM and it is quite difficult to handle. Annual costs for technology maintenance (CTS_{evap}^{mnt}) could be estimated as a percentage from total investment costs (INV) (and in the Eqn. 10.20) are expressed by the maintenance coefficient (k_{evap}^{mnt}), which is estimated to range between 0.05 to 0.15.

$$CTS_{evap}^{mnt} = k_{evap}^{mnt} \cdot INV \tag{10.20}$$

10.2.4.12 Investment Costs

For purposes of this study, total investment costs (INV of a project are calculated from specific investment costs ($sINV$) per tonne of processed LD. This approach allows comparison of BGPs with different capacities, which would vary significantly in terms of total investment costs. Eqn. 10.21 describes the calculation:

$$INV = sINV \cdot \dot{M}_{ld} \tag{10.21}$$

10.2.4.13 Thermal Energy Consumption

For the thickening project to be successful, it is necessary to utilize only as much heat, as it is available in the BGP in a form of cost-free waste heat. Thermal energy is the main source of energy for the evaporation process. Any requirement for extra thermal energy production would significantly reduce the feasibility of the project, most probably it would make it unacceptable. In this study, the feasibility of the project considers the scenarios in which the evaporation system does not consume more than available amount of heat (this is considered as infeasible). This condition is described by Eqn. 10.22, where PC_{bgp}^{th} refers to own heat consumption by the BGP. PC_{bgp}^{th} is determined in accordance with Eqn. 10.23 and is dependent on the coefficient of heat consumption (k_{bgp}^{th}). In BGPs the heat is utilized for the digestion process, eventually for drying processes or heating of social buildings, stables, attached industrial or commercial premises etc (Daniel-Gromke et al., 2017). In certain cases, k_{bgp}^{th}

could reach up to 1, i.e. there is no available waste heat and evaporation project is rated as infeasible.

$$PC_{evap}^{th} \leq PO_{cog}^{th} - PC_{bgp}^{th} \quad (10.22)$$

$$PC_{bgp}^{th} = k_{bgp}^{th} \cdot PO_{cog}^{th} \quad (10.23)$$

Thermal power output of the cogeneration unit was determined using Eqn. 10.24, where η_{cog}^{th} and η_{cog}^{el} are constants, both equal to 42 %.

$$PO_{cog}^{th} = PO_{cog}^{el} / \eta_{cog}^{el} \cdot \eta_{cog}^{th} \quad (10.24)$$

10.2.5 Datasets Generation and Analysis

As introduced earlier, there are two main outputs. Firstly, a support for decision making in a form of a decision tree, which was generated using a "no rules-based" dataset (Chapter 10.2.5.1). Secondly, exploratory data analysis was carried out to provide comprehensive information about feasibility of the system integration. "Rules-based" datasets were used (Chapter 10.2.5.2) for that purpose.

To the authors' knowledge, there is no sufficiently extensive database of parameters of existing BGPs which could be used for the analyses. So, to reproduce variability of real operational parameters, a stochastic approach was deployed to simulate datasets using the Monte-Carlo (MC) method together with the techno-economic model presented in Chapter 10.2.4. Each dataset is obtained by 20,000 simulation runs with randomly generated values of the independent variables. Each combination of variables represents the operational parameters of one particular BGP. This work call these combinations as "samples". The aim of the subsequent data analysis was to identify successful combinations of samples that are favourable for the economy of the project. Feasibility of the samples is tested using three criteria: consumed heat is lesser or equal to available heat, cash flow is positive and payback period is below certain level (3 or 5 years for "no rules-based", 15 years for "rules-based datasets"). The samples, which do not meet these requirements, are not feasible. But they are not excluded from the results. They just represent the set of solutions which are not acceptable for BGP.

10.2.5.1 "No rules-based" Dataset

For the decision analysis, it is necessary to have all combinations of values of the independent variables since PP is investigated as a function. The full decision space for the implementation of ES is generated by MC simulations of the rigorous techno-economic model. In the MC

simulation, each point in the domain space is assigned an equal probability to give a non-bias decision space. This work categorized this generated data set under "no rule-based" dataset and assigned the name "Base case 1" to this scenario. The parameters of the uniform distribution, minimum and maximum value, were estimated according to data found in literature or according to the authors' experience or consultancies with technology suppliers and BGP operators (specified in the Table 10.2).

The tools used for the decision analysis are the Statistica package and development in Python 3.7. This work extended development based-on conventional development libraries such as sklearn, matplotlib, scikit-optimize, dtreeviz. The decision analysis aims to select important variables from a well-trained artificial neural network, then using them to train a decision tree classifier.

The decision tree classifier functions to separate sub-solution spaces of fast payback period (0-3 years) and medium payback period (3-5 years) from the full decision space. This approach is expected to give solutions that are more reliable and robust, as the optimality is not presented as a "point", but as a solution space in the tree's leaf node. The tree is pruned and validated with an alternative set of testing data. Bayesian optimization is also carried out to efficiently tune the hyper-parameters of the decision tree, ensuring no overfitting and high validation performance (see Figure 10.5).

Using 20,000 data points from Monte Carlo simulations, important significant variables were selected by using a simple artificial neural network with the automated search algorithm. Validation for the accuracy of the artificial neural network was performed on a separately prepared 20,000 data points. The artificial neural network architecture was searched between 10 to 25 hidden layers and variable activation functions including Identity, Logistics, Hyperbolic Tangent and Exponential functions. The loss function that was used for training the artificial neural network is the Mean Squared Error (MSE) because it is a globally differentiable loss function that speeds up training time. Sensitivity analysis was carried out for each variable by turning off the respective variables in the neural network and measure the error in payback period prediction. If the shut-off variable causes a high error in prediction, it is higher in importance. The importance factor is shown in Eqn. 10.25:

$$I(x_k) = \sum |y - y'(x_n)| \qquad where\ \forall n \in N, \qquad \mathrm{k} \notin N \tag{10.25}$$

The decision tree classifier was trained with a best split greedy approach by minimizing entropy at each resulting node. The entropy of the node is correlated with the probability of classes within the classification group (see Eqn. 10.26). This ensures that the information gain is maximized when moving down the tree. This work also trained the decision tree with 20,000 Monte Carlo simulation data, while using a separate 20,000 Monte Carlo simulation points for validation.

$$Entropy = -\sum_{j} p_j \ log_2 p_j \tag{10.26}$$

A simple pruning algorithm was used to eliminate repeated classifications of the same class. Nevertheless, the pseudo-code for the algorithm is shown in Table 10.3 for easier understanding.

Table 10.3: Algorithm for decision tree pruning

Algorithm 1 Pseudo-code for decision tree pruning
1: **Repeat**
2: **For each** leaf nodes in decision tree **do** • *Get the leaf nodes of tree*
3: Get parent of leaf node • *Check the decision of the parent of tree node*
4: **If** classification of parent == classification of leaf **node** • *Check if there are trivial leaf nodes*
5: Remove all child of parent **with** classification == parent • *Remove trivial leaf nodes*
6: Update leaf nodes • *Update the new leaf nodes after previous are removed*
7: **Until** all classification of leaf nodes != classification of parent • *Repeat until no more trivial leaf nodes*

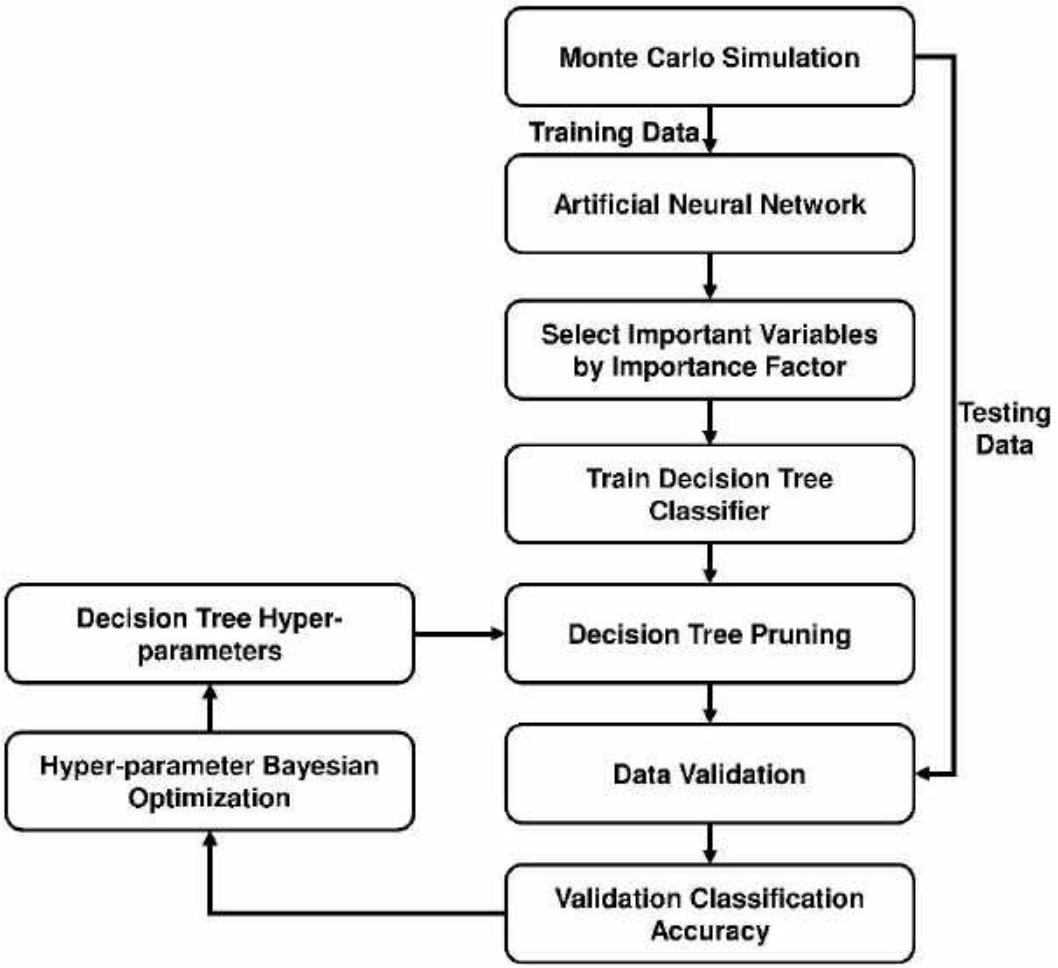


Figure 10.5: Procedure for full decision space analysis

The effects of this pruning algorithm allow the tree to have a reduced number of nodes while maintaining the same level of accuracy within the decision tree. This can be visualized in Figure 10.6, where two leaf nodes of the same classification were effectively pruned to form a tree with reduced branches but same classification accuracy.

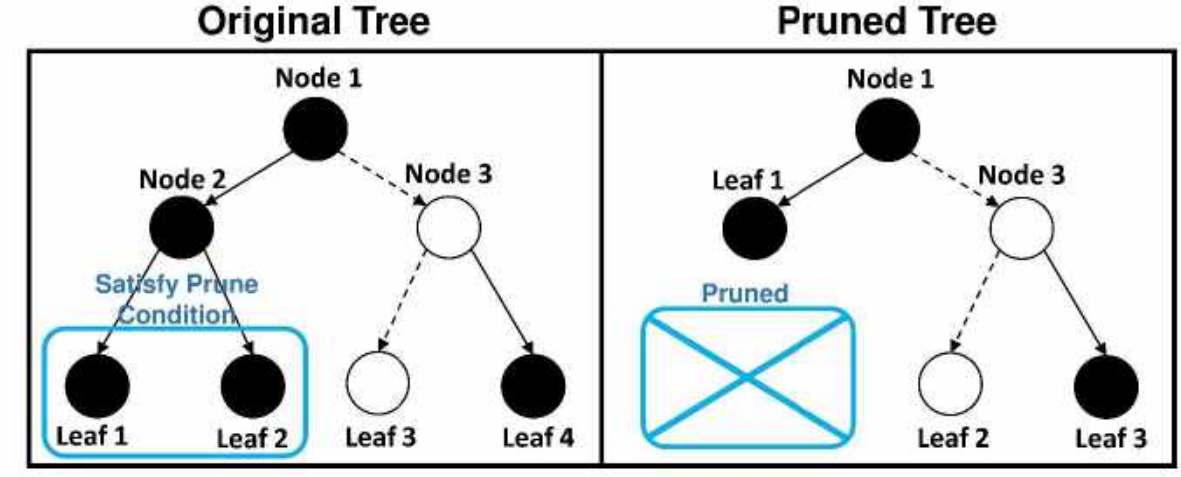


Figure 10.6: Visual decision tree pruning illustration

The hyper-parameters for training the decision tree classifier are the maximum depth of the tree (Max Depth), minimum samples to activate further splitting (Min Sample Split), the minimum weight fraction of leaf node samples with respect to input samples (Min Weight Fraction) and minimum impurity decrease. This work has carefully selected logical limits of the hyper-parameter optimization problem by trial and error to provide faster convergence of the optimization algorithm. The limits for hyper-parameter optimization can be found in Table 10.4:

Table 10.4: Limits for hyper-parameters in the decision tree

Hyper-parameters	Minimum Limit	Maximum Limit
Max Depth	3	50
Min Sample Split	10	40
Min Weight Fraction	0	0.001
Min Impurity Decrease	0	0.001

Gaussian process Bayesian optimization was used for the hyper-parameter optimization. The kernel that was used is the Matern kernel with v =5/2 and small α_0 (1e-6) was according to Eqn. 10.27:

$$C_v(d) = \sum_0 (x, x') = \alpha_0 \frac{2^{1-v}}{\Gamma(v)} \left(\sqrt{2v}\, \|x - x'\|\right)^v K_v \left(\sqrt{2v}\, \|x - x'\|\right) \quad (10.27)$$

Where K_v is the modified Bessel function and Γ is the gamma function. The kernel serves as the prior of the function of tree accuracy with respect to hyper-parameters. The posterior distribution of the function is then estimated using all available data, and the maximal point is estimated and measured. This process repeats iteratively until convergence.

10.2.5.2 "Rules-based" Datasets

The "rules-based" datasets consists of three scenarios with different sets of generated data. The "Base case 2" acts as a reference scenario for the "rules-based" datasets. The "No ASS sale" scenario examines the impact of the marketability of an ammonium sulphate solution. The "CHP bonus" scenario analysis the effect of bonuses for CHP production (Chapter 10.2.4.2).

In contrast with decision analysis (Chapter 10.2.5.1), the MC simulation mode used for generation of "rules-based" datasets require data which reflects the real operational conditions. For this case, simply including unrealistic combinations would lead to biased results. That is why some rules towards realistic combinations were introduced.

The objective of these rules is to modify the probability distribution or the functional dependency of the variables so that it gives a better representation of the real situation. However, there is still some fluctuation along this dependency determined by the uniform distribution. The rules are described in more detail below and are visualized in Figure 10.7. Variables which are strictly independent are generated from the same uniform distribution used in the case of the decision analysis.

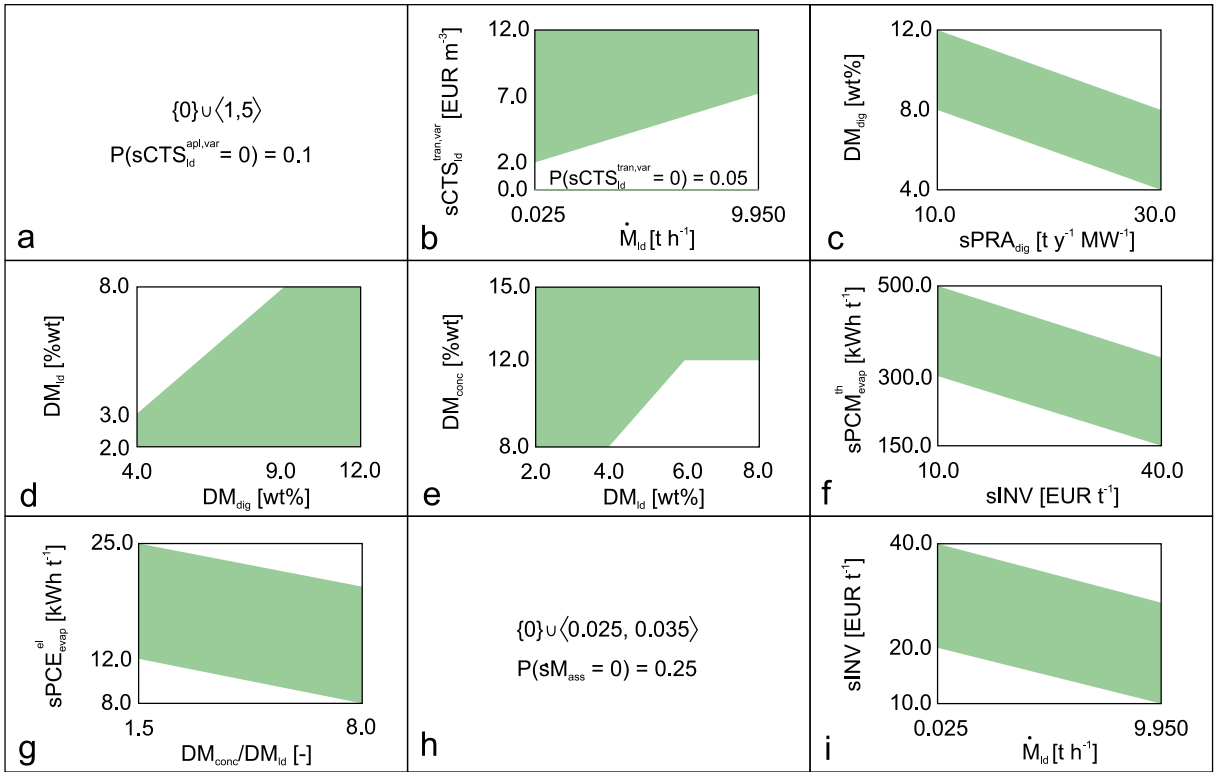


Figure 10.7: Matrix for Monte Carlo simulation limits of the "rules-based" scenarios, h) is not valid for the "No ASS sale" scenario

The first of the rules considers the fact that in some cases, the application costs may be zero. In these cases, known to us from the Czech Republic, the fertilizer is collected by local farms free of charge. This work estimates that the chance for zero application costs is 10 % (Figure 10.7(a)). Otherwise, this cost is between 1 and 5 EUR/m^3.

Certain BGPs do not have any costs related to LD transport outside the plant itself. This situation is typical when there is some kind of barter exchange between closely located farmers (who provide biomass substrate for free or with discount) and BGP, which processes the biomass and provides a valuable natural fertilizer. Such a BGP may not have any agricultural production at all, hence the cost-free collection of LD by local farmers is highly appreciated. This work supposes that 5 % of BGPs fulfil this scenario ($sCTS_{ld}^{apl,var}$= 0 EUR/m^3). If the transport cost is non-zero another rule is applied. Since BGPs with high installed capacity (high LD production) need a larger area of land for discharge LD, it is expected that also their average transport distance will be higher. Hence this work sets a rule for all "rules-based" scenarios, which generates higher $sCTS_{ld}^{tran,var}$ for high capacity BGPs (Figure 10.7(b)). For instance, if $\dot{M}_{ld}$ = 0.025 t/h the value of $sCTS_{ld}^{tran,var}$ can be generated from the range between 2 and 12 EUR/m^3. If $\dot{M}_{ld}$ = 9.95 t/h the value of $sCTS_{ld}^{tran,var}$ can be generated from the range between 7 and 12 EUR/m^3.

A rule of high DM_{dig} – low $sPRA_{dig}$ and low DM_{dig} – high $sPRA_{dig}$ was applied (Figure 10.7(c)) since high water content results in high volumes of digestate (Drosg et al., 2015). DM in LD (DM_{ld}) is estimated to be between 2 and 8 % with respect to Møller et al. (2002), Drosg et al. (2015) and own experience and is restricted to be always less than DM_{dig} according to Figure 10.7(d). ES should provide reasonably higher DM content in a concentrate compared to LD. To guarantee reasonable difference between DM_{conc} and DM_{ld} the rule depicted in Figure 10.7(e) is introduced so the difference lesser than 4 % is impossible. With the increase in thermal efficiency of the ES ($sPCM_{evap}^{th}$), this work expects high specific investment costs ($sINV$) as a result of more intensive heat recovery and increased requirements for heat transfer area. This rule for "rule-based" scenarios is applied according to Figure 10.7(f).

Since the $sPCE_{evap}^{el}$ is often declining with the increase of volume reduction (Vondra et al., 2018a), i.e. with high freshwater production, a restriction is acquired as a part of the "rules-based" scenarios (Figure 10.7(g)). For the "No ASS sale" scenario this work assumes that no ASS is sold. In this way, this work demonstrates the influence of N recovery and marketability on the project feasibility. For the remaining scenarios, this work still assumes 25% probability that no ASS is sold (i.e. $s\dot{M}_{ass}$ = 0). However, if ASS is successfully sold its value is between 0.025 and 0.035 EUR/t (Figure 10.7(h)). Since high capacity projects are generally favourable

for specific investment costs, this work assumes for the "rules-based" scenarios that $sINV$ will decrease with the mass flow rate of LD (Figure 10.7(i)).

The CHP bonus can be used to support BGP operation and motivate operators to usefully utilize waste heat at the same time. It is up to authorities what is considered as useful utilization. In this study, it is assumed that ES represents useful utilization. There may be many strategies for the CHP bonus application. For instance, only the amount of the net electricity production corresponding to the usefully utilized heat is subsidized. Or, if more than 50 % of the heat is usefully utilized the whole net electricity production is subsidized, etc. It is not possible to consider all possible strategies in the model.

Therefore, in order to simplify it, the whole net electricity production is subsidized in the "CHP bonus" scenario. However, the value of the CHP bonus is influenced by the feed-in-tariff and the heat utilization prior to ES integration. It is assumed that the higher the feed-in-tariff the lower the CHP bonus because high support by a government is already provided by high feed-in-tariff. Further, it is assumed that the higher heat utilization the lower the CHP bonus because those operators who currently utilize high amount of heat already have (probably) CHP bonus virtually included in the feed-in-tariff. The model for CHP bonus is depicted in Figure 10.8. It is assumed that BGP with the current heat consumption higher than 80 % of the total production already has the CHP bonus included in the feed-in-tariff even when its value is only 0.06 EUR/kWh. On the other hand, if BGP has the heat utilization lower than 40 % of the total production and the feed-in-tariff lesser than 0.2025 it gets the CHP bonus and its value depends on the precise value of the feed-in-tariff. The highest CHP bonus of 0.05 EUR/kWh was inspired by the support scheme in Finland (European Commission, 2019b). In this study, the CHP bonus could be claimed only in the "CHP bonus" scenario.

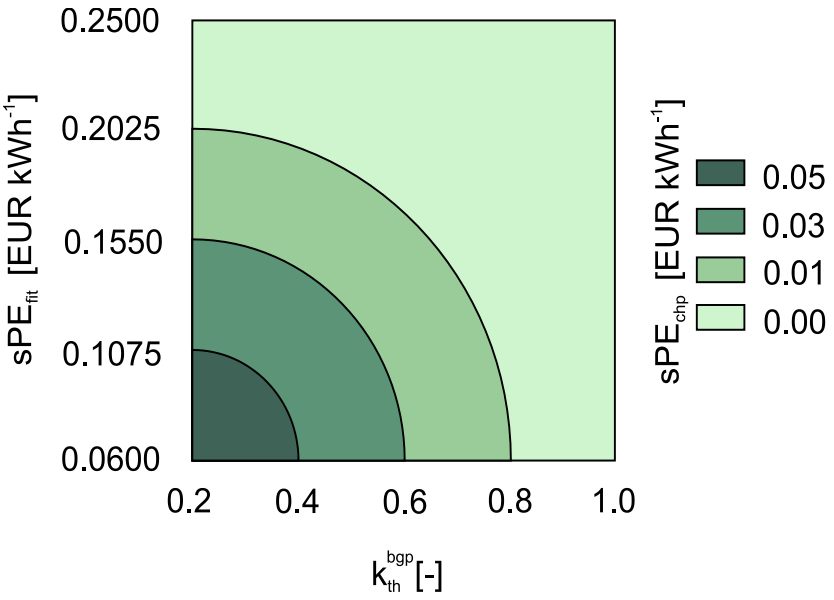
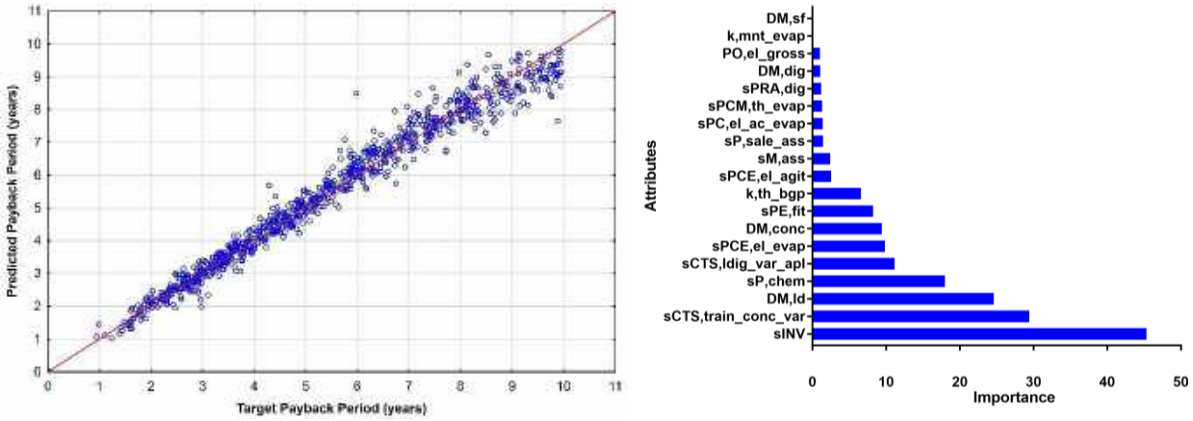


Figure 10.8: Monte Carlo simulation limit for the CHP bonus variable

10.3 Results and Discussion

10.3.1 Base case 1 (“no rules-based" dataset)

Using the dataset from the Monte Carlo simulation, a neural network was trained to predict the payback period of the implementation of evaporation systems given the conditions and criteria. The automated neural network architectural search found a configuration of MLP 17-10-1 (exp-exp), where variables with insignificant variance were automatically ignored. The simple artificial neural network validated with 0.996 Mean Squared Error accuracy (see Figure 10.9(a)) while showing the importance value of each variable by sensitivity analysis (see Figure 10.9(b)). There are two “knee points” within the importance plots, one being after the 4th variable (from bottom) and another after the 9th variable. The drastic change of gradients in these “knee points” signifies that those are the threshold for variable screening. This work chooses the first 4 variables instead of 9 variables considering the complexity and user experience for the decision maker. The four most significant variables are the specific investment cost per LD ($sINV$), the specific cost for concentrated LD transport ($sCTS_{ld}^{tran,var}$), DM concentration in LD (DM_{ld}), and the specific price of chemicals (sP_{chem}). These variables are used to train the decision tree classifier.



(a) (b)

Figure 10.9: (a) Neural network expected validation result against actual result (b) Importance of variables in scree plot

Hyper-parameter optimization was implemented for training the decision tree classifier to achieve high validation accuracy. Using Bayesian optimization, the maximum validation error was found in 5 iterations. The algorithm was further iterated for 25 steps and the validation error remains unchanged (Figure 10.10). This shows that the tree generated has at least an epsilon-optimal hyper-parameter setting to achieve high validation performance. The tree was able to achieve 98.14 % of classification accuracy in a stochastic environment.

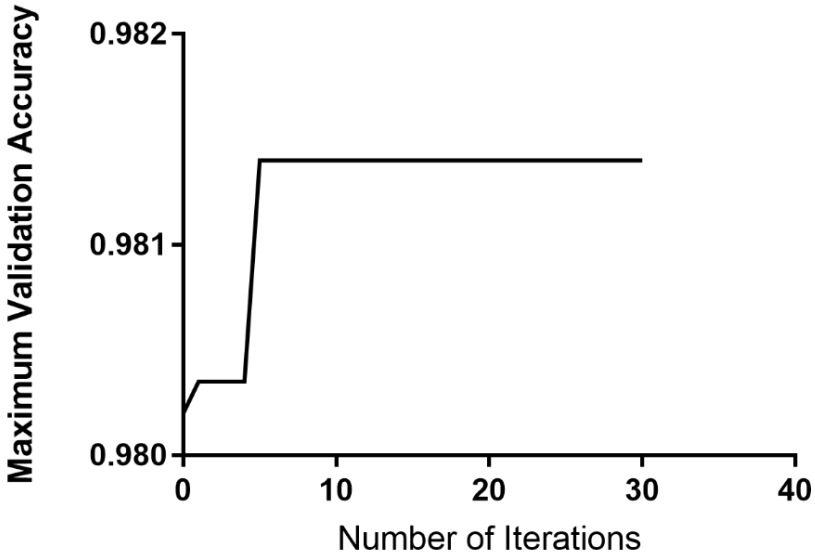


Figure 10.10: Maximum validation classification accuracy and number of iterations by Bayesian optimization

The full decision tree diagram can be found in the Appendix A. Statistically, there is more than 95% chance of getting infeasible economics by evaluating the full decision space and their outcomes when implementing the ES. This reflects that the economics for treating digestate in the European Union is particularly difficult and generally unfavourable as of now. Successful biogas projects require significant effort from both policy and management (Gebrezgabher et al., 2010). Searching for a good heuristic for stakeholders to implement such evaporation system is absurdly difficult and can be described as "searching for a needle in a haystack". The main management effort of such project is to market the recovered ASS, while CHP incentives are provided by governments (European Commission, 2018) to encourage such efforts. The effect of these important but inadequate efforts will be comprehensively analysed in Chapter 10.3.2. Nevertheless, by screening through variables with the decision tree, this study can essentially improve this situation to benefit the decision maker. This is computed by the decision tree classifier to separate the full decision space into subspace that are representative of fast PP (0-3 years), average PP (3-5 years) and infeasible implementation. The decision tree classifier can elegantly classify such datasets by only using sequence of univariate inequality constraints (see Table 10.4). The simplicity of using such univariable inequality constraints as a heuristic for decision making is highly effective and convenient for stakeholders. Selecting two of the highest probability pathways for fastest payback (0-3 years), a branch of the tree diagram is generated and shown in Figure 11.

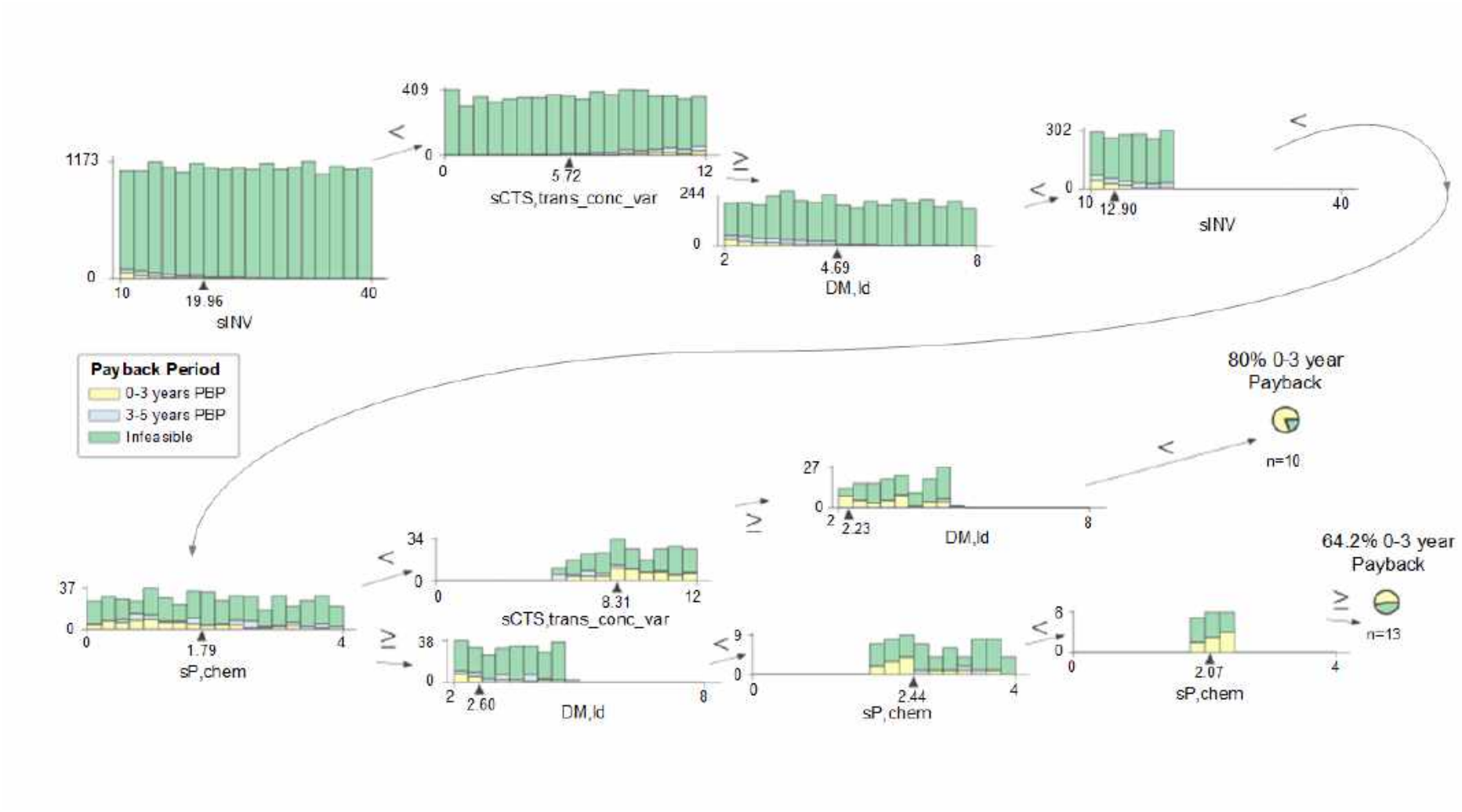


Figure 10.11: Branch of decision tree that indicates interesting pathways for fast payback period

An interesting pathway that was generated is shown in Figure 10.11. The tree suggests that there is a compelling opportunity of between 0-3 years payback with 80% probability. This can be achieved by having specific investment costs per LD below 12.90 EUR/t, DM concentration in LD below 2.23 wt%, specific cost for LD transport and specific price of chemicals over 8.31 EUR/m^3 and below 1.79 EUR/t respectively.

An alternative pathway can also be identified in Figure 10.11, where there is a 64.2% probability of achieving 0-3 years payback. This can be achieved by having similar a specific investment costs as the previous pathway, maintaining the specific price of chemicals between 2.07-2.44 EUR/t while having specific cost for LD transportation and DM concentration in LD below 5.72 EUR/m^3 and 2.6 wt% respectively. Although there is a range for the specific price of chemicals, decision makers should only consider the upper limit, as lower chemical prices give better economic feasibility. Technically, the lower boundary of sP_{chem} gives confinement of the subspace so that the probability of success can be qualitatively assessed. The criteria for each evaporation implementation decision pathway is tabulated in Table 10.5 for a visualized comparison.

Table 10.5: Criteria and reward for pathways

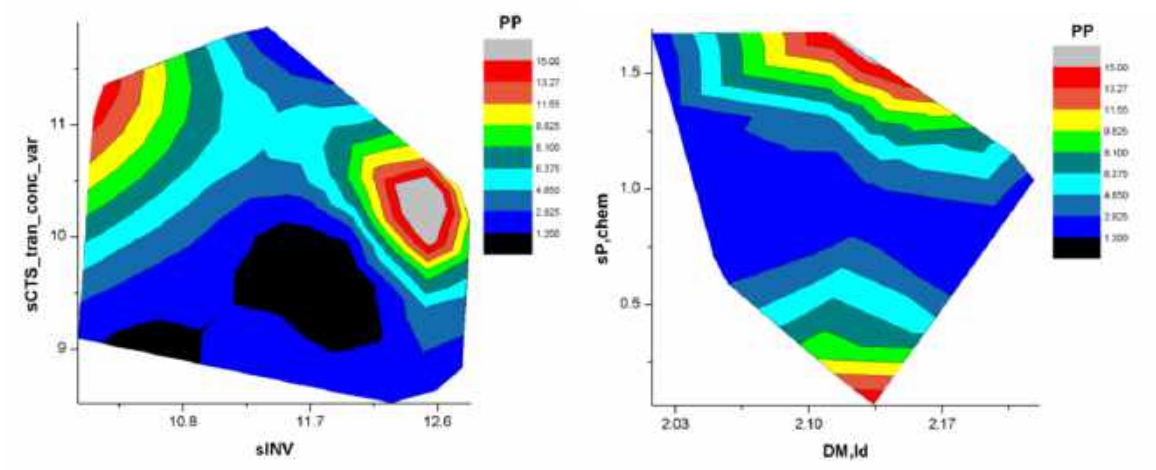
Solutions	**Pathway 1**	**Pathway 2**
Criteria	Specific investment cost per tonne of liquid digestate (EUR/t) <12.90 Specific cost for liquid digestate transport (EUR/m^3) >8.31 Dry matter concentration in liquid digestate (wt%) < 2.23 Specific price of chemicals (EUR/t) <1.79	Specific investment cost per tonne of liquid digestate (EUR/t) <12.90 Specific cost for liquid digestate transport (EUR/m^3>5.72 Dry matter concentration in liquid digestate (wt%) < 2.6 2.07< Specific price of chemicals (EUR/t) <2.44
Payback	0-3 years payback with 80% probability	0-3 years payback with 64.2%

To achieve short PP within an ES implementation project, specific investment cost must be generally lower than a threshold of 12.90 EUR/t. Having a higher specific cost for concentrated LD transport improves the stability of the project as the probability of success improves. The

implementation of ES is also more economically favourable for low DM content of LD and specific price of chemicals below 1.79 and 2.44 EUR/t respectively. Low DM content (< 2.23%, <2.6% respectively) could be achieved by acquiring mechanical separator with high separation efficiency (decanter centrifuge, screw press with fine sieve). The specific price for chemicals depends on the amount of ammonia that must be removed during the thickening process. Avoiding a nitrogen-rich substrate (e.g. poultry manure, grass silage) and requiring efficient management of chemical dosage in ES could be considered as beneficial steps for achieving short PP.

Using transportation distance estimation from Ghafoori et al. (2007), this work estimates that the minimum truck transport distance for pathway 1 is 31.94 km and pathway 2 is 18.7 km. These minimum transportation distance acts as a reference for decision makers if the specific cost for LD transport is unknown. Vilanova Plana and Noche (2016) discussed that the mode of transportation and storage time also affects the specific transportation costs, and it is recommended that such considerations are made by decision makers.

A probabilistic result was presented in the decision tree, this work performs further analysis on the best pathway (Pathway 1) to show the nature and causes of success and failure within the class. In order to study the micro-behaviour of variables within the class of the pathway, the respective solution space has been plotted as contour plots in Figure 10.12.



(a) (b)

Figure 10.12: Solution space of pathway 1 considering (a) specific transportation costs of LD and specific investment costs per tonne of LD (b) specific price of chemicals and DM concentration in LD

The solution space for Pathway 1 is plotted in Figure 10.12, there is an obvious multi-dimensional saddle. The decision tree has successfully found a solution space where there is a stable multi-dimensional local minimum. This saddle can be found in Figure 10.12(a) (where there is a light blue gap) and Figure 10.12(b) (where there is a dark blue gap). For both plots, the contours are smooth and move towards high PP towards the edges of the solution space. This shows that making decisions that are bounded within this solution space has a smooth and resilient outcome and do not cause sudden infeasibility due to dynamic changes. This attribute of the pathway is critically important for real-world implementation, as it gives allowance for slight error during actual project implementation. Pathway 2, on the other hand, only has a 64.2% of having a fast PP (see Table 10.5) which is not robust. Although Pathway 2 can potentially be a "high-risk high-reward" decision option, this work highly recommends making decisions that follow Pathway 1 as it is much more stable (as presented by the solution space contour plots).

10.3.2 "Rules-based" Datasets

Figure 10.13 shows the distribution of PP for all "rules-based" scenarios. Note that only PP from 0 to 15 years is assumed to be feasible. They represent approx. 2.7% and 6.7% of all samples generated in "Base case 2" and "CHP bonus" scenarios, respectively. Long PP, negative cash flow or lack of available heat for evaporation are denoted as infeasible. Obviously, there is a dramatic increase of feasible scenarios in case of "CHP bonus". This is valid for short PP (< 5 years) in particular. Clearly, CHP bonus introduction significantly supports a digestate integration into BGP. In comparison with the "Base case 2" scenario, there is an increase of feasible samples by 251%. Missing market with ASS ("No ASS sale" scenario) does not influence PP significantly. There is a decrease of feasible simulations by 16% against the "Base case 2". This result suggests that ASS marketing is not necessarily a decisive factor for the LD processing project. To be more impactful, the LD treatment must be capable of producing more valuable products or exploit more fertilizer or otherwise marketable substances.

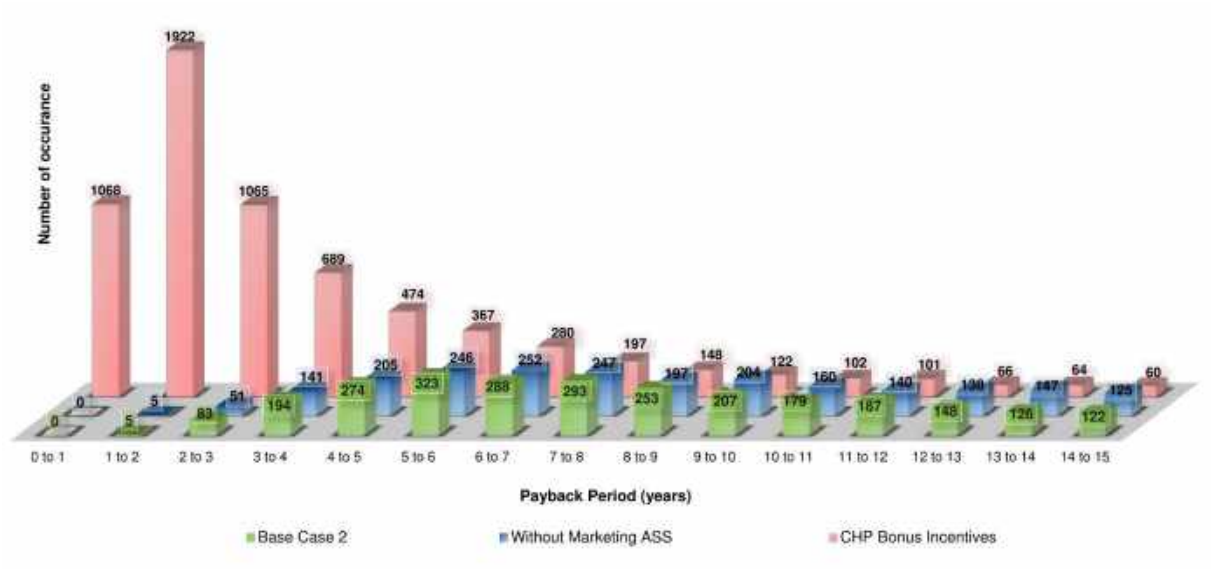
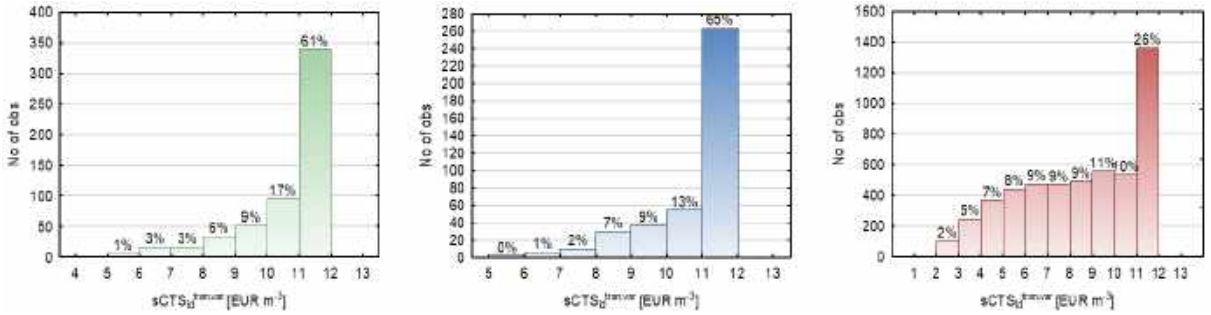


Figure 10.13: Overall difference in payback period occurrence with different scenarios

In the following part, the distribution of individual independent variables is discussed together with the impact of each scenario. The histograms are shown for PP from 0 to 5 years as it is the most attractive range for decision makers. Moreover, the same range was used for classification in the decision tree design.

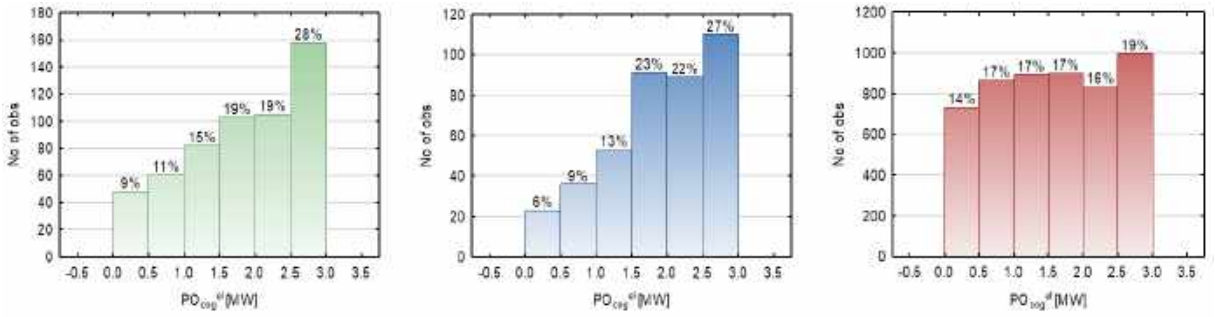
Specific transport costs ($sCTS_{ld}^{tran,var}$) is one of the most important variables. CHP bonus decreases its importance (Figure 10.14); however, it remains quite big - 26 % of feasible samples has a value higher than 11 EUR/m^3. For "Base case 2" and "No ASS sale", the values higher than 11 EUR/m^3 are dominant with more than 60 % of feasible samples. Clearly, without CHP bonus, the LD processing is valuable mainly in regions where high transportation cost due to longer distances is expected, e.g. regions with high density of agriculture production, with a high geographical density of BGPs or regions with legal restrictions.



(a) (b) (c)

Figure 10.14: Distribution of specific LD transportation price in the successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives

Another variable with interesting results is the installed power output of a BGP (PO_{cog}^{el}), which basically describes its capacity. BGP with higher installed electricity output have also high digestate production and therefore lower $sINV$ (see rules – Figure 10.7). At the same time, it is likely that transport cost is higher because the amount of LD is higher and must be transported between longer distances. Since the average installed capacity in EU is around 0.56 MW (Deremince and Königsberger, 2017), the simulated results (Figure 10.15) has shown that there is approx. a 3 times lower chance of a successful implementation of evaporation system into the current industrial settings in comparison with BGPs with installed capacity higher than 2.5 MW. On the other hand, it is obvious that due to economies of scale BGP with high capacities (> 1.0 MW) have higher chance of successful implementation - only 20% of BGPs under 1 MW were able to succeed within the feasible samples in the "Base case 2". Thus, the LD thickening project in an average BGP in the EU will probably be hardly feasible. Centralization of digestate treatment for small (< 1.0 MW) neighbouring BGPs could be advantageous in this regard. Nevertheless, referring Figure 15c, the aforementioned difficult project implementation can be improved with increased CHP bonus which allows smaller capacity BGP to be more successful.

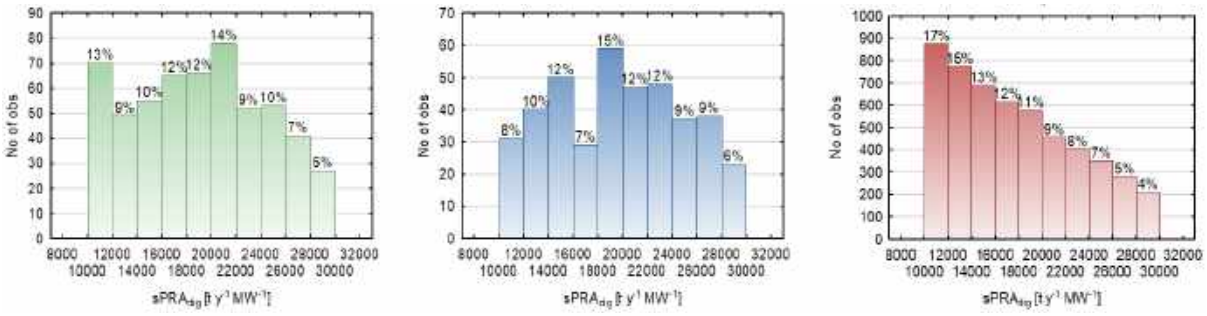


(a) (b) (c)

Figure 10.15: Distribution of installed electric power output in the successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives

Furthermore, the distribution of the values of $sPRA_{dig}$ was investigated (see Figure 10.16). Once again, the effect of increased CHP incentives has a large impact on the distribution. The explanation is as follows, the higher $sPRA_{dig}$ the lower DM_{dig} and therefore more water need to be evaporated. This results in heat demand increase. The samples with high $sPRA_{dig}$ (especially those with > 20 000 t $^{y-1}$ MW^{-1}) can easily become infeasible due to lack of available

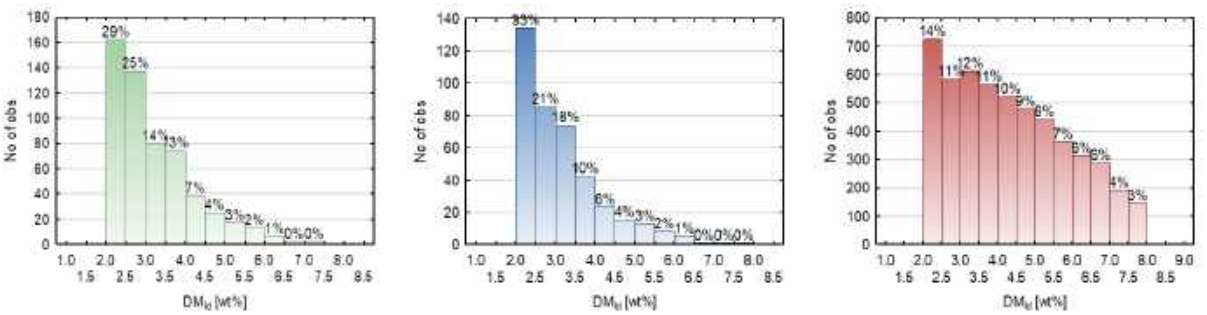
heat compared to the samples with low $sPRA_{dig}$. Evidently, this work identifies that the number of observations is decreasing with increasing $sPRA_{dig}$. This nature is not observable in case of "Base case 2" and "No ASS sale" scenarios. The reason is that CHP bonus makes feasible even those BGPs with low potential of LD volume reduction (high DM_{dig}). So, BGPs with low $sPRA_{dig}$ (high DM_{dig}) are more prone to be infeasible in case of "Base case 2" and "No ASS sale" while they are more feasible in case of "CHP bonus".



(a) (b) (c)

Figure 10.16: Distribution of specific production of digestate in the successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives

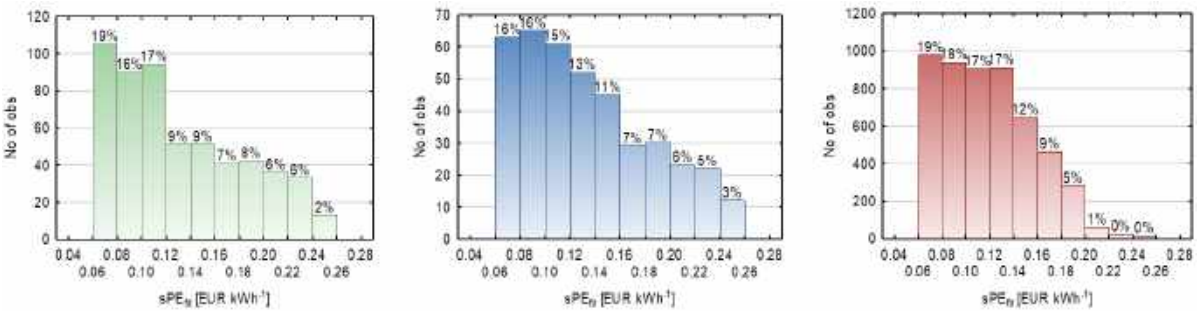
Figure 10.17 shows the distribution of DM_{ld}. Obviously, more efficient mechanical separation (less DM in LD) significantly increases the probability towards a feasible solution. Nearly 90 % of feasible samples has DM_{ld} up to 4.5 wt%. CHP bonus decreases the significance of this variable, but it remains important. Low DM_{ld} enables significant volume reduction in the ES thus contributes to high savings in transport and spreading. For BGPs with high DM content (> 4.5 wt%) in LD it could be beneficial to accordingly adjust the operation of their mechanical separator, e.g. using a finer screen or increasing the resistance force in case of screw press separator.



(a) (b) (c)

Figure 10.17: Distribution of DM concentration in LD in successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives

Electricity feed-in tariffs (sPE_{fit}) play an important role in terms of return on investment. In the "Base case 2" and "No ASS sale" scenarios around 50% of successful samples lie below the value of 0.12 EUR/kWh (Figure 10.18). This suggests that high incentives for electricity production do not promote projects on LD treatment as the power consumption of newly acquired technologies is too costly for BGP operators. Since high feed-in tariffs (> 0.1075 EUR/kWh) are penalized in the "CHP bonus" scenario (see Figure 10.8), the feasibility of samples with low value of sPE_{fit} is even more prevalent in this scenario (Figure 10.18(c)).



(a) (b) (c)

Figure 10.18: Distribution of specific price for electricity (feed-in tariffs) in successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives

The other variables' histograms can be found in the Appendix B, but their distributions are briefly commented in the following part. Majority of the feasible samples (about 90 %) has $sCTS_{ld}^{apl,var}$ higher than 2 EUR/m^3 and, at the same time, the higher $sCTS_{ld}^{apl,var}$ the more feasible samples (Figure B10.1). However, when CHP bonus is introduced, the distribution of the feasible samples becomes uniform. So, the importance of this variable is decreased by CHP. It is also the case of k_{evap}^{mnt} (Figure B10.2). The only difference is that the higher value (up to 0.15) the lesser feasible samples. sP_{chem} (Figure B10.3) is very similar to k_{evap}^{mnt} following the trend the higher value (up to 4 EUR/t) the lesser feasible samples. All these variables have one feature in common – CHP bonus significantly decreases their impact.

On the other hand, the following discussion of results includes the variables with small sensitivity to the introduction of CHP bonus. DM_{dig} (Figure B10.4) seems to be normally distributed along the most frequent value, which is 8 wt%, so no impact is observed. This is also valid for $sPCM_{evap}^{th}$ (Figure B.5) where the most frequent value is about 300 kWh/t. In any case, the availability of a sufficient amount of heat with regard to LD production and ES thermal efficiency is a prerequisite for the success of the project. As far as DM_{conc} is concerned, it follows the rule the higher content (up to 15 wt%) the more feasible samples (Figure B10.6). DM_{sf} (Figure B10.7) has rather uniform distribution within the range from 20 to 35 wt% and its importance is negligible. In the case of k_{bgp}^{th} (Figure B10.8), it holds that the higher value (up to 1) the lesser feasible samples. This is an expected result as the lack of available heat makes a sample infeasible. $sPCE_{ac}^{el}$ has an almost uniform distribution and negligible effect (Figure B10.9). $\Delta sPCE_{agit}^{el}$ has the same nature (Figure B10.10). $s\dot{M}_{ass}$ (Figure B10.11) distribution is influenced by the "No ASS sale" scenario. Naturally, there are more feasible samples when $s\dot{M}_{ass}$ is higher (up to 0.035 wt%) because the product can be sold. It is distributed in the opposite way for the other two scenarios. sP_{ass}^{sale} (Figure B10.12) is only relevant for "Base case 2" and "CHP bonus" scenarios and is almost uniformly distributed with small effect. Specific investment cost $sINV$ (Figure B10.13) is normally distributed among the most frequent value of 25 EUR t^{-1} in the "CHP bonus" scenario. However, it is slightly biased towards lower values (down to 10 EUR/t). In the "Base case 2" and "No ASS sale" scenarios, around 93% and 97% of successful samples respectively lie below the value of 30 EUR per tonne, which makes this value a significant threshold.

10.4 Conclusion

Advancement in digestate treatment systems is the next step towards creating a circular economy. This work recommends the utilization of vacuum evaporator systems as they are robust and effective. Nevertheless, the decision in implementing evaporator systems within existing biogas facilities and infrastructure is non-straightforward. Stakeholders are required to make decisions considering multiple dimensions of varying factors that will affect the plant economics. A novel techno-economic analysis tool was developed for assessing scenarios of the economic feasibility of such retrofitting of evaporator systems. Monte Carlo simulations resulted in a high percentage of infeasible solutions (more than 95% probability), which suggests that project feasibility is highly sensitive to operational parameters and BGPs have a

low randomized probability of being economically justified when investing into the thickening technology. Performing critical variable selection on a trained artificial neural network, four of the most influential variables were identified as the specific investment cost per LD, specific cost for LD transport, DM concentration in LD and specific price of consumed chemicals. A novel decision tree was utilized to search for a set of decision that would yield reliable and fast the payback period. Stakeholders can also conveniently utilize the decision tree as it provides decision-making heuristic in the form of sequential univariate inequality. Two feasible pathways were suggested by the decision tree to achieve a fast payback period (below 3 years). Analysing the solution space, one of the pathways gave a resilient and stable solution with 0-3 year payback period and probability of 80%, while the other pathway gave a "high-risk high-reward" solution with 0-3 year payback period and probability of 64.2%. These pathways can be easily achieved by planning the technical implementation and financial investment into such project carefully.

Statistical evaluation of the impact of each scenario for rules-based datasets was also investigated. It showed that only about 2.5% and 2% of samples are feasible with PP up to 5 years (the most attractive) for "Base case 2" and "No ASS sale" scenario. On the other hand, it is about 10% for "CHP bonus" scenario. Detailed analysis of the distribution of independent variables was also carried out. It revealed several main conclusions:

- high transport cost (> 8 EUR m^{-3}) is almost the necessary condition without CHP bonus,
- neighbouring small-scale BGPs should implement centralized treatment, which has lower specific investment/maintenance cost,
- even BGPs with low potential for reduction of LD volume (DM concentration in LD > 4.5 wt%) have short PP when CHP bonus is introduced; generally, the introduction of CHP bonus has a strongly positive effect (increase of feasible projects by 251%);
- efficient mechanical separation (DM concentration in LD < 4.5 wt%) has a very positive impact on PP,
- high feed-in tariffs (> 0.12 EUR/kWh) have a negative impact on the project's feasibility,
- ASS production and sale have slightly positive effect on PP, but in most cases not the decisive effect.

The results of this work are published in hopes to aid the decision making for the integration of evaporation systems, thus reducing untreated LD residues within the biogas industry.

Furthermore, it is shown that a data-driven approach (using Monte Carlo, neural networks, and decision tree) is effective in evaluating the techno-economics of the integration of a new operational unit in a existing industry or facility.

Nomenclature

Abbreviation	**Definition**
$\boldsymbol{CTS}$	Costs, EUR
$\boldsymbol{DM}$	Dry matter content, wt%
$\boldsymbol{INV}$	Investment costs, EUR
$\boldsymbol{k}$	Coefficient, -
$\boldsymbol{\dot{M}}$	Mass flow rate, kg h^{-1}
$\boldsymbol{PC}$	Power consumption, kW
$\boldsymbol{PO}$	Power output, MW
$\boldsymbol{PP}$	Payback period, y
$\boldsymbol{REV}$	Revenues, EUR
$\boldsymbol{SH}$	Service hours, h y^{-1}
$\boldsymbol{sCTS}$	Specific costs, EUR m^{-3}
$\boldsymbol{sINV}$	Specific investment costs, EUR t^{-1}
$\boldsymbol{s\dot{M}}$	Specific mass flow rate, wt%
$\boldsymbol{sP}$	Specific price, EUR t^{-1}
$\boldsymbol{sPCE}$	Specific power consumption, kW MW^{-1}
$\boldsymbol{sPCM}$	Specific power consumption, kWh t^{-1}
$\boldsymbol{sPE}$	Specific price of electricity, EUR kWh^{-1}
$\boldsymbol{sPRA}$	Specific production annual, t y^{-1} MW^{-1}
$\boldsymbol{\dot{V}}$	Volumetric flow rate, m^3 h^{-1}
$\boldsymbol{\eta}$	Efficiency, -
$\boldsymbol{\rho}$	Density, kg m^{-3}
Mathematical Indices	**Definition**
$\boldsymbol{ac}$	Air-cooled chillers
$\boldsymbol{agit}$	Agitators in storage tanks
$\boldsymbol{apl}$	Field application

$\boldsymbol{ass}$	Ammonium Sulphate solution
$\boldsymbol{bgp}$	Biogas plant
$\boldsymbol{chem}$	Chemicals
$\boldsymbol{chp}$	Combined heat and power
$\boldsymbol{conc}$	Concentrate
$\boldsymbol{cog}$	Cogeneration unit
$\boldsymbol{dig}$	Digestate
$\boldsymbol{el}$	Electric power
$\boldsymbol{evap}$	Evaporation technology
$\boldsymbol{fit}$	Feed-in tariff
$\boldsymbol{fw}$	Fresh water
$\boldsymbol{ld}$	Liquid digestate
$\boldsymbol{mnt}$	Technology maintenance
$\boldsymbol{new}$	New scenario (with $evap$)
$\boldsymbol{orig}$	Original scenario (without $evap$)
$\boldsymbol{sale}$	Sale of product
$\boldsymbol{sf}$	Solid fraction
$\boldsymbol{strg}$	Storage
$\boldsymbol{th}$	Thermal power
$\boldsymbol{tran}$	Transportation
$\boldsymbol{var}$	Variable (costs)

References

Al Seadi, T., Drosg, B., Fuchs, W., Rutz, D., Janssen, R., 2013. Biogas digestate quality and utilization, in: The Biogas Handbook. Elsevier, pp. 267–301. doi.org/10.1533/9780857097415.2.267

Auburger, S., Wustholz, R., Petig, E., Bahrs, E., 2015. Biogas digestate and its economic impact on farms and biogas plants according to the upper limit for nitrogen spreading—the case of nutrient-burdened areas in north-west Germany. AIMS Energy 3, 740–759. doi.org/10.3934/energy.2015.4.740

Azouma, Y.O., Jegla, Z., Reppich, M., Turek, V., Weiß, M., 2018. Using agricultural waste for biogas production as a sustainable energy supply for developing countries. Chem. Eng. Trans. 70, 445–450. doi.org/10.3303/CET1870075

Bamelis, L., Blancke, M.A., Camargo-Valero, L., De Clercq, A., Haumont, B., De Keulenaere, F., Delvigne, E., 2015. Techniques for nutrient recovery from digestate derivatives. www.biorefine.eu/sites/default/files/publication-uploads/wp2a5_gxabt_20151216_recovery_techniques_-_digestate.pdf (accessed 7 May 2019).

Berglund, M., Börjesson, P., 2006. Assessment of energy performance in the life-cycle of biogas production. Biomass Bioenergy 30, 254–266. doi.org/10.1016/j.biombioe.2005.11.011

Bolzonella, D., Fatone, F., Gottardo, M. and Frison, N., 2018. Nutrients recovery from anaerobic digestate of agro-waste: Techno-economic assessment of full scale applications. J Environ Manage, 216, pp.111-119. doi: 10.1016/j.jenvman.2017.08.026

CDPQ Decision Tools, 2018. Centre de developpement du porc du Quedec inc. //www.cdpq.ca/outils-d-aide-a-la-decision.aspx (accessed 9 May 2019)

Chen, Y.C., Yang, Z.M., Chen, Q.H., Jiang, X.L., Gao, M. and Xia, Q., 2009. An overview on disposal of anaerobic digestate for large scale biogas engineering. China biogas, *28*(1), pp.14-20.

Chiumenti, A., da Borso, F., Chiumenti, R., Teri, F., Segantin, P., 2013. Treatment of digestate from a co-digestion biogas plant by means of vacuum evaporation: Tests for process optimization and environmental sustainability. Waste Manag. 33, 1339–1344. doi.org/10.1016/j.wasman.2013.02.023

Cordes, F., 2018. Vacuum evaporation as system provider for flexible biogas plants, in: Progress in the Treatment and Application of Manure and Digestate Products. Presented at the Progress Manure and Digestate 2018, German Biogas and Bioenergy Society (GERBIO), Schwäbisch Hall, Germany.

CZBA - Czech Biogas Association, 2014. Strategická výzkumná agenda oboru bioplyn. www.czba.cz/files/ceska-bioplynova-asociace/uploads/files/SVA_CzBA_2014_FINAL.pdf (accessed 7 May 2019).

Dahlin, J., Herbes, C. and Nelles, M., 2015. Biogas digestate marketing: Qualitative insights into the supply side. Resour Conserv Recycl, *104*, pp.152-161. doi: 10.1016/j.resconrec.2015.08.013

Daniel-Gromke, J., Rensberg, N., Denysenko, V., Trommler, M., Reinholz, T., Völler, K., Beil, M., Beyrich, W., 2017. Anlagenbestand Biogas und Biomethan – Biogaserzeugung und nutzung in Deutschland. DBFZ Deutsches Biomasseforschungszentrum gemeinnützige GmbH, Leipzig.

Dasa, K.T., Westman, S.Y., Millati, R., Cahyanto, M.N., Taherzadeh, M.J., Niklasson, C., 2016. Inhibitory Effect of Long-Chain Fatty Acids on Biogas Production and the Protective Effect of Membrane Bioreactor. BioMed Res. Int. 2016, 1–9. doi.org/10.1155/2016/7263974

Deremince, B., Königsberger, S., 2017. Statistical Report of the European Biogas Association 2017. Brussels, Belgium. european-biogas.eu/wp-content/uploads/2017/12/Statistical-report-of-the-European-Biogas-Association_

excerpt-web.pdf (accessed 14 December 2018).

Drosg, B., Fuchs, W., Al Seadi, T., Madsen, M., Linke, B., 2015. Nutrient Recovery by Biogas Digestate Processing. IEA Bioenergy.

Đurđević, D., Blecich, P., Lenić, K., 2018. Energy Potential of Digestate Produced by Anaerobic Digestion in Biogas Power Plants: The Case Study of Croatia. Environ. Eng. Sci. 35, 1286–1293. doi.org/10.1089/ees.2018.0123

European Commission, 2018. Energy - Cogeneration of heat and power. ec.europa.eu/energy/en/topics/energy-efficiency/cogeneration-heat-and-power (accessed 21 January 2019).

European Commission, 2019a. Horizon 2020, Work Programme 2018-2020. ec.europa.eu/programmes/horizon2020/en/what-work-programme (accessed 10 May 2019).

European Commission, 2019b. Renewable energy policy database and support. www.res-legal.eu/search-by-country/finland/single/ (accessed 21 January 2019).

Fagerström, A., Seadi, T.A., Rasi, S., Briseid, T., 2018. The role of Anaerobic Digestion and Biogas in the Circular Economy. IEA Bioenergy Task 37 2018 8.

Frischmann, P., 2012. Enhancement and treatment of digestates from anaerobic digestion (Desk top study on digestate enchancement and treatment). WRAP. www.wrap.org.uk/content/enhancement-and-treatment-digestates-anaerobic-digestion (accessed 10 May 2019).

Fuchs, W., Drosg, B., 2013. Assessment of the state of the art of technologies for the processing of digestate residue from anaerobic digesters. Water Sci. Technol. 67, 1984–1993. doi.org/10.2166/wst.2013.075

Gebrezgabher, S., Meuwissen, M., Prins, B. and Lansink, A., 2010. Economic analysis of anaerobic digestion—A case of Green power biogas plant in The Netherlands. NJAS - Wageningen Journal of Life Sciences, 57, 109-115.

Ghafoori, E., Flynn, P. and Feddes, J., 2007. Pipeline vs. truck transport of beef cattle manure. Biomass Bioenergy, 31, 168-175. doi: 10.1016/j.biombioe.2006.07.007

Golkowska, K., Vázquez-Rowe, I., Lebuf, V., Accoe, F. and Koster, D., 2014. Assessing the treatment costs and the fertilizing value of the output products in digestate treatment systems. Wat Sci Tech, 69(3), pp.656-662. doi: 10.2166/wst.2013.742

Gómez, X., Cuetos, M.J., Cara, J., Morán, A., García, A.I., 2006. Anaerobic co-digestion of primary sludge and the fruit and vegetable fraction of the municipal solid wastes. Renew. Energy 31, 2017–2024. doi.org/10.1016/j.renene.2005.09.029

Guercini, S., Castelli, G., Rumor, C., 2014. Vacuum evaporation treatment of digestate: full exploitation of cogeneration heat to process the whole digestate production. Wat Sci Tech 70, 479–485. doi.org/10.2166/wst.2014.247

Hung, C.Y., Tsai, W.T., Chen, J.W., Lin, Y.Q. and Chang, Y.M., 2017. Characterization of biochar prepared from biogas digestate. Waste Manage, 66, pp.53-60. doi: 10.1016/j.wasman.2017.04.034

Kirchherr, J., Reike, D., Hekkert, M., 2017. Conceptualizing the circular economy: An analysis of 114 definitions. Resour. Conserv. Recycl. 127, 221–232. doi.org/10.1016/j.resconrec.2017.09.005

Koszel, M. and Lorencowicz, E., 2015. Agricultural use of biogas digestate as a replacement fertilizers. Agric. Agric. Sci. Procedia, 7, pp.119-124. doi.org/10.1016/j.aaspro.2015.12.004

Liguori, R., Faraco, V., 2016. Biological processes for advancing lignocellulosic waste biorefinery by advocating circular economy. Bioresour. Technol. 215, 13–20. doi.org/10.1016/j.biortech.2016.04.054

Ma, H., Guo, Y., Qin, Y., Li, Y.-Y., 2018. Nutrient recovery technologies integrated with energy recovery by waste biomass anaerobic digestion. Bioresour. Technol. 269, 520–531. doi.org/10.1016/j.biortech.2018.08.114

Maier, C., 2018. Vapogant - digestate evaporator, in: Progress in the Treatment and Application of Manure and Digestate Products. Presented at the Progress Manure and Digestate 2018, German Biogas and Bioenergy Society (GERBIO), Schwäbisch Hall, Germany.

Máša, V., Bobák, P., Kuba, P., Stehlík, P., 2013. Analysis of energy efficient and environmentally friendly technologies in professional laundry service. Clean Techn Environ Policy 15, 445–457. doi.org/10.1007/s10098-013-0618-2

Melse, R.W., Verdoes, N., 2005. Evaluation of Four Farm-scale Systems for the Treatment of Liquid Pig Manure. Biosyst. Eng. 92, 47–57. doi.org/10.1016/j.biosystemseng.2005.05.004

Miltner, M., Makaruk, A., Harasek, M., 2013. Biomethane-Calculator, IEEE-project BioMethane Regions. bio.methan.at./?q=de/download_biomethane-calculator (accessed 10 May 2019).

Møller, H.B., Lund, I., Sommer, S.G., 2000. Solid–liquid separation of livestock slurry: efficiency and cost. Bioresour. Technol. 74, 223–229. doi.org/10.1016/S0960-8524(00)00016-X

Møller, H.B., Sommer, S.G., Ahring, B.K., 2002. Separation efficiency and particle size distribution in relation to manure type and storage conditions. Bioresour. Technol. 85, 189–196. doi.org/10.1016/S0960-8524(02)00047-0

Monfet, E., Aubry, G., Ramirez, A.A., 2018. Nutrient removal and recovery from digestate: a review of the technology. Biofuels 9, 247–262. doi.org/10.1080/17597269.2017.1336348

Monlau, F., Sambusiti, C., Ficara, E., Aboulkas, A., Barakat, A., Carrère, H., 2015. New opportunities for agricultural digestate valorization: current situation and perspectives. Energy Environ. Sci. 8, 2600–2621. doi.org/10.1039/C5EE01633A

Noel, V., Fourcroy, J., 2017. Eurovent rating standard for DX air coolers, air cooled condensers, dry coolers: RS 7/C/008 - 2017. Eurovent Certita Certification SAS. www.eurovent-certification.com/fic_bdd/en/1494596601_S02_D04_ECP-HE_2017_RS-7C008.pdf (accessed 18 December 2018)

Pablo-Romero, M. del P., Sánchez-Braza, A., Salvador-Ponce, J., Sánchez-Labrador, N., 2017. An overview of feed-in tariffs, premiums and tenders to promote electricity from biogas in the EU-28. Renewable Sustainable Energy Rev. 73, 1366–1379. doi.org/10.1016/j.rser.2017.01.132

Potting, J., Hekkert, M., Worrell, E., Hanemaaijer, A., 2017. Circular Economy: Measuring innovation in the product chain (Policy Report No. 2544). PBL Netherlands Environmental Assessment Agency. www.pbl.nl/sites/default/files/cms/publicaties/pbl-2016-circular-economy-measuring-innovation-in-product-chains-2544.pdf (accessed 10 May 2019).

Ragazzi, M., Maniscalco, M., Torretta, V., Ferronato, N., Rada, E.C., 2017. Anaerobic digestion as sustainable source of energy: A dynamic approach for improving the recovery of organic waste. Energy Procedia 119, 602–614. doi.org/10.1016/j.egypro.2017.07.086

Rehl, T., Müller, J., 2011. Life cycle assessment of biogas digestate processing technologies. Resour. Conserv. Recycl. 56, 92–104. doi.org/10.1016/j.resconrec.2011.08.007

Scarlat, N., Dallemand, J.-F., Fahl, F., 2018. Biogas: Developments and perspectives in Europe. Renew. Energy 129, 457–472. doi.org/10.1016/j.renene.2018.03.006

Tampio, E., Marttinen, S., Rintala, J., 2016. Liquid fertilizer products from anaerobic digestion of food waste: mass, nutrient and energy balance of four digestate liquid treatment systems. J. Clean. Prod. 125, 22–32. doi.org/10.1016/j.jclepro.2016.03.127

Uçkun Kiran, E., Stamatelatou, K., Antonopoulou, G., Lyberatos, G., 2016. Production of biogas via anaerobic digestion, in: Handbook of Biofuels Production. Elsevier, pp. 259–301. doi.org/10.1016/B978-0-08-100455-5.00010-2

van Haeff, J., 2015. The Role of Bio-waste in the Emerging Circular Economy. www.compostnetwork.info/wordpress/wp-content/uploads/John-van-Haeff.pdf (accessed 10 May 2019).

Vaneeckhaute, C., Lebuf, V., Michels, E., Belia, E., Vanrolleghem, P.A., Tack, F.M.G., Meers, E., 2017. Nutrient Recovery from Digestate: Systematic Technology Review and Product Classification. Waste Biomass Valorization 8, 21–40. doi.org/10.1007/s12649-016-9642-x

Venkata Mohan, S., Nikhil, G.N., Chiranjeevi, P., Nagendranatha Reddy, C., Rohit, M.V., Kumar, A.N., Sarkar, O., 2016. Waste biorefinery models towards sustainable circular bioeconomy: Critical review and future perspectives. Bioresour. Technol. 215, 2–12. doi.org/10.1016/j.biortech.2016.03.130

Vilanova Plana, P. and Noche, B. (2016). A review of the current digestate distribution models: storage and transport. Waste Manage 202, pp.345-357. doi: 10.2495/WM160311

Vondra, M., Máša, V., Bobak, P., 2016. The potential for digestate thickening in biogas plants and evaluation of possible evaporation methods. Chem. Eng. Trans. 787–792. doi.org/10.3303/CET1652132

Vondra, M., Máša, V., Bobák, P., 2018a. The energy performance of vacuum evaporators for liquid digestate treatment in biogas plants. Energy, Process Integration for Energy Saving and Pollution Reduction – PRES 2016 146, 141–155. doi.org/10.1016/j.energy.2017.06.135

Vondra, M., Máša, V., Tous, M., Konecna, E., 2018b. Vacuum evaporation of a liquid digestate from anaerobic digestion: a techno-economic assessment. Chem. Eng. Trans. 769–774. doi.org/10.3303/CET1870129.

Xia, A. and Murphy, J.D., 2016. Microalgal cultivation in treating liquid digestate from biogas systems. Trends Biotechnol., *34*(4), pp.264-275. doi: 10.1016/j.tibtech.2015.12.010

Zhang, Q., Hu, J., Lee, D.-J., 2016. Biogas from anaerobic digestion processes: Research updates. Renew. Energy 98, 108–119. doi.org/10.1016/j.renene.2016.02.029

Appendix

A. Full Decision Tree Diagrams

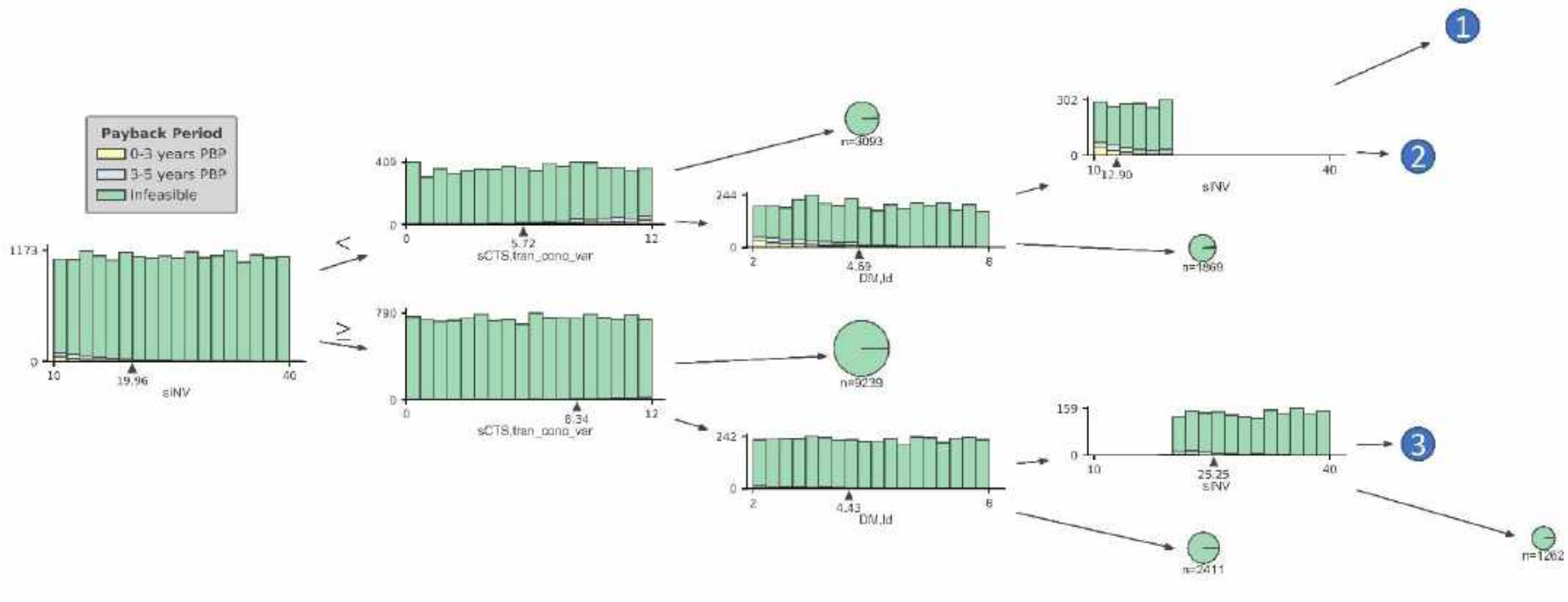


Figure A10.1: Part 1 of the full decision tree classifier

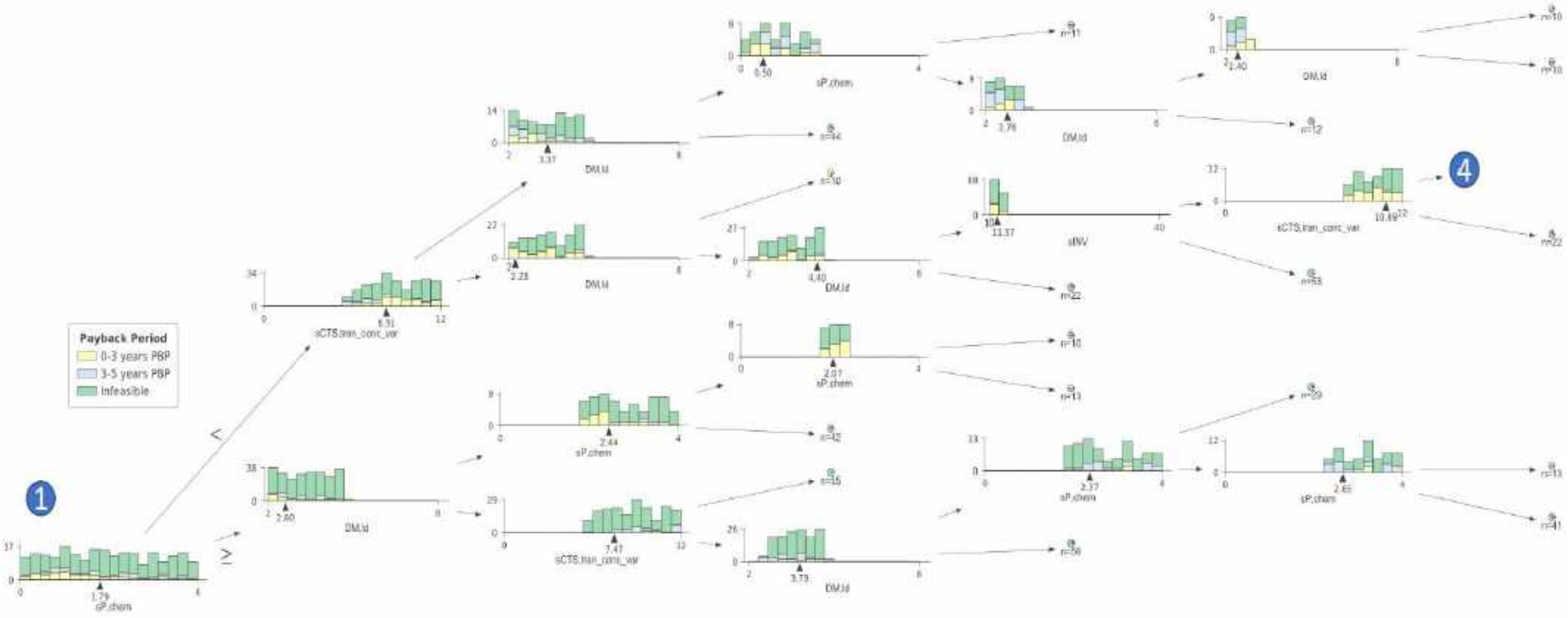


Figure A10.2: Part 2 of the full decision tree classifier

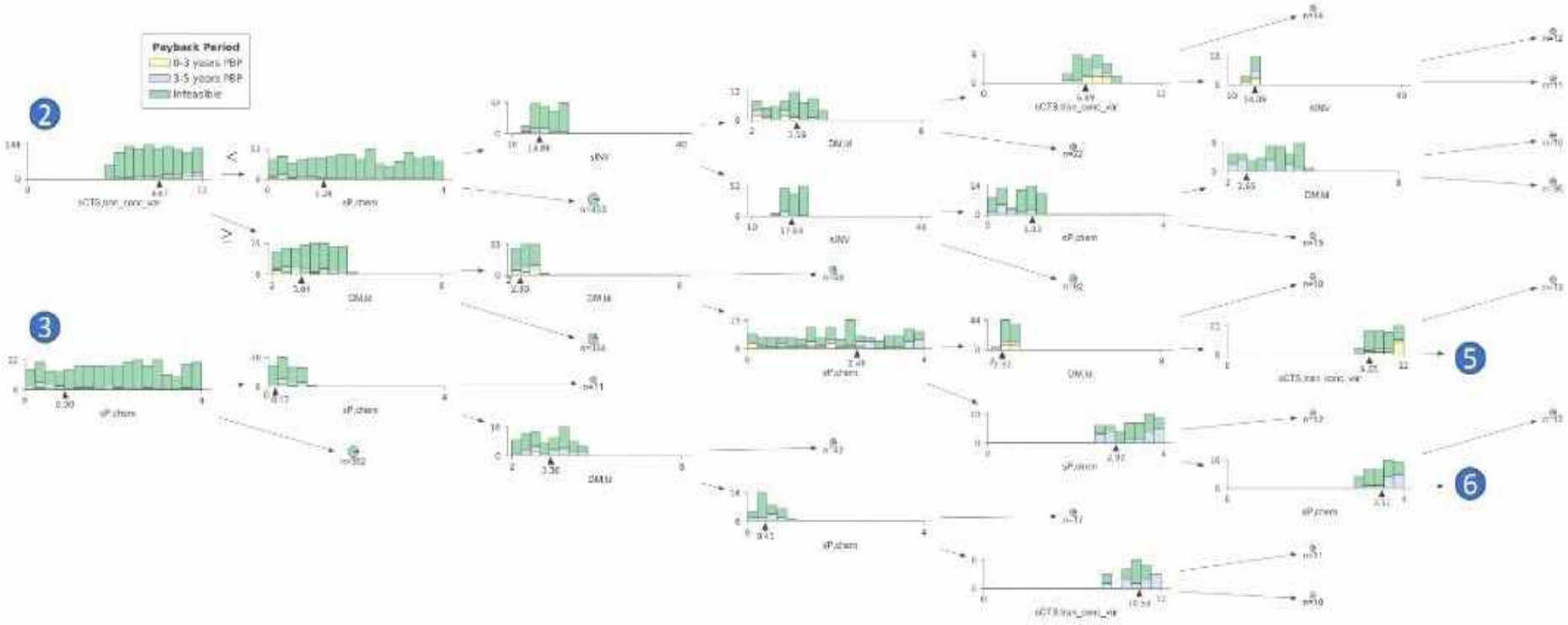


Figure A10.3: Part 3 of the full decision tree classifier

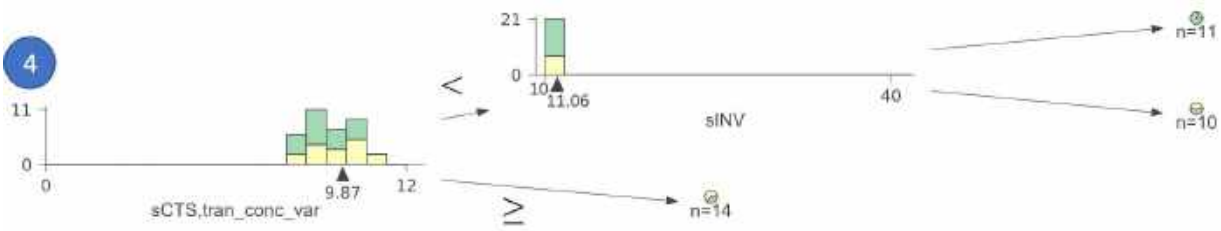


Figure A10.4: Part 4 of the full decision tree classifier

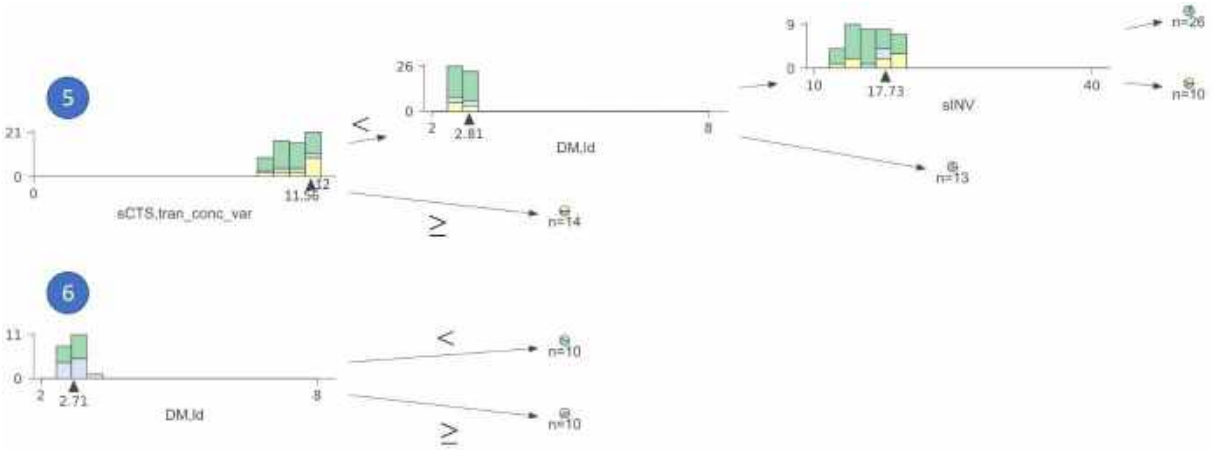
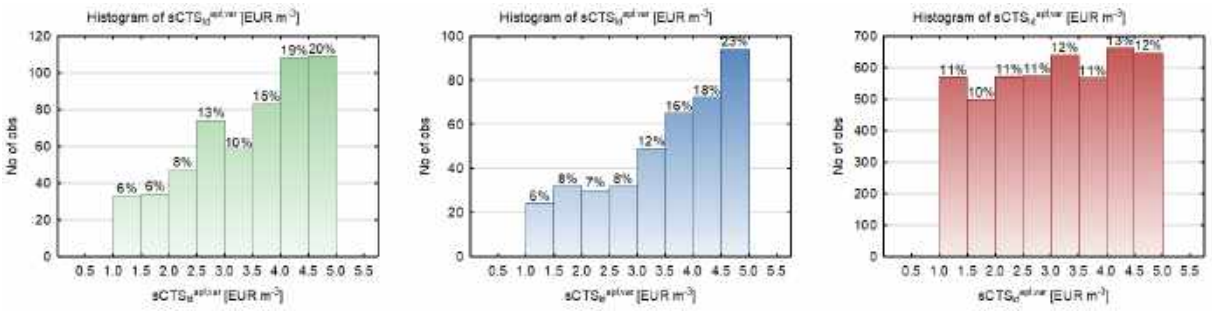


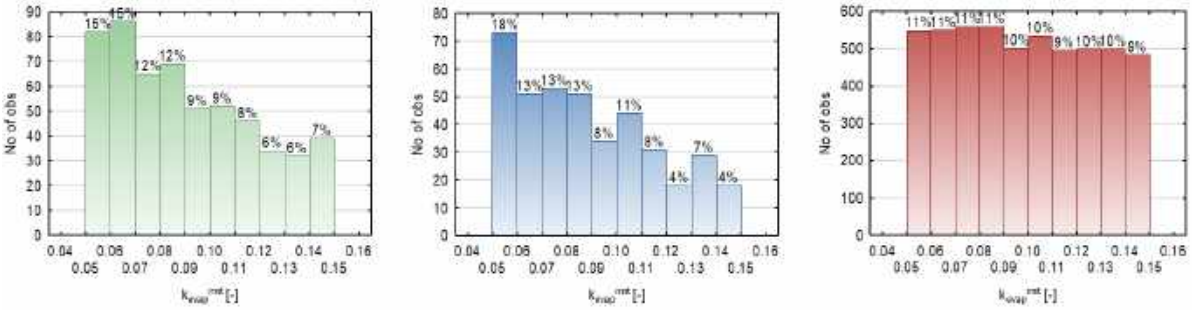
Figure A10.5: Part 5 of the full decision tree classifier

B. Histograms for the "rules-based" datasets



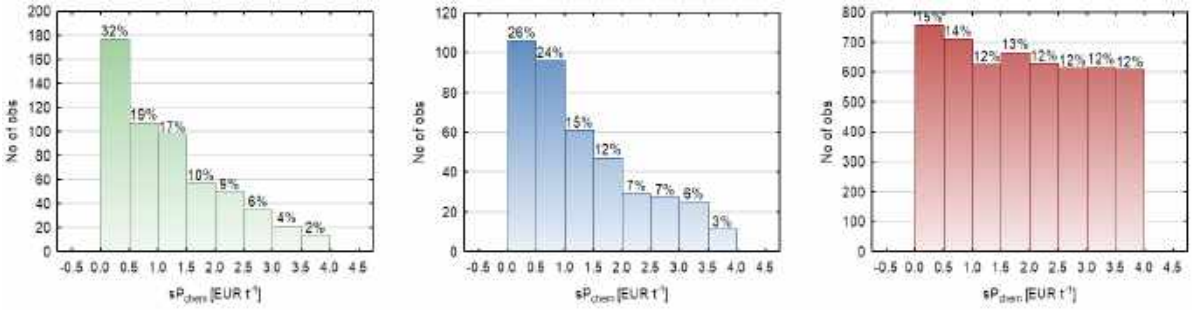
(a) (b) (c)

Figure B10.1: Distribution of specific costs for LD application (spreading) in successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives



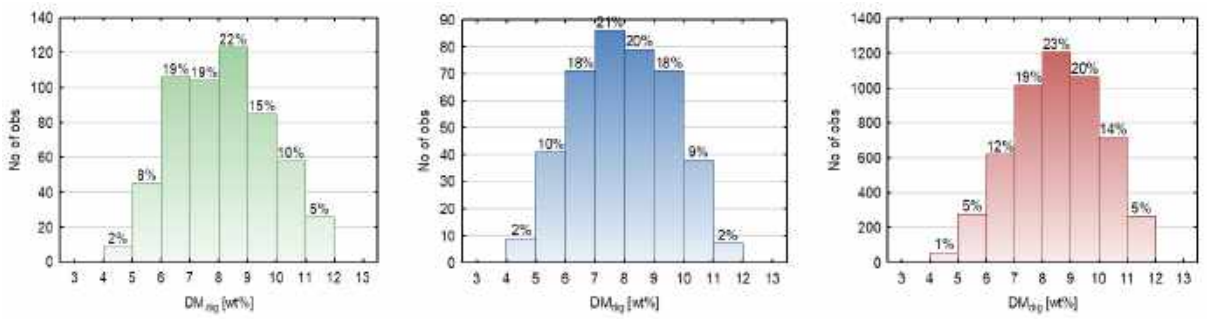
(a) (b) (c)

Figure B10.2: Distribution of values of the maintenance coefficient in successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives



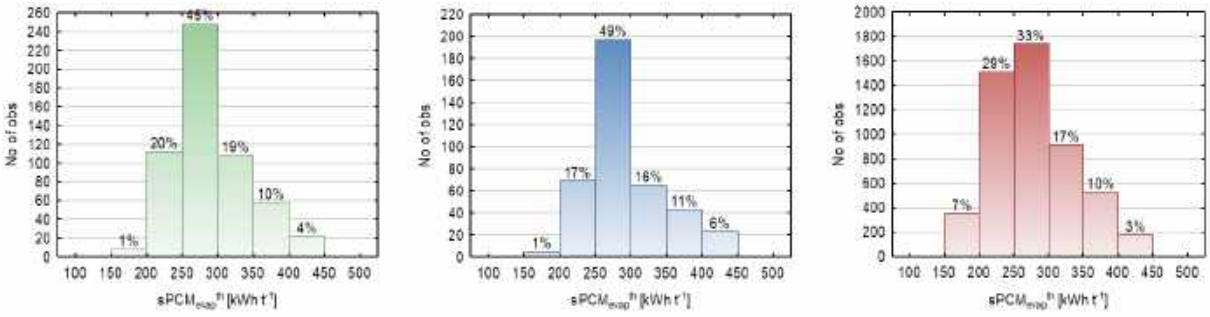
(a) (b) (c)

Figure B10.3: Distribution of the specific price of chemicals in successful scenario of (a) "Base case 2" (b) "No ASS sale" (c) "CHP bonus" incentives



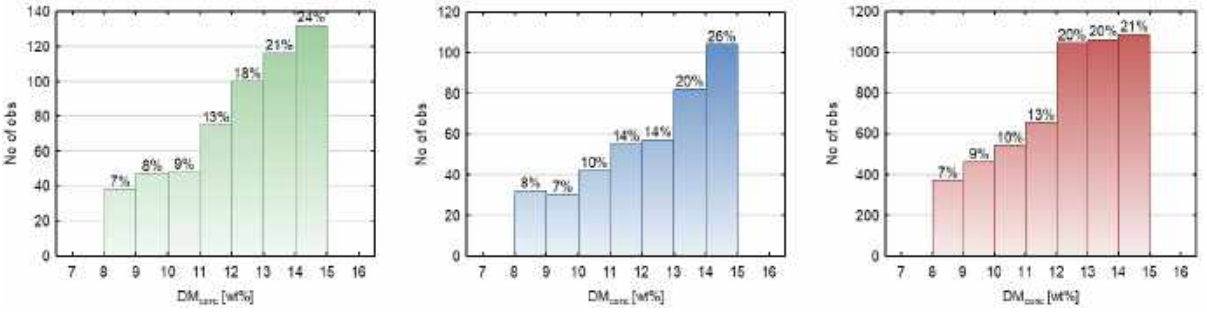
(a) (b) (c)

Figure B10.4: Distribution of dry matter concentration in digestate in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



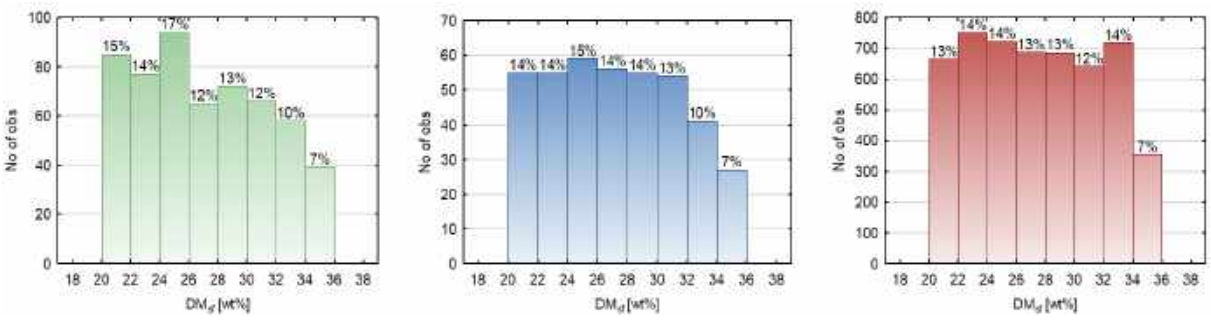
(a) (b) (c)

Figure B10.5: Distribution of specific thermal energy consumption of an evaporation system in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



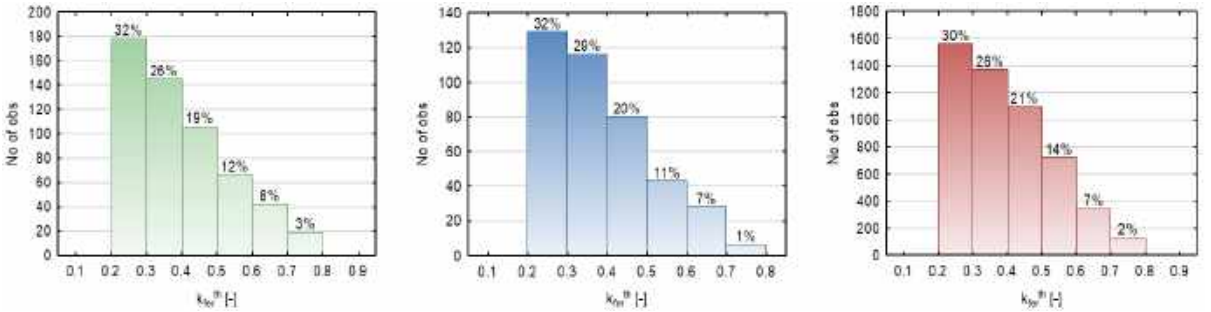
(a) (b) (c)

Figure B10.6: Distribution of dry matter concentration in concentrated (thickened) LD in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



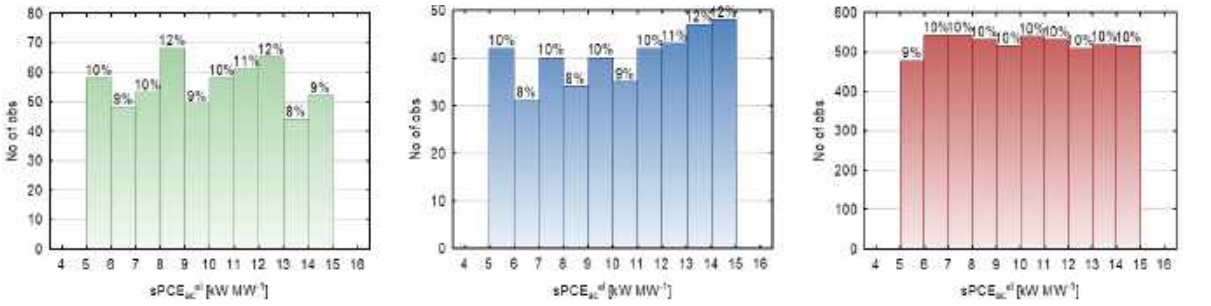
(a) (b) (c)

Figure B10.7: Distribution of dry matter concentration in solid fraction of digestate in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



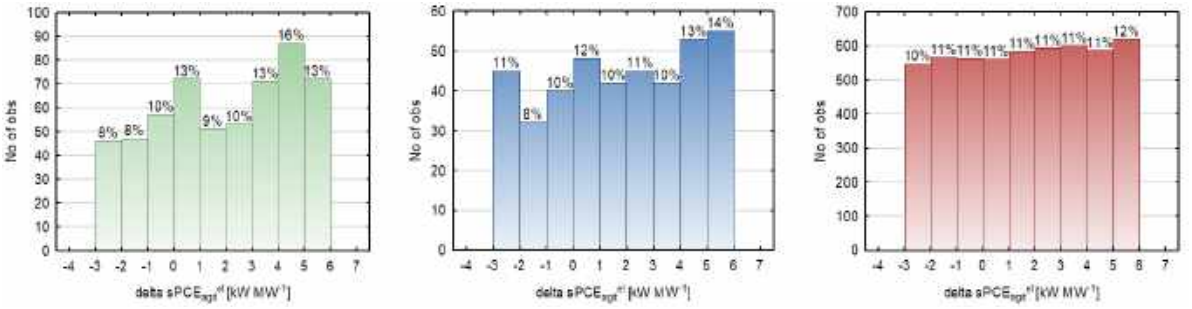
(a) (b) (c)

Figure B10.8: Distribution of the coefficient of heat consumption in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



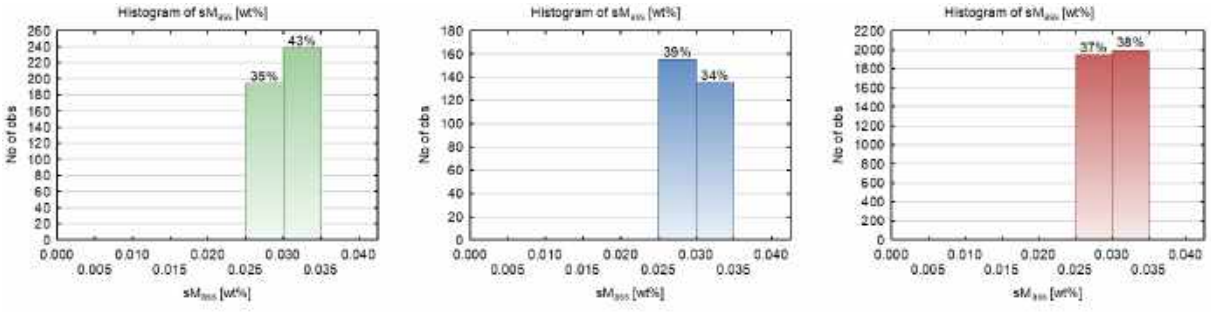
(a) (b) (c)

Figure B10.9: Distribution of specific electricity consumption of air-cooled chillers in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



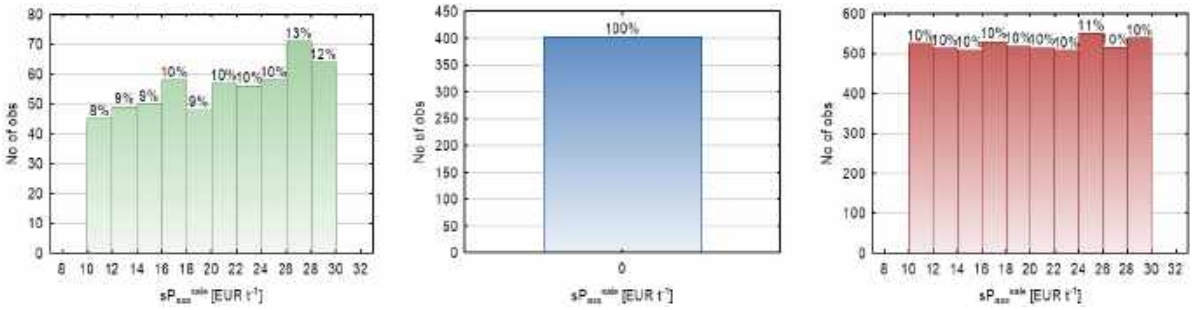
(a) (b) (c)

Figure B10.10: Distribution of the change in specific electricity consumption of agitators in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



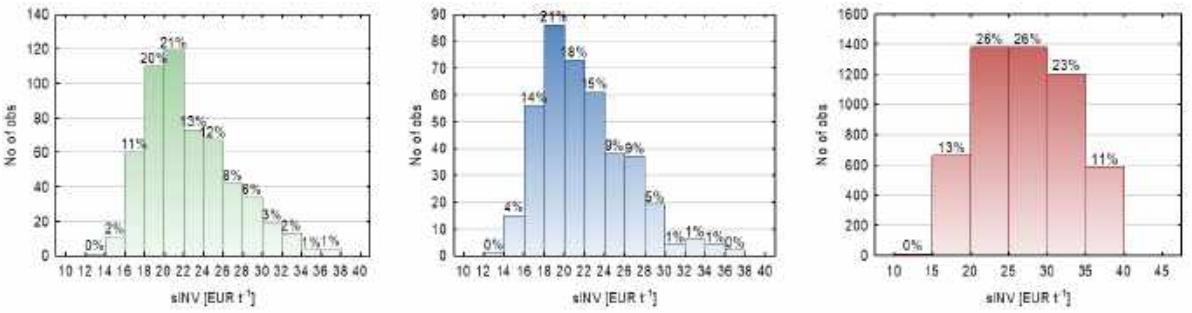
(a) (b) (c)

Figure B10.11: Distribution of specific production of ammonium sulphate solution in successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives



(a) (b) (c)

Figure B10.12: Distribution of specific market price for ammonium sulphate solution in successful scenario of (a) “Base case 2” (b) “No ASS sale”- not relevant (c) “CHP bonus” incentives



(a) (b) (c)

Figure B10.13: Distribution of specific investment cost for an evaporation system successful scenario of (a) “Base case 2” (b) “No ASS sale” (c) “CHP bonus” incentives

CHAPTER 11 ANOMALY-AWARE PREDICTIVE ANALYSIS FOR OPTIMAL REAL-TIME PROCESS MANAGEMENT

This paper is currently under review in Waste Management Journal.

Abstract

Current research shows that operational management of waste-to-energy systems are done using methodologies such as generic regression models, Fourier analysis or deep neural networks. However, traditional methods fall short in simultaneously providing accuracy, adaptability, and robustness for industrial standards. This work proposes a smart energy framework based-on hierarchical temporal memory (HTM), which is a machine learning methodology that was developed as a step towards achieving artificial general intelligence (AGI). As the counterpart of deep neural networks, the HTM model replicates the mechanism in cerebral neocortex of the human brain (which is used for logical reasoning). HTM uses sparse distributed representation and spatial pooling to learn from continuous streams of unlabelled data with great robustness and high capacity. In this work, the HTM-based smart energy framework is demonstrated in an industrial case study for the turbine electricity generation of a waste-to-energy cogeneration system. The smart energy framework is also implemented on an industrial illustrative case study to demonstrate its effectiveness in predicting energy requirements and operational anomalies. This work benchmarked that HTM was able to achieve mean squared error (MSE) of 0.08466 % while giving 35450 Euro profit in half a year. Coupled with a novel dual-mode optimization procedure, HTM demonstrated 11 % improvement with respect to only predictive optimization (with HTM) and a 27 % improvement in estimated gross profit with respect to a human operator in a waste-to-energy plant.

Keywords: Hierarchical Temporal Memory (HTM), machine learning, neural networks, waste-to-energy, cortical learning algorithms, energy production optimisation.

11.1 Introduction

Problems related to energy allocation has shown great importance with the recent integration of renewable sources (Zobaa and Bansal, 2011) and waste utilization (Chae et al., 2010) within existing energy networks. Within the current energy industry, Stehlík (2016) highlighted the potential for waste-to-energy application and demonstrated their potential for waste utilization in a large region. Waste-to-energy also provides emission minimization in the direction of sustainable development by effectively processing municipal and biomass wastes (Stehlík, 2009a). For industrial application, Máša et al. (2018) have shown the importance of energy management models in achieving energy-savings during practical engineering project implementations. Similarly, Yao et al. (2016) have also agreed with this approach from a mathematical perspective, showing that predictive analytics can be used in energy-savings, resource planning and optimization. With the ever-growing demands of energy management, the needs for better energy prediction and optimization models are required for renewable energy and waste utilization (Kuznetsova et al., 2019; Lindberg et al., 2014).

The nature of energy prediction models is very much different than other prediction models (e.g. stock market prediction), in the sense that there is a need for robustness to ensure energy security (Winzer, 2012). Energy prediction models operate in a robust manner to avoid catastrophic failures in energy systems (Arabian-Hoseynabadi et al., 2010). Shoreh et al. (2016) discussed that the main barriers to implementing such a responsive structure for industrial applications possess three main barriers, which are financial barriers, regulatory barriers and knowledge-based barriers. The implementation of energy management systems within waste-to-energy industries is critical as it is one of the most crucial industry supporting sustainable waste management (Brunner and Rechberger, 2015). Kuznetsova et al. (2019) discussed the importance of decision-support tools in waste-to-energy management systems, showing that 50% total operational expenses reduction can be achieved. For the purpose of simulating waste-to-energy systems using first principles, specialized software such as W2E (Touš et al., 2009) were developed to provide steady-state estimations. Stehlík (2009b) added that computation simulation, mathematical modelling, process integration and optimization studies provides excellence in waste-to-energy system design, operation, and management. Nevertheless, operational management for waste-to-energy systems provides a complex and stochastic (Touš et al., 2015a) system dynamics that is very difficult to be forecasted due to inherently varying properties in the waste feed, such as lower heating value (Touš et al., 2013). Later, Putna et al. (2018) used correlation coefficients to statistically estimate these fluctuation parameters in

waste-to-energy plant demand modelling. At this stage, waste-to-energy energy prediction mainly relies on stochastic programming (Šomplák et al., 2012), regression analysis (Younes et al., 2014), statistical and uncertainty model (S. Wang et al., 2016). Nevertheless, from author's first handed experience, most waste-to-energy plants are manually managed by human operators and more advanced energy management models should be deployed in the waste-to-energy industry to provide more accurate and robust operational management.

Looking into energy management tools for generic energy industries, the core enabler of the industrial energy management system is the implemented predictive and optimization algorithm which allows for efficient allocation of energy to sources. A classical work from Marchetti and Parigi (2002) compared regression models, such as combinations of linear, log-linear, exponential, autoregressive integrated moving average, to predict industrial trends in Italy. Moreover, Perrin and Sibly (2003) shown that dynamic optimization methods can be practically applied for energy management systems.

From many research works (Bao et al., 2018; Touš et al., 2015a), the two fundamental algorithmic components of industrial energy management are forecasting (Chen et al., 2011) and decision optimization (Yu and Hong, 2017). For forecasting analysis, there are three major directions which includes regression-based analysis (Marchetti and Parigi, 2002), Fourier analysis (Tsai et al., 2016) and neural networks (Srinivasan, 2008). Autoregressive Integrated Moving Average (ARIMA) is one of the more well-known and reliable regression-based methods for forecasting. Even though the method was invented in the 1970s, ARIMA-based methods are still popular and effective today and often used as the standard of benchmarking (Sen et al., 2016). Yuan et al. demonstrated that ARIMA (Yuan et al., 2016) is effective when being use to predict the energy consumption in China, achieving 4.62% of mean average percentage error (MAPE) on predictions. Discrete Fourier Transformation-based methods was also used for energy forecasting by Beiraghi and Ranjbar (2011). The work showed that Fourier-based methods are good at forecasting periodic data. A recent work from Oh and Son (Oh and Son, 2018) also utilized Discrete Fourier Transformation for reliable wind-power forecasting. Alternatively, neural networks are also popularly used for energy forecasting. Works of González-Romera et al. (2008) indicated that neural networks are better in trend predictions than Fourier series, vice-versa Fourier methods are better in periodic behaviour forecasting. The work studied a novel hybrid method with comparison to ARIMA obtaining results lower than 2% MAPE for electric energy demand forecasting. He et al. (2019) used a

Least Absolute Shrinkage and Selection Operator-Quantile Regression Neural Network (LASSO-QRNN) to simultaneously predict the energy consumption in different quantiles in the next several years.

Day-ahead energy predictions (with hourly resolution) is much more challenging than single time-step short-term predictions. For this, Lan et al. (2019) studied the effects of using backpropagation-based neural networks, Elman neural networks and ARIMA with a spatial-temporal approach. Due to the difficulty of day-ahead energy prediction, Sun et al. (Sun et al., 2019) proposed a Q-learning enhanced probabilistic model for the spatial-temporal wind power forecasting. Another state-of-the-art approach for this type of prediction is by using Long Short-Term Memory (LSTM) neural networks (Hochreiter and Schmidhuber, 1997). Recent work from Qing and Niu (2018) demonstrated that LSTM neural networks can perform better than persistence (human) for predicting solar energy, achieving 76.245 root mean squared error (RMSE). Researchers also made effort to improve the performance of LSTM neural networks for specific energy forecasting applications. For example, Y. Wang et al. (2019) used a pinball loss to improve the probabilistic prediction performance of the LSTM neural network for the task of consumer energy demands. Some researchers also forcefully combine multiple models (ensemble method), giving weighted outputs from each prediction individual model. An example can be seen in the works of Y. Liu et al. (2019) who used an ensemble of multiple models based on deep learning and variational Bayesian inference on forecasting of solar energy. More energy forecasting methods can be found in the review paper of Suganthi and Samuel (2012).

Methods for real-time optimization for industrial energy management are highly dependent on the case study for implementation. Ahmad and Khan (2019) demonstrated that an effective and practical real-time energy management system should not rely on existing models, prior knowledge and have low computation complexity. If the system can be effectively modelled, model predictive control (MPC) methods are good methods which can provide applicable optimization solutions (Bao et al., 2018). Nevertheless most systems in the real world are stochastic or grey-box (Máša et al., 2016), this makes model-based method very difficult to be implemented. A combination of predictive models is commonly used together with mathematical programming. For example, Xenos et al. (2015) used a combination of regression analysis, mix-integer linear programming (MILP) and non-linear programming (NLP) for the real-time optimization of compressor load distribution. Gradient-based or

approximated gradient methods are also popular for real-time optimization. Atthajariyakul and Leephakpreeda (2004) proved that gradient-based optimization methods were effective for the energy optimization of a 24-hour operating HVAC system. Gradient methods are effective (Marchetti et al., 2009), and they can be implemented in a model-free fashion. Probabilistic methods such as Monte Carlo optimization (Máša et al., 2016) can also be used for real-time optimization if the nature of the problem is highly stochastic. More detailed discussion on alternative methods for real-time optimization can be found in works of Chachuat et al. (2009). Some significant work related to both energy prediction and optimization for generic energy management models (Table 11.1).

Table 11.1. Energy prediction and optimization works for the generic energy industry

Research Work	Year	Contribution
Srinivasan (2008)	2003	Proposed the use of group method data handling (GMDH) neural network for energy demand predictions.
Chen et al. (2011)	2011	Proposed the use of a simple neural network for forecasting with matrix real-coded genetic algorithm for optimization.
Ding et al. (2014)	2014	Optimized smart energy network using a day-ahead approach and considered market selling/purchasing activity.
Touš et al. (2015a)	2015	Industrial implemented production planning strategy based-on Artificial Neural Network and Monte Carlo simulations
Máša et al. (2016)	2016	Data-driven grey box model which combines Artificial Neural Network and input-output models for Monte Carlo optimization
Yu et al. (2016)	2016	Real-time robust optimization of energy demand responses.
Tsai et al. (2016)	2016	Fourier series-based model for analysing and predicting energy consumptions.
Bao et al. (2018)	2018	Bi-level optimization with model predictive control (MPC) considering energy storage systems.

To author's knowledge, the simultaneous day-ahead prediction with hourly resolution and real-time optimization of waste-to-energy systems is rarely carried out. For example, Touš et al., (2015a) used an artificial neural network with Monte-Carlo simulation to carry out hour-ahead operational management for waste-to-energy systems. Šomplák et al. (2013) considered

uncertainty with stochastic optimization for waste-to-energy operational planning. Dai et al. (2011) used a two-stage support-vector regression and interval-parameter mixed integer linear programming (IMILP) for the yearly prediction of waste and optimization of waste-to-energy supply chain in Beijing, China. Furthermore, Namuli et al. (2013) used dynamic simulation and a metaheuristic algorithm (Tabu search) for the optimization of biomass-based waste-to-energy systems. These work have provided useful development for the operational management of waste-to-energy systems, however the requirement of a management system with accurate day-ahead prediction (with hourly resolution) and real-time optimization for a waste-to-energy plant is still not available. This is prominent due to municipal waste feed containing difficult-to-predict characteristics (Touš et al., 2013), noises in real-time operational data, and temporally-dependant incentive scheme (Touš et al., 2015b). Evidently, there is much challenge in such waste-to-energy management system such as the requirement of: (i) 24 time-steps-ahead prediction (for day-ahead with hourly resolution) rather than 1 time-step ahead, (ii) dealing with anomaly and knowing when to trust the prediction, and (iii) targeting the economics of the waste-to-energy plant.

In this work, a novel smart waste-to-energy framework is proposed with HTM acting as the core day-ahead (with hourly resolution) prediction and anomaly detection algorithm (see Chapter 11.2.1 and 11.2.2). HTM uses a scalar prediction error of cells in the latent space to construct the anomaly score, instead of residual error which makes it inherently able to predict anomaly ahead of time. Furthermore, a dual-mode optimization algorithm which targets the system economics is used to switch from predictive optimization to robust optimization upon suspecting anomaly. To elucidate the effectiveness of this framework, a waste-to-energy processing plant in Czech Republic is used as the case study. It is demonstrated in the following chapters that the robustness of the framework and its improvement in energy efficiency is improved.

11.2 Methodologies and Algorithms

In this chapter, the framework of the proposed HTM-based energy management system is described. The formulation of the HTM algorithm, other comparative methods and aspects of the energy management system will be also discussed.

11.2.1 Industrial Energy Management Framework

The energy management system depends on the HTM algorithm to learn the datastreams of energy demand data in an on-line setting. The HTM algorithm detects the anomalism of the energy demand and decides the mode of optimization algorithm to be used. Two modes of

optimization are proposed in this framework, being real-time predictive and real-time robust optimization. Only one of the optimization algorithms will run at a single time. Robust optimization has objectives to reduce the uncertainties of the energy system while maximizing energy efficiency. On the other hand, predictive optimization maximizes energy efficiency based on reliable predictive data. Both optimization systems have their objectives instantiated by an economic model which is suitable to the energy system. Results from the optimization will be input to the demand response system to declare energy production from energy source in the future. Energy source is transferred to energy sinks, which generates the energy demand data stream. The flow diagram of this energy management framework is shown in Figure 11.1. The novelty of this framework is that HTM is being used as a predictive algorithm with inherent latent anomaly prediction while coupled with an automatic switching predictive and robust optimization procedure (dual-mode optimization).

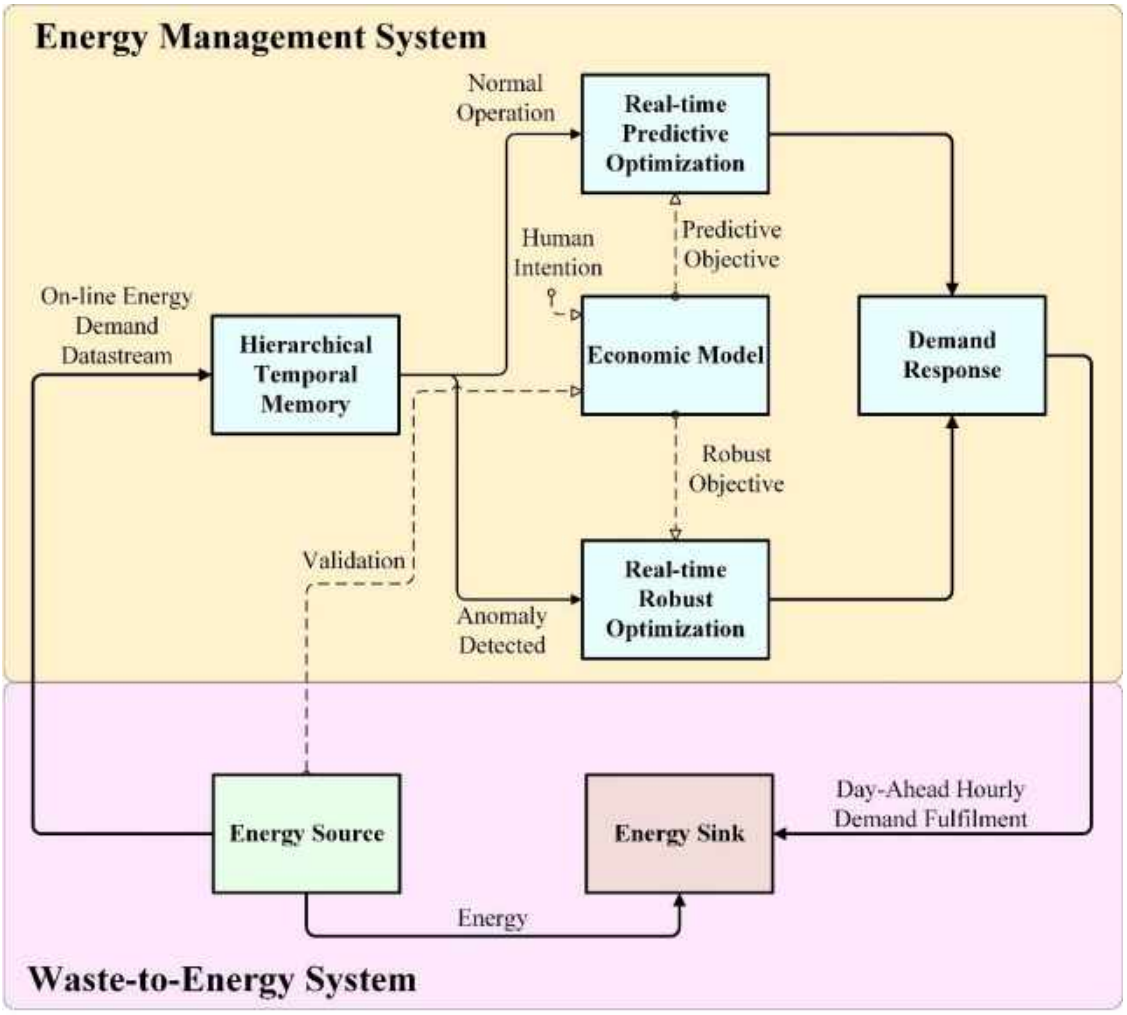


Figure 11.1. Proposed novel energy management system framework with HTM and dual-mode optimization for waste-to-energy systems

11.2.2 Hierarchical Temporal Memory (HTM)

This work proposed the use of Hierarchical Temporal Memory (HTM) for the application of energy allocation problems. HTM is a novel machine learning technique that was formulated based on the Cortical Learning Algorithm (CLA). HTM can be considered as a special form of Bayesian neural networks, as it consists of node collections arranged in a tree-shaped structure while using the Bayesian-belief propagation mechanism (Chen et al., 2012). The method was first proposed by Hawkins and Sandra (2007) which was inspired by the architecture of the neocortex of the brain (used for sensory perception, cognition, language and human reasoning). There are many implementations of HTM and the currently most developed one is from Numenta, which was founded by Jeff Hawkins (2006). The first generation of HTM was focused on the fundamental concepts and terminologies of HTM (Chen et al., 2012). The work from George (2009) discusses the concepts of spatio-temporal learning and the replication of the neocortex behaviours using HTM. The HTM model was later improved using sparse distributed representations as shown by Ahmad and Hawkins (2015). In this second generation of HTM algorithm, the architecture of the neocortex is further embedded into the HTM model by extending nodes in each level to a column (Chen et al., 2012; Hawkins et al., 2017). Since then, there have been many incremental improvements to the HTM algorithm such as Rozado et al. (2010), which optimized the HTM algorithm for multi-variable time series. Kostavelis and Gasteratos (2012) improved the HTM algorithm to provide more accurate performance and faster training speed. The spatial pooler of HTM has also been mathematically formalized in works of Mnatzaganian et al. (2016). HTM has also been shown to provide robust predictions with a considerably small computation complexity in real-world applications by Melis et al. (2009). Recent work from Ahmad et al. (2017) shown that HTM can also be simultaneously used for anomaly detection while performing real-time predictions.

The main advantage of using a HTM model for the purpose of energy prediction are: (i) Early anomaly detection within underlying systems while providing online predictions (Ahmad et al., 2017) (ii) On-line learning instead of batch training (Hawkins, 2006) (iii) Unsupervised spatio-temporal learning mode (Mnatzaganian et al., 2016) (iv) Robust to variations, changes and noises (Hawkins and Ahmad, 2016). In addition, on-line learning of HTM provides an adaptive learning feature gives great importance for waste-to-energy plants due to constant changes in the heating values of input wastes. Moreover, HTM uses a neocortex-inspired neuron model instead of the conventional McCulloch-Pitts neuron model (See Figure 11.2). As the purpose of artificial intelligence is to mimic nature's intelligence, Figure 1 demonstrates the fundamental inspiration

of HTM which originates from the biological neuron with high conceptual similarity. Such characteristics of the HTM algorithm makes it a suitable and elegant method to be implemented in real-world waste-to-energy management systems.

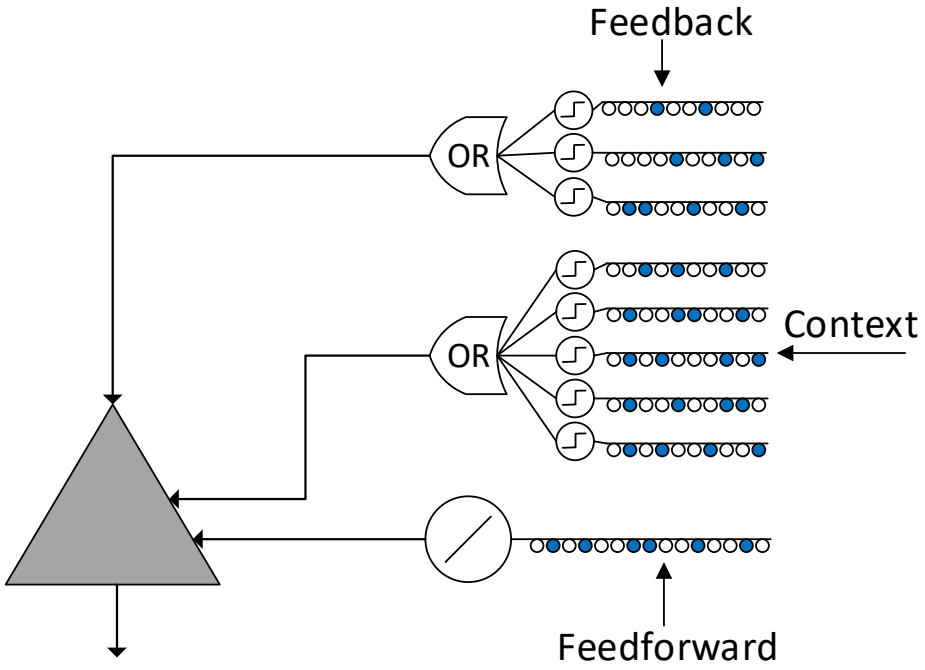


Figure 11.2. Structure of HTM model neuron. (Adapted from (Hawkins and Ahmad, 2016))

The HTM is a biological constraint model which consists of interconnected cells in a levelled hierarchical framework. The cells of a same hierarchical level are arranged in columns, which allows HTM to learn the spatio-temporal correlational properties of its inputs. Hawkins and Ahmad (2016) mentions that the cells in an HTM region are conceptually equivalent to the biological neurons in the layer 3 of the neocortex region, which is one of the primary feed-forward layers of neurons. The size of column and each layer in the HTM is not manually coded, but it is optimized and learned (see Fig. 11.3).

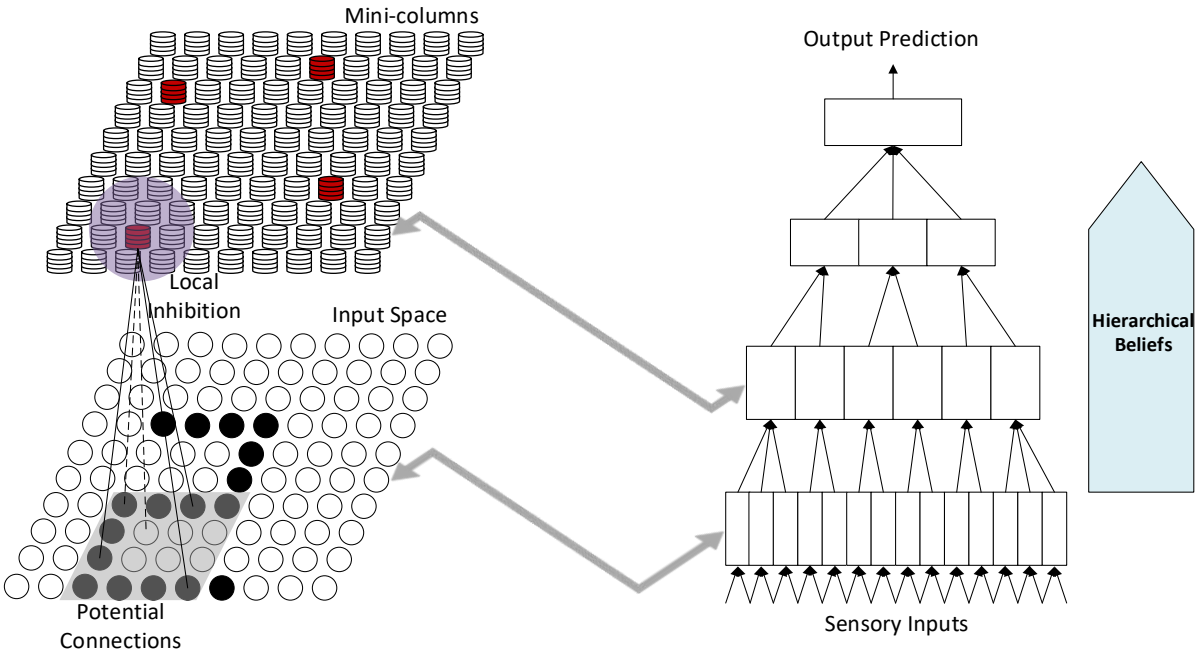


Figure 11.3. Structure of Hierarchical Temporal Memory (HTM) (Modified from Cui et al. (2017))

For the case of this work, the NuPic (Numenta Platform for Intelligent Computing) implementation of HTM was used, which applied a particle swarm optimization-based method for the learning of such structures and network properties. This method is also known as swarming. In HTM structure, there are 3 types of considered variables, which are field selection, scalar parameter values, and enumerate parameter values. The swarming algorithm treats these variables differently. Concisely, the swarming algorithm uses a greedy-based field search logic as an outer optimization loop, while other variable types are searched in more "mini-swarms". Scalar parameter values are optimized using the modified particle swarm optimization algorithm (PSO) by Shi and Eberhart (Shi and Eberhart, 2002) which its equations are shown as following:

$$v_{i+1} = v_i + C_1 \times rnd(0,1) \times (p_i - x_i) + C_2 \times rnd(0,1) \times (g_{best} - x_i)$$
$$x_{i+1} = x_i + v_i \tag{11.1}$$

From Equation 11.1, v_i is the velocity of particle at step i, x_i is the position of particle at step i, p_i is the local best position, g_{best} is the global best position, $rnd(0,1)$ is a random generator function for values between 0 and 1, C_1 and C_2 are both constant and 2.0 is used for both values. On the other hand, enumerated parameters are initialized in swarms based on their probability of average score errors. This gives higher probability of search in better-performing enumerated parameters.

In this work, three categories of encoders are being used, which are scalar encoders, date encoders and spatial encoder. These encoders convert variables to binary representations in the structure of HTM. Layers are connected using separate weighted sparse connectors for spatial dimension and temporal dimension, which are referred as spatial poolers and temporal poolers. The output layer is a classification layer, which uses bucketing to convert binary encoded data to rounded continuous variables. HTM uses a Hebbian-based learning which consists of competitive learning rules and homeostatic excitability control (Cui et al., 2017). The formalization of the learning rules in HTM has been originally proposed in the works of Hawkins and Ahmad (2016). The learning process can be concluded in three steps:

1) Initialization: Each segment within the network is initialized with some potential synapses which connects to randomly selected cells within the layer. The potential synapses have permanence values which are assigned randomly.
2) Computing cell states: From the inhibitory process, a set of winning columns denoted as W^t will be determined by matching the current feed-forward input pattern. The active state for each cell in the columns is computed as shown in Eqn. 11.2.

$$a_{ij}^t = \begin{cases} 1 \quad if \ \ j \in W^t \quad and \quad \pi_{ij}^{t-1} = 1 \\ 1 \quad if \quad j \in W^t \quad and \quad \sum_i \pi_{ij}^{t-1} = 0 \\ 0 \quad otherwise \end{cases} \tag{11.2}$$

where a^t_{ij} refers to the activity of the i^{th} cell in j^{th} column with an active and non-active state of 1 and 0 respectively, π^{t-1}_{ij} refers to the predictive activity of the i^{th} cell in j^{th} column with predictive and non-predictive state of 1 and 0 respectively. Subsequently, the predicted step for the current time step, t will be computed using Eqn. 11.3.

$$\pi_{ij}^t = \begin{cases} 1 \quad if \quad \exists_d \left\| \tilde{D}_{ij}^d \circ A^t \right\|_1 > \theta \\ 0 \quad otherwise \end{cases} \tag{11.3}$$

where θ represents the threshold for segment activation, $\circ$ is the operator for element-wise multiplication. $\tilde{D}_{ij}^d$ represents the connected synapses of d^{th} segment of i^{th} cell in j^{th} column, this is a binary value which is 1 when connected and 0 otherwise. A^t is a binary matrix which represents the set of active cells at time t, in other words, $A^t=\{a^t_{ij}\}$. If there are more than the threshold value of connected synapses with active presynaptic

cells, then the segment will be activated. If at least one segment is active, the cell will be depolarized.

(3) Updating segments and synapses: The segment selection criteria for update procedure is as described in Eqn. 11.4.

$$\forall_{j \in W^t} (\pi_{ij}^{t-1} > 0) \quad and \quad \left\| \tilde{D}_{ij}^{d} \circ A^t \right\|_1 > \theta \tag{11.4}$$

The equation selects only columns that both were active during previous time step and provided correct prediction. If the winning column gave incorrect predictions, the cell will be selected to give the closest context if the current sequence transition repeats. This imply that the segment that has the most synapses with active cells is selected and reinforced even if it is below threshold. This is formulated in Eqn. 11.5.

$$\forall_{j \in W^t} (\pi_{ij}^{t-1} > 0) \quad and \quad \left\| \dot{D}_{ij}^{d} \circ A^t \right\|_1 = \max_i \left(\left\| \dot{D}_{ij}^{d} \circ A^t \right\|_1 \right) \tag{11.5}$$

where $\dot{D}_{ij}^{d}$ represents the binary matrix containing only the positive entries in distal segments, D_{ij}^{d}. To reinforce the segments, synapse with active cells are rewarded while synapse with inactive cells are punished. This is implemented using a small decrease in penalty p^- and a large permanence value reward p^+ to update the permanence value of synapses within active cells:

$$\Delta D_{ij}^{d} = p^+ (D_{ij}^{d} \circ A^{t-1}) - p^- \dot{D}_{ij}^{d} \tag{11.6}$$

For cells that are currently inactive, a very small decay is implemented to trigger their activation. Biologically, this happens if segments were mistakenly reinforced by chance. This phenomenon is described in Eqn. 11.7.

$$\Delta D_{ij}^{d} = p^{--} \dot{D}_{ij}^{d} \quad where \quad a_{ij}^{t} = 0 \quad and \quad \left\| \tilde{D}_{ij}^{d} \circ A^{t-1} \right\|_1 > \theta \tag{11.7}$$

As the update step, matrices ΔD^{d}_{ij} are added to the current matrices of permanence values for each time step. The overall procedure of the learning process is illustrated in Figure 4 for better visual understanding.

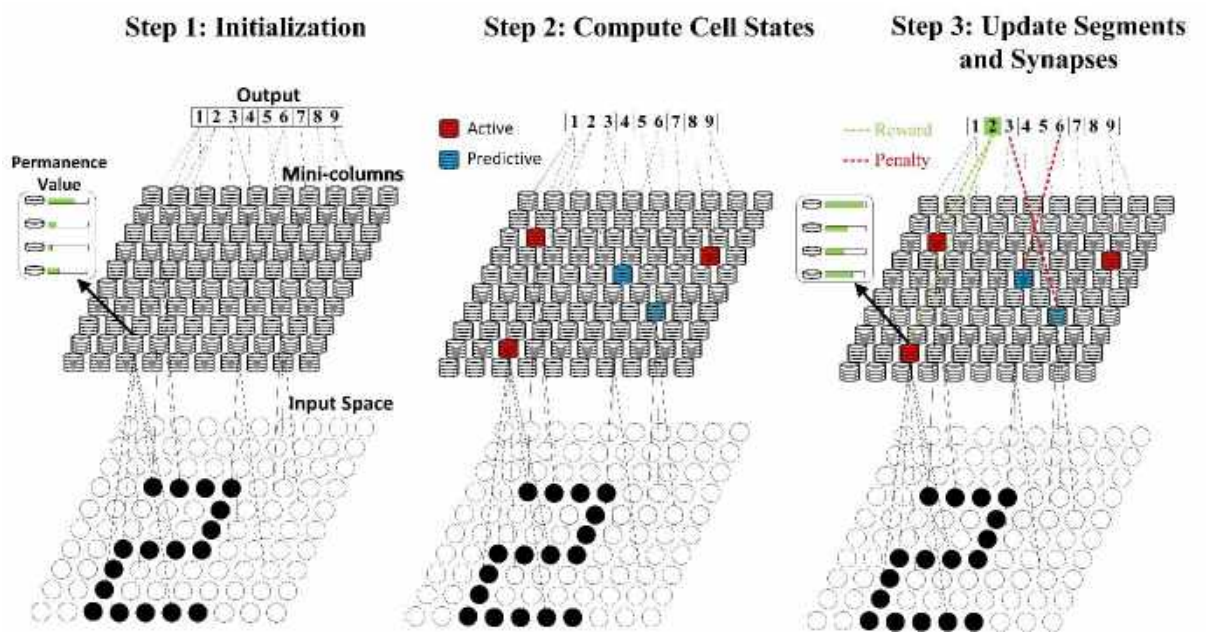


Figure 11.4: Illustration of the training and inference procedure of HTM

One of the main success of the HTM terminology is the use of Sparse Distributed Representations (SDR). SDR is consistent with the biological cortical representation in which many neurons from the same region of activated neurons remain relatively inactively, this biological activity is called as *sparse*. From the works of Ahmad and Hawkins (2015), this deceptively simple encoding method is effective for the application of HTM as it serves as a biological constraint for the model architecture when compared to dense representations. The number of unique SDR encodings for an SDR with *n* length and *w* active cells can be calculated using the choose function (Ahmad and Hawkins, 2015):

$$\binom{n}{w} = \frac{n!}{w!\,(n-w)!} \tag{11.8}$$

As a comparison, an SDR with length of 40 and 4 active cell has 93,390 possible encoding, while dense representation with length of 40 has 1.1 x 10^{12} possible encoding. This simple example of SDR reduces the search space by 1.2 x 10^{7} folds.

For anomaly detection, HTM uses a scalar value for the prediction error of each of the cell, s_t to give an instantaneous measure of how well the current HTM model predicts the input. This is computed using the following equation as described by Ahmad et al. (2017):

$$s_t = 1 - \frac{\pi(x_{t-1}) \cdot a(x_t)}{\left|a(x_t)\right|} \tag{11.8}$$

where $a(x_t)$ is the sparse encoding of the current input, x_t, while $\pi(x_{t-1})$ is the sparse vector representation in the internal workings of the HTM. $|a(x_t)|$ is the scalar norm and is equivalent to the number of 1 bits in $a(x_t)$. To highlight, the anomaly prediction is carried out in the latent space (internal working) of the HTM.

The prediction error is modelled to follow a rolling normal distribution, where the mean, μ_t and variance, σ_t^2 are updated from previous errors in the fixed rolling window, *W*. For the short-term prediction, a separate short-term window, *W'* is also used to compute the short-term moving average, $\tilde{\mu}_t$. This is shown in Eqn. 11.9-11.11.

$$\mu_t = \frac{\sum_{i=0}^{i=W-1} s_{t-i}}{W} \tag{11.9}$$

$$\sigma_t^2 = \frac{\sum_{i=0}^{i=W-1} (s_{t-i} - \mu_t)^2}{W-1} \tag{11.10}$$

$$\tilde{\mu}_t = \frac{\sum_{i=0}^{i=W-1} s_{t-i}}{W'} \tag{11.11}$$

The key criteria to declare an anomaly is the anomaly likelihood, L_t. It can be computed using a Gaussian tail probability, $Q(x)$. When the anomaly likelihood is larger than the threshold, $1-\varepsilon$, the mode of the energy system switches to RO. In this work, ε is set to 0.1 (see Eqn. 11.12).

$$L_t = 1 - Q\left(\frac{\tilde{\mu}_t - \mu_t}{\sigma_t}\right)$$
$$anomaly \quad declared : L_t \geq 1 - \varepsilon \tag{11.12}$$

11.2.3 Comparative Methods

In this work, 5 other methods will be used as comparison with the predictions of HTM, which includes Long-Short-Term Memory (LSTM) neural networks, autoregressive integrated moving average (ARIMA), Fourier Transformation Extrapolation, persistence, and comparison with human operator. The first three methods represent the most popular methods from the respective field of neural networks, regression analysis and Fourier transformation, while the latter two represent practical and commonly used methods in a waste-to-energy facility.

11.2.3.1 Long Short-Term Memory (LSTM) Neural Network

The LSTM neural network was first invented by Hochreiter and Schmidhuber (1997). The LSTM unit consists of an input gate, output gate and a forget gate. The popularity of the LSTM unit was due to the unit being able to be mathematically differentiated while simultaneously dealing with exploding and vanishing gradient problems. Furthermore, LSTM unit is elegant in such a way that it allows the neural network to "forget" some less important data such as the human brain forgets information (Gers et al., 2000). The mathematical formulation of a LSTM unit with a forget gate is shown as the below (Gers et al., 2000; Hochreiter and Schmidhuber, 1997):

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \tag{11.13}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{11.14}$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \tag{11.15}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \tag{11.16}$$

$$h_t = o_g \circ \sigma_h(c_t) \tag{11.17}$$

Where x_t represents the input vector to the LSTM unit, h_t is the output vector to the unit, c_t is the cell state vector, f_t is the forget gate's activation vector, i_t is the input gate's activation vector, o_t is the output gate's activation vector, σ is the sigmoid function. W and U are weight matrices while b is bias matrix, all to be learned during training. For illustration for LSTM neural network can be found in Figure 11.5. The gradient descent-based Adam optimizer (Kingma and Ba, 2015) was used for weight updates of the LSTM neural network with learning rate of 0.001, B_1 parameter of 0.9 and B_2 parameter of 0.999. For the optimization of hyperparameters (ie. cells in layers, number of layers), particle swarm optimization (PSO) (Shi and Eberhart, 2002) is used for fair comparison (see Eqn. 11.1). The PSO algorithm was set to optimize the validation loss (MSE) with a population size of 50 and stopping criteria of no changes on objective within 10 iterations.

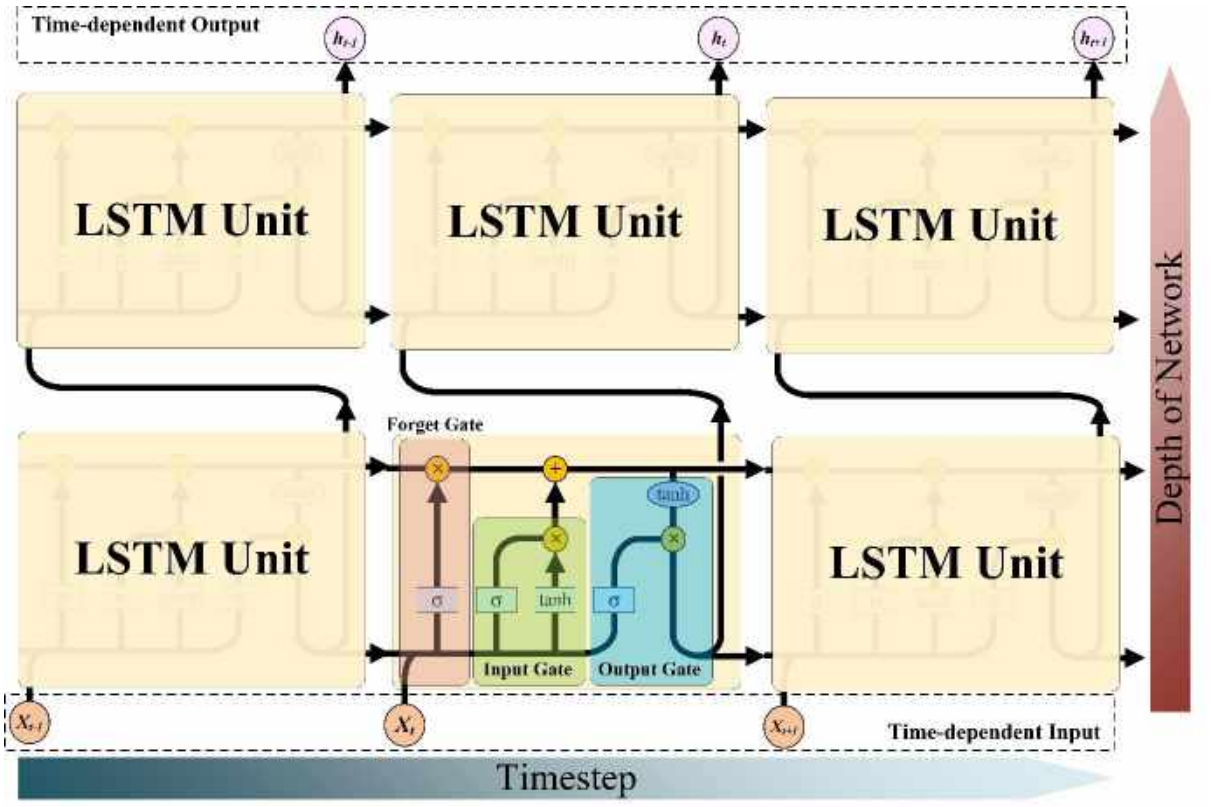


Figure 11.5: The architecture of a LSTM neural network

11.2.3.2 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA was first popularized by Box and Jenkins (Geurts et al., 1977) to obtain the best-fit of a time series model by using the past values of that series. The term "AR" (auto-regression) from ARIMA often corresponds to an order, *p* which indicates the autocorrection of weighted moving average over past observation. The term "I" (integration) from ARIMA corresponds to an order, *d* which is the order of differencing. This parameter controls the type of the trend, such as linear or polynomial. Lastly, the term "MA" (moving average) from ARIMA corresponds to an order, *q* which considers the weighted moving average over past errors. Therefore, the function for this method is commonly represented as ARIMA(p,d,q). In this study, the orders (p,d,q) are optimized by a similar PSO algorithm as used for the LSTM neural network. The general mathematical formulation for this method can be found in Equation 11.18.

$$(1-\sum_{k=1}^{p}\alpha_k L^k)(1-L)^d X_t = (1+\sum_{k=1}^{q} B_k L^k)\varepsilon_t \tag{11.18}$$

In the equation, α is the auto-regressive coefficients, B is the moving average coefficients, ε_t is the error term, X_t is the predicted variable at time t, L is an operator which obtains the previous time-step of the variable such that $L \cdot X_t = X_{t-1}$.

11.2.3.3 Fourier Transformation Extrapolation (FTE)

Fourier transformation extrapolation (FTE) is a method that studies the periodic behaviours of a time-series using the famous discrete Fourier transform. This famous method is used in many forecasting applications, including energy forecasting (Beiraghi and Ranjbar, 2011). A Fast Fourier Transformation (Cooley and Tukey, 1965) algorithm is used to accelerate computation. In this work, the time-series data was first detrended using a simple first-order method (see Eqn. 11.19), the first-order trend will be recompensated after the inverse Fourier transform.

$$x'_t = x_t - p_0 \Delta t \tag{11.19}$$

where x' is the detrended variable, x is the variable, p_0 is the first-order trend coefficient, t represents the time-step. The discrete Fourier transform is represented in Eqn. 11.20.

$$X_k = \sum_{n=0}^{N-1} x'_t \, e^{-\frac{i2\pi}{N}kn} = \sum_{n=0}^{N-1} x'_t \left[\cos\left(\frac{2\pi}{N}kn\right) - i \cdot \sin\left(\frac{2\pi}{N}kn\right)\right] \tag{11.20}$$

where X_k is the Fourier transformed variable, k is the phase angle, n is a whole number, 1/N is the frequency. Since N is the only hyperparameter, a brute force search was used for hyperparameter optimization, and the global optimality of this hyperparameter can be confirmed. Subsequently, the inverse transformation can be obtained (see Eqn. 11.21) for an extrapolation in phase, $\Delta\theta$.

$$x'_{t+1} = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cdot e^{\frac{i2\pi kn + \Delta\theta}{N}} \tag{11.21}$$

Finally, the first-order trend is recompensated using Eqn. 11.22:

$$x_t = x'_{t+1} + p_0 \Delta t \tag{11.22}$$

11.2.3.4 Persistence and Direct Comparison with Human Operator

A practical method for day-ahead forecasting is by using persistence forecasting. Although simple, this method has provided statistical evidence to give a logical and effective forecasting result. This approach has been studied in many energy management research (Abdel-Aal et al., 2009; Lan et al., 2019) and is commonly used as a benchmark for practical forecasting. The persistence forecasting method for day-ahead prediction can be expressed as following, where $x(t)$ is the predicted variable in time t, and $x(t+24|\ t)$ is the predicted variable in day-ahead given the information of time t.

$$x(t + 24 \,|\, t) = x(t) \tag{11.23}$$

Another practical method for the benchmark of waste-to-energy management is by direct comparison with existing human operator. This is the most robust benchmark due to this

forecast being made by a waste-to-energy operational expert, and directly impact the economics of the company. Therefore, the operator commonly has years of operational experience, and can provide profitable forecast from the perspective of an expert human in the task.

11.2.4 Dual Mode Optimization

A dual-mode optimization algorithm which consists of two types of optimization formulation (predictive and robust optimization) are used for this type of problem. Under normal circumstances (when the HTM does not detect anomaly), the predictive optimization (PO) mode is applied. For comparative method, only PO was used due to these methods not being able to provide anomaly ahead of time. The formulation of predictive two-step optimization is shown in Eqn. 11.24 and 11.25. An approximated gradient method was proposed due to two characteristics of the realistic waste-to-energy management problem: (i) the problem being a stochastic black-box and (ii) impossible to perform multiple sampling from the real-world system for optimization. The first characteristic makes exact algorithms not suitable for this problem, while the second characteristic make optimization procedure that require multiple sampling (such as metaheuristics, heuristics, and direct search method) infeasible. Eqn. 11.24 is the outer optimization problem to obtain an optimized α value which can maximize the prior accumulated economic potential of the energy system. Eqn. 11.25 describes that the maximal expectation of economic potential by assigning an adaptive correction factor ε with respect to the prediction algorithm's 24-hour-ahead prediction for the declared energy production.

$$\begin{aligned}
&\max \quad \sum_{i=1}^{t} \frac{G_i}{P_i} \\
&s.t. \quad 0 \le \alpha_{t+1} \le 1 \\
&\quad P^s{}_i = P^*{}_i + \varepsilon_i \qquad \forall i \in 1,2,3,...,t \\
&\quad L_L \cdot P^s{}_i \le P_i \le L_H \cdot P^s{}_i \qquad \forall i \in 1,2,3,...,t \\
&\quad \varepsilon_i = \varepsilon_{i-1} + \alpha_{i-1} \frac{\Delta f}{\varepsilon_{i-1}} \qquad \forall i \in 1,2,3,...,t
\end{aligned} \tag{11.24}$$

$$\begin{aligned}
&\max \quad f = \frac{G_{t+24}}{P_{t+24}} \\
&s.t. \quad P^s{}_{t+24} = P^*{}_{t+24} + \varepsilon_t \\
&\quad P^s{}_{t+24} - v \le P_{t+24} \le P^s{}_{t+24} + v
\end{aligned} \tag{11.25}$$

Where G is the economic net gross profit (Euros), P is the energy production, P^s is the declared energy production (MWh) that was set 24 hours prior to actual production, P^* is the predicted energy production (MWh), ε is the adaptive correction factor (MWh), α is the update coefficient, v is the allowed variability for the energy production (MWh). Initial values for ε is 0 and α *is 0.05*. In this case, the prediction of prediction algorithm was taken as the ground-truth for optimization. Due to the employment of a realistic modeless approach, the exact gradients from the waste-to-energy system (black-box) is unknown. Therefore, a first order approximation of the gradient is used on the adaptive correction factor to fine-tune the prediction with respect to economic factor, formulated in a two-step optimization problem.

Robust optimization (RO) is carried out instead of predictive optimization for when HTM detects an anomaly. With this approach, the anomaly points can be targeted and optimized using a custom algorithm. This work uses a straightforward and effective RO method by applying the central limit theorem and assuming Gaussian anomalies. For this case, the expected value of the process economics is maximized by shifting the declared energy limits. The formulation of the RO can be found below:

$$
\begin{aligned}
\max \quad & E(f) \\
s.t. \quad & E(f) = \int_{P^s{}_{low}}^{P^s{}_{high}} f(P^s)[P^s + N(\bar{\delta}, \sigma^2)] \cdot dP^s \\
& P^s{}_{low} = \overline{P^s} - z(p)\frac{\sigma_{P^s}}{\sqrt{n}} \\
& P^s{}_{high} = \overline{P^s} + z(p)\frac{\sigma_{P^s}}{\sqrt{n}} \\
& \delta = P_a - P^s{}_a \sim N(\bar{\delta}, \sigma_\delta^2) \\
& P^s \sim N(\overline{P^s}, \sigma_{P^s}^2)
\end{aligned}
\tag{11.26}
$$

where *E(f)* is the overall expected value for the economic potential (Euros), $P^s{}_{high}$ and $P^s{}_{low}$ are the limits for the declared energy production (MWh), p is the Gaussian distribution probability, z is standard score function for normal distribution, the N is Gaussian distribution, δ is the difference between actual and declared anomaly points (MWh). The dual mode optimization operates with synchronous to the prediction algorithm (i.e. HTM) with the use of the anomaly binary to automatically trigger switching between modes. The overall concept of the interaction between HTM and the dual-mode optimization strategy can be found in Figure 11.6.

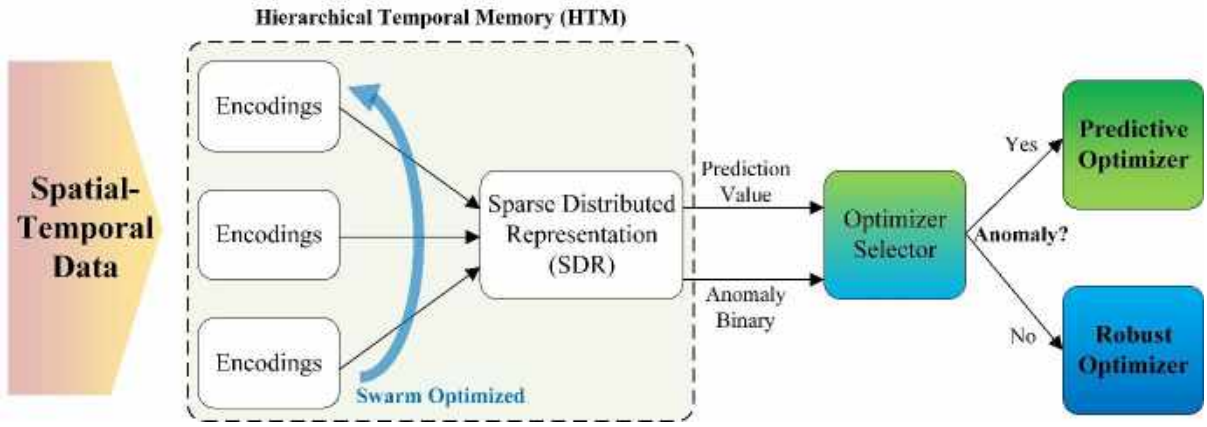


Figure 11.6: Flow diagram illustrating interaction between HTM and optimization strategy

11.3 Case Study

An industrial case study (Company A) from Czech Republic is used for demonstration. The case study is a waste-to-energy plant that collects municipal wastes for combustion to generate superheated steam. For this case study, electricity produced by the steam turbine is sold to the electrical grid. Although heat is recovered from the system, the processing plant has surplus heat, and therefore, it is not the interest of prediction and optimization. The process flow diagram of the waste-to-energy plant is illustrated in Figure 11.7.

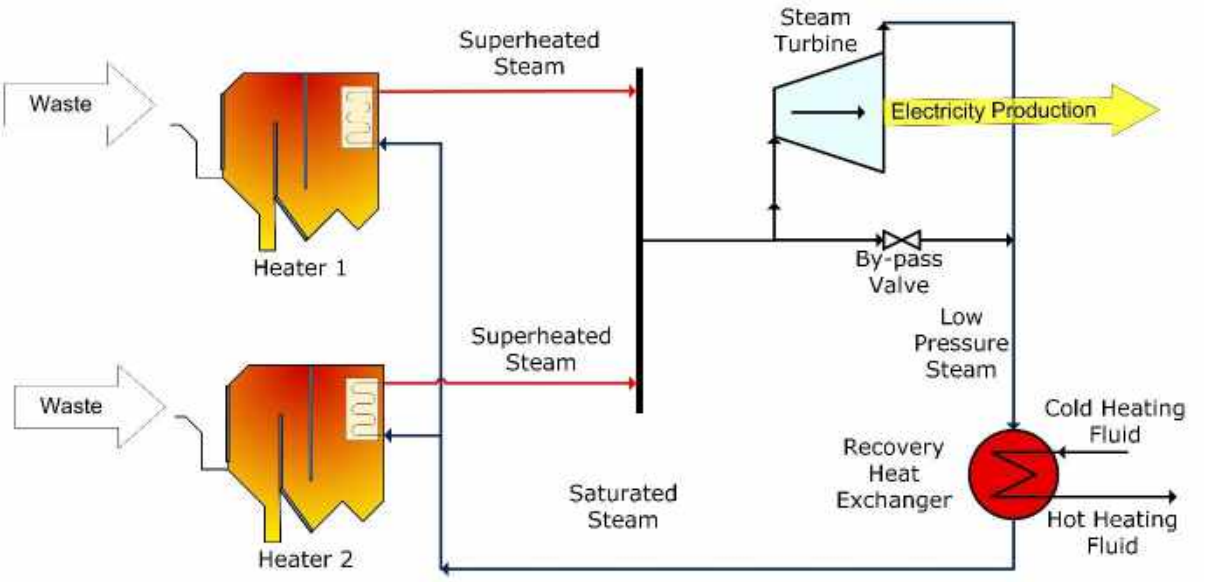


Figure 11.7. Process flow diagram of Waste-to-Energy processing plant

The processing plant can declare its energy production accurate to an hourly interval, however, this can only be done 24 hours pre-emptively. This also indicates that this energy predictions will be day-ahead with hourly resolution. Economically, if the actual electricity production is lesser than the declared energy production, large penalties will occur by contract. On the other

hand, if the actual electricity is more than the declared limits, electricity production must be decreased by by-passing steam, which valuable electric power is lost (See Table 11.2 for details).

Table 11.2. Case study specific economic parameters for evaluation

Parameter	Value	Unit
Electric tariff to grid	1.50	Euro/ MWh
Penalty for low energy production	200	Euro/h
Production allowed variability	0.5	MWh
Estimated loss of revenue due to steam by-pass	0.2	Euro/ MWh

Source: Rounded averaged values from case study

Under these circumstances, the prediction of electricity production becomes critical. However, waste characteristics are commonly not measured in a waste-to-energy plant, as it is often less hygiene for workers. On top of this, even if waste characterisation is possible, there are several critical problems such as low data frequency and expensive analytical equipment. Due to this matter, variation of heating values in waste causes large deviation in superheated steam generation, which leads to inconsistent electric production. Furthermore, there are frequent operational anomalies, such as shutdown and "tripping" of heaters, which further elevates the difficulty for prediction. This situation makes an input-output modelling strategy unsuccessful. Nevertheless, this work uses a novel approach of giving up on input-output correlations and concentrate the predictive model on dealing with spatial-temporal predictions (see Fig. 11.8(a)). This reduces the complexity of the problem in dealing with data acquisition and intermediate uncertainties while maintaining high prediction accuracy with awareness on anomaly. The dual-mode optimization is also able to switch modes when anomaly is detected, performing more conservative decisions. Figure 11.8(b) shows the economic factors for the optimization, giving a high penalization cost for under-generation of energy, while having slight profit loss for over-generation (Refer to Table 11.1).

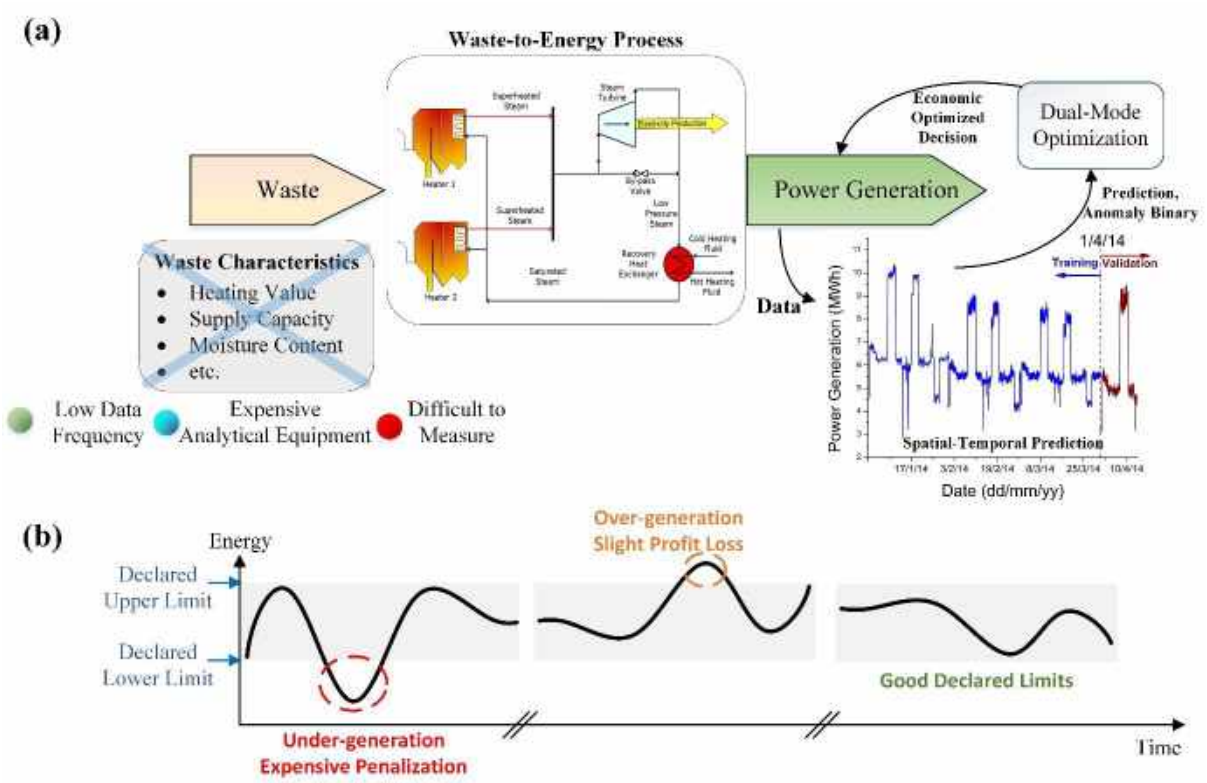


Figure 11.8: Waste-to-energy case study with (a) ideology of the implemented management system (b) demonstration of penalization and profit loss in energy response.

11.4 Results and Discussion

From the industrial case study, electric power production data of 6 months was used. Although HTM learns the data using online prediction, the swarming of the HTM structure must be applied on a restricted dataset and validated on full dataset. Hence, these data were split into the first 3 months and last months for training and prediction simulation. Figure 11.9 shows a snapshot illustration of the HTM structure and action of cells during online prediction.

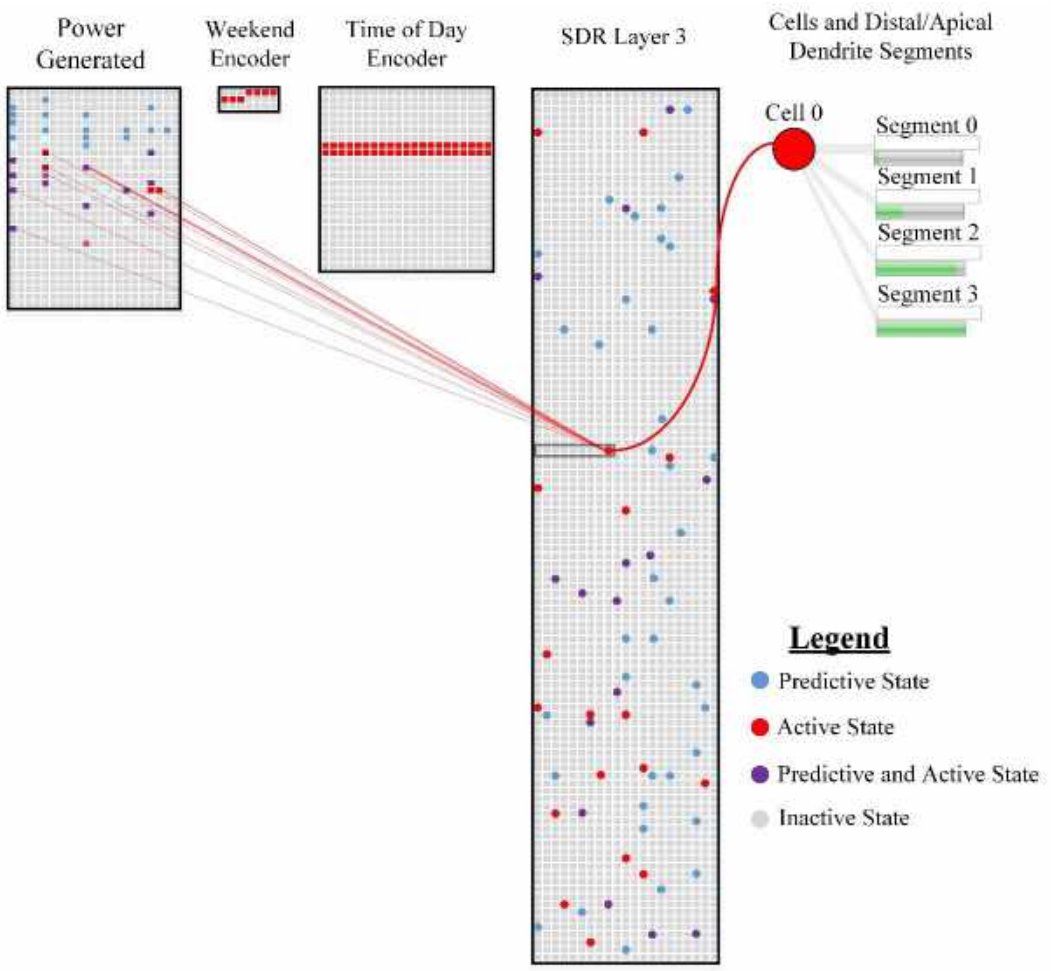


Figure 11.9. Action of cells during online prediction of Hierarchical Temporal Memory (HTM)

The prediction results of the HTM has shown an overall average mean squared error (MSE) of 0.8466 %. The predicted and actual turbine electricity generation is very similar except for the first few days, where there is a start-up error (see Fig. 11.10(a)). The graph highlighted with yellow shows the detection of anomaly based on the anomaly likelihood determined by HTM (Fig. 11.10(b)). The anomaly threshold for its likelihood is set at the recommended likelihood of 0.9. The HTM is also able to predict data 24 hour in advance with Mean Absolute Error (MAE) of 1.2429 %, which relates to R-squared of 0.9982 (See Fig. 11.11)).

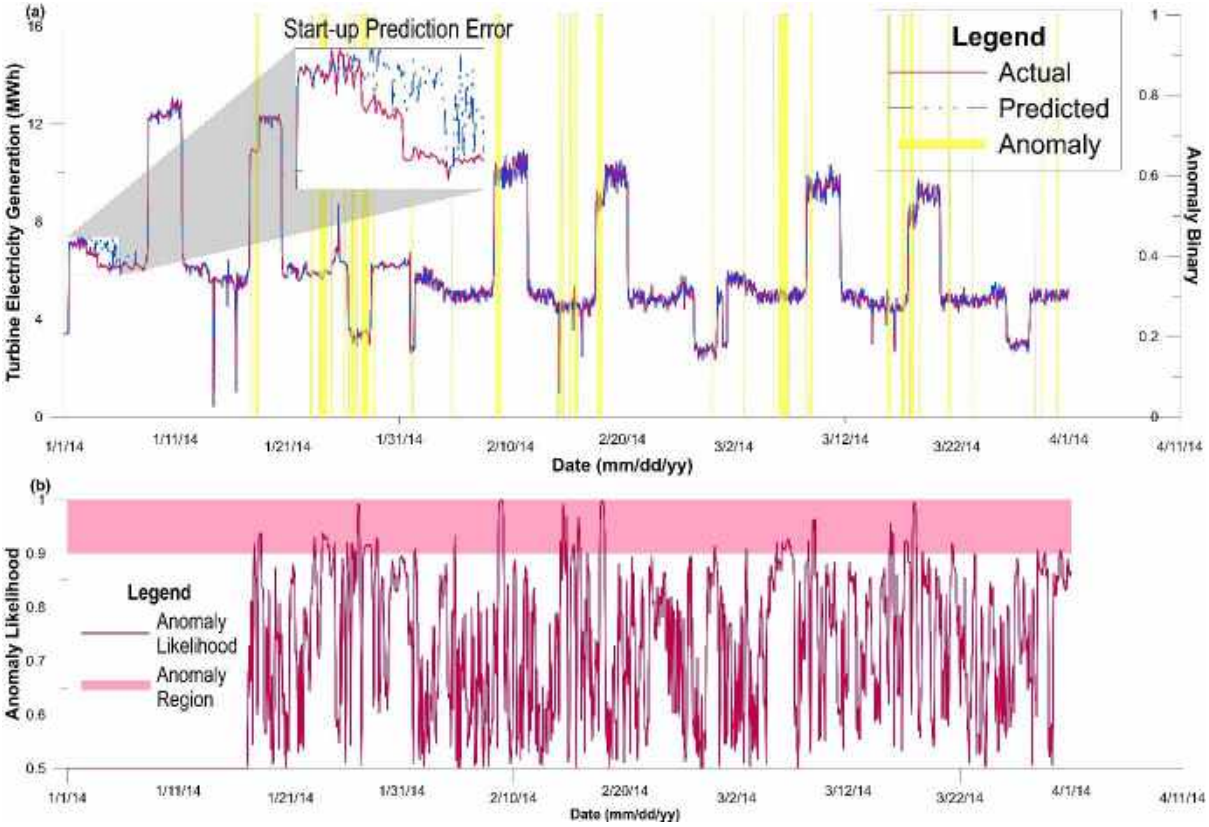


Figure 11.10. Prediction results of the HTM for first three months: (a) prediction of energy (b) anomaly likelihood

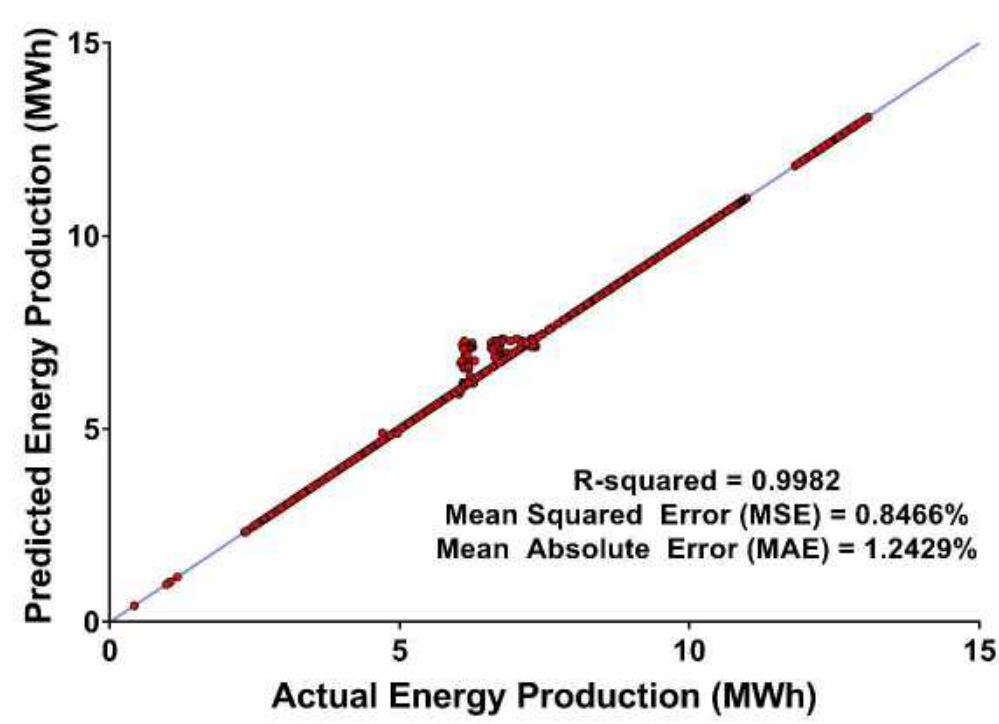


Figure 11.11. Predicted against actual energy production (MWh)

As comparison, three other types of well-known methods were compared during validation with HTM. This work chooses established methods from the fields of Fourier analysis, regression analysis, and recurrent neural networks to act as a benchmark for the results of HTM. The results of the prediction of each method are shown in Figure 9. It is worth mentioning that these real-time results are predicted 24 time-steps (24 hours) ahead of time, and therefore may have worsen prediction performance when compared with results of 1 time-step-ahead predictions. Furthermore, the hyper-parameters of each method were fine-tuned by PSO or brute force search to give reasonable results. Polynomial Fourier Transformation Extrapolation with 10 harmonic functions was used, which provided a good indication of the monthly temporal trend. However, the magnitude of prediction was deteriorated due to the repetition of 24 time-steps. Using the regression approach, differenced first-order autoregressive integrated moving average (ARIMA) correctly identified the magnitude of the energy peaks giving reasonable spatial predictions. Unfortunately, due to the averaging nature of the regression iteration, the method is not effective when repeatedly used to predict 24 time-steps ahead. This causes the prediction to have a time lag in predictions. Contrarily, Long Short-Term Memory (LSTM) neural networks (with 128 cells and 4 depth) provides very solid results, with only a slight offset in spatial predictions. The long-term and short-term temporal trend that are predicted by the LSTM neural network shows good correlation with the actual energy generated. Due to the spatial-temporal learning mechanisms in HTM, it is shown to give very close predictions with relative to the actual energy generated. Although anomalies and errors are inevitable, HTM gives a robust prediction such that the actual data is practically indistinguishable from the predicted.

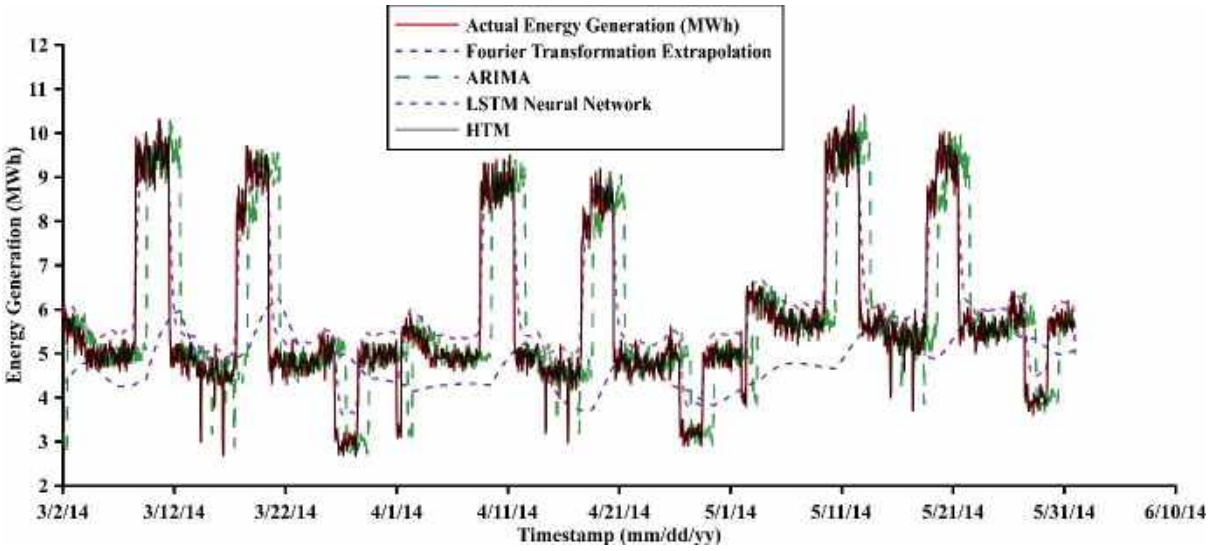


Figure 11.12. Comparison of HTM with other methods during validation

Three accuracy metrics were used for benchmarked in Figure 9, which are Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R^2). In all cases, HTM out-performs the conventional methods. HTM has resulted in a magnitude smaller of MAE and MSE (note that Fig. 9 is plotted in logarithmic scale) while providing R-squared values close to 1.

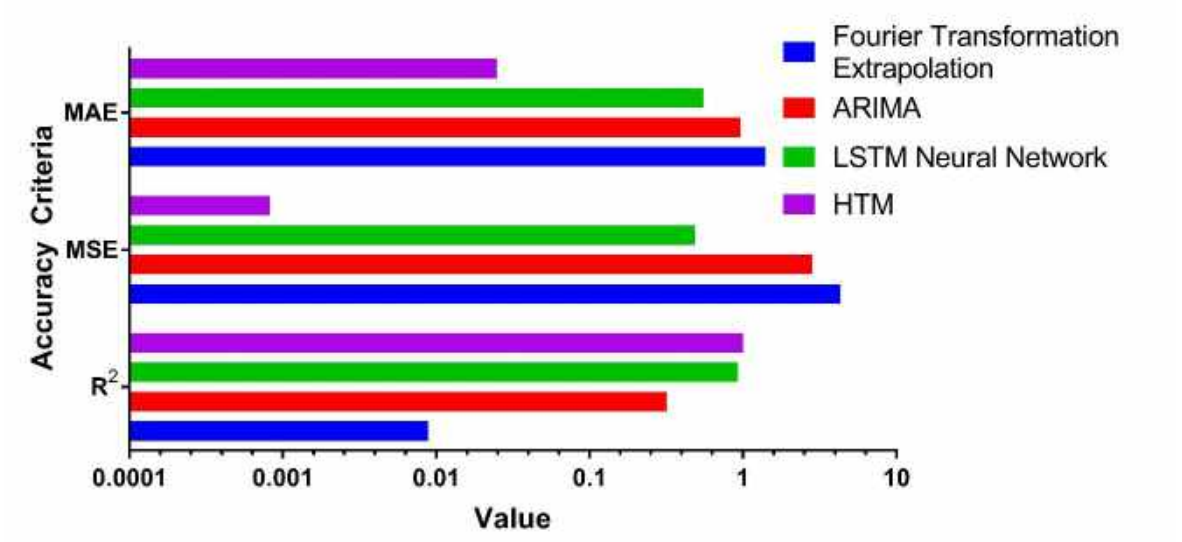


Figure 11.13. Validation accuracy of HTM compared with other methods (log-scaled for value)

Three cases of simulation were carried out for optimization benchmarking, which are HTM with optimization, HTM only and the actual energy response from human. The declared limits of the energy production are done 24 hours prior to the actual data point. Having the actual energy generation exceeding the upper limit will lead to a by-pass operation, which reduces the revenue in the plant. Meanwhile, if the energy generation drops below the lower limit, a heavy penalty of 200 Euros will be carried out. From Figure 11.14(a), the optimization algorithm shifts the predicted energy towards the upper limit to leverage more profits and avoid the heavy penalty costs at the same time. At some rare points, the energy generation exceeds the upper limits, however, the costs are justified. The HTM in this case is optimized by a dual-mode optimization (DMO) algorithm which determines whether the prediction data from the HTM is reliable or in anomaly. If an anomaly is expected, the optimization algorithm switches to a robust algorithm and optimizes the declared energy limits based on maximum expectation of effective gross profits. In normal cases, the predictions from HTM is used as the ground truth and the decision of the declared energy limits is optimized to maximize the direct effective gross profit.

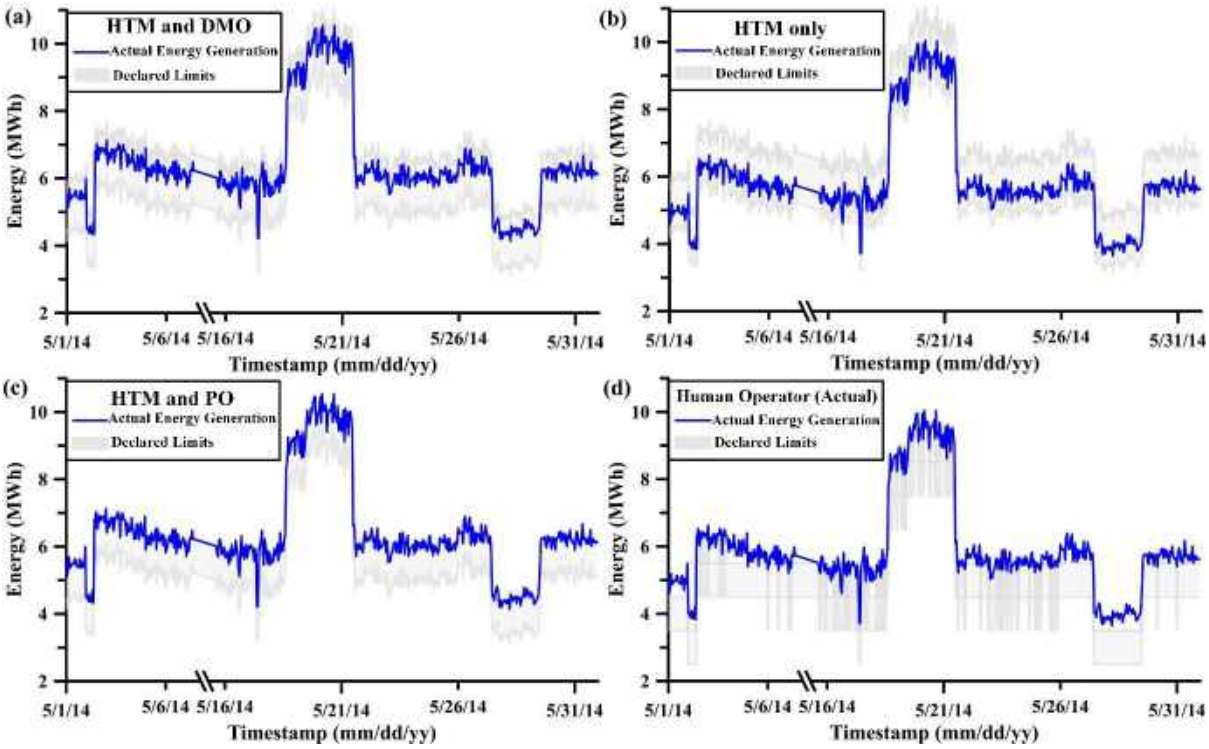


Figure 11.14: Energy response and declared limits of waste-to-energy plant of using (a) HTM and DMO (b) HTM only (c) HTM and PO (d) Human Operator

Using only HTM without the optimization algorithm, the energy generation is relatively centralized in between the limits (see Fig. 11.14(b)). However, this does not maximize the economic potential of the energy generation system. In this case, the model is only trying to obtain a reliable prediction by minimizing the error between the day ahead estimated and the real value. This also means that the decision-making aspect of the energy management system is taking the prediction of the machine learning algorithm as the ground truth without considering the uncertainty and costs importance. It is worth mentioning that the most significant difference using HTM with the optimization algorithm is that the energy management system does not make more conservative actions (i.e. lower energy limits) when the prediction model detects anomaly. Next, with only the use of HTM and predictive optimization (Fig. 11.14(c)), the upper limit is predictively set to the energy to avoid large amount of penalty due to under-generation of power. This strategy can avoid the heaviest penalty (under-generation) at all time, however, frequently gives overproduction of energy due to fluctuations, causing a slightly lowered plant economics.

As a practical benchmark, the actual energy limits that were specified by a human operator are shown in Figure 11.14(d). In the waste-to-energy plant, the operator can observe the waste input, weather, machine condition and other operational conditions to provide educated

estimations on the power generated. In this aspect, the human operator has more useful information for energy prediction than HTM-based predictors however HTM predictions were more efficient. From the energy response, it can be observed that the frequency of limit changes of the human operator is lesser (than the proposed algorithms), while limits are predicted in a less aggressive manner. It is also worth highlighting that the human operator often over-produces the energy (or under specify the lower limit) to avoid getting heavily penalized by not meeting the declared lower limit.

Finally, the total estimated gross profit from each method is presented for HTM and dual-mode optimization (DMO), HTM and predictive optimization (PO), ARIMA and PO, FTE and PO, LSTM and PO, persistence forecasting and the actual human operator from company A (see Fig. 11.15(a)). The highest total gross profit for 6 months was achieved by HTM and DMO with a value of 35450 Euro. The use of DMO which switches from PO to RO under anomaly detection has improved the results by 11.6 % compared to HTM and only PO. Since the case study, company A originally relies on the expert human operator, the original energy response from the operator is used as the primary benchmark. From the result, it is also apparent that the energy response from a human expert can provide gross profit which is approximately 40% better than simple persistence forecasting, thus explaining the common necessity of experienced operators in waste-to-energy plants. The use of HTM with DMO over human operator has improved the gross profit by approximate 7600 Euros within 6 months which is equivalent to a 27.3 % improvement. When observing the increase of total gross profit over time in Figure 11.15(b), the seesaw pattern in the graph generally indicate drop in profit due to anomaly. Methods that suffer from anomaly are ARIMA and persistence forecasting. Moreover, in Figure 11.15(b), methods which have a better prediction accuracy generally provides higher rate of increase in profits.

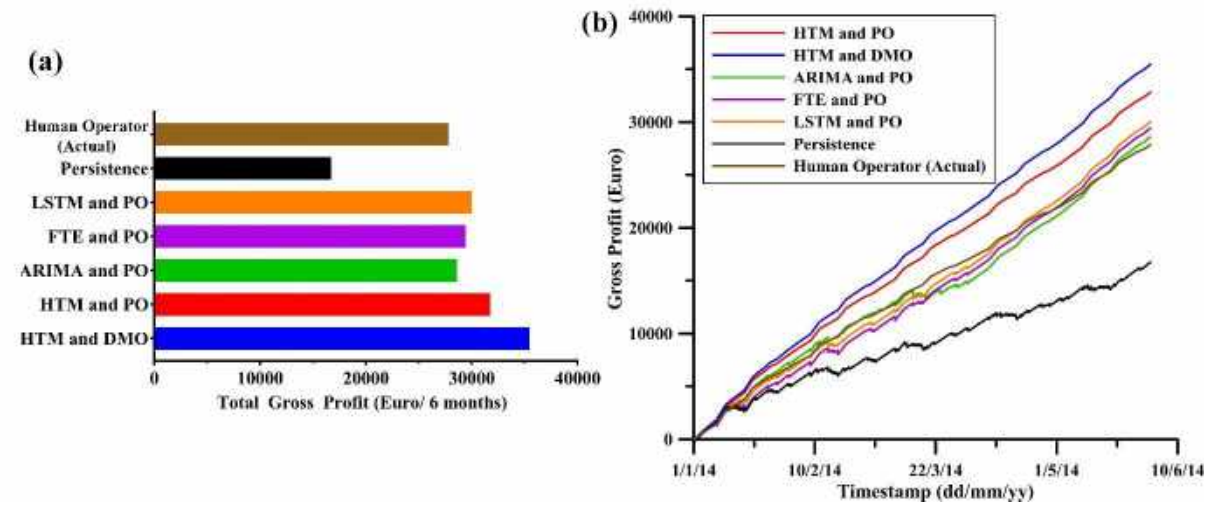


Figure 11.15 Figure of different waste-to-energy management method and their (a) total gross profit (b) gross profit with time

With such results, this work further highlights the usefulness of the combined HTM and DMO algorithm in energy management systems. Firstly, one of the main challenges of implementing predictive algorithms within existing energy managing facilities is on the robustness of such algorithms under anomalous conditions (Máša et al., 2018). By utilizing HTM, the algorithm was able to effectively update and learn its definition of anomaly from energy data. Furthermore, the HTM model can send a signal in the form of an anomaly binary to the novel optimization algorithm which acts differently between robust and predictive mode. This mimics a human's behaviour to take safer decisions when the scenario is unfamiliar or in uncertainty and greatly improves the model's adaptability with anomaly. Next, this model provides a data-driven algorithmic strategy towards energy management. This greatly reduces the tedious workload of the human operator in energy demand response and can automatically provide decisions. Lastly, it is important to highlight that this model is cost efficient and non-process invasive. The model only requires real-time data on the energy production, which is very commonly available in energy managing facilities. It is possible that predictive algorithms require extra data sensors (Herrmann et al., 2011) that might be expensive and would require invasive engineering installation. The proposed model proves practicality and effectiveness and can be highly implementable in various energy management systems.

11.5 Conclusion

The use of biology-inspired Hierarchical Temporal Memory (HTM) coupled with an optimization algorithm has yield accurate predictions and provided anomaly detection for a waste-to-energy

industrial case study. A dual-mode algorithm (DMO) which switches between PO and RO is proposed with the use of HTM. It is demonstrated that the HTM is robust enough to provide sensible data prediction under noise. In terms of validation accuracy, HTM outperforms Fourier Transformation Extrapolation, autoregressive integrated moving average (ARIMA) regression and Long Short-Term Memory (LSTM) neural network with Mean Square Error of 0.8466 %. While simulating different strategies using HTM and multiple different methods, it was found that the proposed dual-mode optimization algorithm can leverage more economic potential within the boundaries of the energy production limits. This led to an increase of 11.6% for total estimated gross profit when compared to HTM with predictive optimization and an increase of 27.4% with comparison to an expert operator in the waste-to-energy cogeneration plant.

Nomenclature

Abbreviations	Definition
AGI	Artificial General Intelligence
ARIMA	Autoregressive Integrated Moving Average
DMO	Dual-Mode Optimization
ELM	Extreme Learning Machine
FTE	Fourier Transform Extrapolation
GMDH	Group Method Data Handling
HTM	Hierarchical Temporal Memory
LSTM	Long Short-Term Memory
MPC	Model Predictive Control
PO	Predictive Optimization
RO	Robust Optimization

References

Abdel-Aal, R.E., Elhadidy, M.A., Shaahid, S.M., 2009. Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks. Renew. Energy. https://doi.org/10.1016/j.renene.2009.01.001

Ahmad, A., Khan, J.Y., 2019. Real-Time Load Scheduling , Energy Storage Control and Comfort Management for Grid-Connected Solar Integrated Smart Buildings. Appl. Energy 114208. https://doi.org/10.1016/j.apenergy.2019.114208

Ahmad, S., Hawkins, J., 2015. Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory 1–18.

Ahmad, S., Lavin, A., Purdy, S., Agha, Z., 2017. Unsupervised real-time anomaly detection for streaming data. Neurocomputing 262, 134–147. https://doi.org/10.1016/j.neucom.2017.04.070

Arabian-Hoseynabadi, H., Oraee, H., Tavner, P.J., 2010. Failure Modes and Effects Analysis (FMEA) for wind turbines. Int. J. Electr. Power Energy Syst. 32, 817–824. https://doi.org/10.1016/j.ijepes.2010.01.019

Atthajariyakul, S., Leephakpreeda, T., 2004. Real-time determination of optimal indoor-air condition for thermal comfort, air quality and efficient energy usage, in: Energy and Buildings. https://doi.org/10.1016/j.enbuild.2004.01.017

Bao, Y., Luo, Y., Zhang, W., Huang, M., Wang, L.Y., Jiang, J., 2018. A bi-level optimization approach to charging load regulation of electric vehicle fast charging stations based on a battery energy storage system. Energies 11. https://doi.org/10.3390/en11010229

Beiraghi, M., Ranjbar, A.M., 2011. Discrete Fourier Transform based approach to forecast monthly peak load, in: Asia-Pacific Power and Energy Engineering Conference, APPEEC. https://doi.org/10.1109/APPEEC.2011.5748585

Brunner, P.H., Rechberger, H., 2015. Waste to energy - key element for sustainable waste management. Waste Manag. https://doi.org/10.1016/j.wasman.2014.02.003

Chachuat, B., Srinivasan, B., Bonvin, D., 2009. Adaptation strategies for real-time optimization. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2009.04.014

Chae, S.H., Kim, S.H., Yoon, S.G., Park, S., 2010. Optimization of a waste heat utilization network in an eco-industrial park. Appl. Energy. https://doi.org/10.1016/j.apenergy.2009.12.003

Chen, C., Duan, S., Cai, T., Liu, B., Hu, G., 2011. Smart energy management system for optimal microgrid economic operation. IET Renew. Power Gener. 5, 258. https://doi.org/10.1049/iet-rpg.2010.0052

Chen, X., Wang, W., Li, W., 2012. An overview of Hierarchical Temporal Memory: A new neocortex algorithm. Model. Identif. Control (ICMIC), 2012 Proc. Int. Conf. 1004–1010.

Cooley, J.W., Tukey, J.W., 1965. An Algorithm for the Machine Calculation of Complex Fourier Series. Math. Comput. https://doi.org/10.2307/2003354

Cui, Y., Ahmad, S., Hawkins, J., 2017. The HTM Spatial Pooler—A Neocortical Algorithm for Online Sparse Distributed Coding. Front. Comput. Neurosci. 11, 1–15. https://doi.org/10.3389/fncom.2017.00111

Dai, C., Li, Y.P., Huang, G.H., 2011. A two-stage support-vector-regression optimization model for municipal solid waste management - A case study of Beijing, China. J. Environ. Manage. https://doi.org/10.1016/j.jenvman.2011.06.038

Ding, Y.M., Hong, S.H., Li, X.H., 2014. A demand response energy management scheme for industrial facilities in smart grid. IEEE Trans. Ind. Informatics 10, 2257–2269. https://doi.org/10.1109/TII.2014.2330995

George, D., 2009. How to make computers that work like the brain. 2009 46th ACM/IEEE Des. Autom. Conf. 420. https://doi.org/10.1145/1629911.1630024

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. Neural Comput. https://doi.org/10.1162/089976600300015015

Geurts, M., Box, G.E.P., Jenkins, G.M., 1977. Time Series Analysis: Forecasting and Control. J. Mark. Res. https://doi.org/10.2307/3150485

González-Romera, E., Jaramillo-Morán, M.A., Carmona-Fernández, D., 2008. Monthly electric energy demand forecasting with neural networks and Fourier series. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2008.06.004

Hawkins, Jeff George, D., 2006. Hierarchical Temporal Memory Concepts, Theory, and Terminology.

Hawkins, J., Ahmad, S., 2016. Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex. Front. Neural Circuits 10, 1–13. https://doi.org/10.3389/fncir.2016.00023

Hawkins, J., Ahmad, S., Cui, Y., 2017. A Theory of How Columns in the Neocortex Enable Learning the Structure of the World. Front. Neural Circuits 11, 1–18. https://doi.org/10.3389/fncir.2017.00081

Hawkins, J., Sandra, B., 2007. On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines. Macmillan.

He, Y., Qin, Y., Wang, S., Wang, X., Wang, C., 2019. Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network. Appl. Energy. https://doi.org/10.1016/j.apenergy.2018.10.061

Herrmann, C., Suh, S., Bogdanski, G., Zein, A., Cha, J., Um, J., Jeong, S., Guzman, A., 2011. Context-Aware Analysis Approach to Enhance Industrial Smart Metering. https://doi.org/10.1007/978-3-642-19692-8

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. https://doi.org/10.1162/neco.1997.9.8.1735

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.

Kostavelis, I., Gasteratos, A., 2012. On the optimization of Hierarchical Temporal Memory. Pattern Recognit. Lett. 33, 670–676. https://doi.org/10.1016/j.patrec.2011.11.017

Kuznetsova, E., Cardin, M.A., Diao, M., Zhang, S., 2019. Integrated decision-support methodology for combined centralized-decentralized waste-to-energy management systems design. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2018.12.020

Lan, H., Zhang, C., Hong, Y.Y., He, Y., Wen, S., 2019. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. Appl. Energy. https://doi.org/10.1016/j.apenergy.2019.04.056

Lindberg, C.F., Zahedian, K., Solgi, M., Lindkvist, R., 2014. Potential and limitations for industrial demand side management. Energy Procedia 61, 415–418. https://doi.org/10.1016/j.egypro.2014.11.1138

Liu, Y., Qin, H., Zhang, Z., Pei, S., Wang, C., Yu, X., Jiang, Z., Zhou, J., 2019. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. Appl. Energy. https://doi.org/10.1016/j.apenergy.2019.113596

Marchetti, A., Chachuat, B., Bonvin, D., 2009. Modifier-adaptation methodology for real-time optimization. Ind. Eng. Chem. Res. https://doi.org/10.1021/ie801352x

Marchetti, D.J., Parigi, G., 2002. Energy consumption, survey data and the prediction of industrial production in Italy: a comparison and combination of different models. J. Forecast. 19, 419–440. https://doi.org/10.1002/1099-131x(200009)19:5<419::aid-for749>3.3.co;2-a

Máša, V., Stehlík, P., Touš, M., Vondra, M., 2018. Key pillars of successful energy saving projects in small and medium industrial enterprises. Energy 158, 293–304. https://doi.org/10.1016/j.energy.2018.06.018

Máša, V., Touš, M., Pavlas, M., 2016. Using a utility system grey-box model as a support tool for progressive energy management and automation of buildings. Clean Technol. Environ. Policy 18, 195–208. https://doi.org/10.1007/s10098-015-1006-x

Melis, W.J.C., Chizuwa, S., Kameyama, M., 2009. Evaluation of hierarchical temporal memory for a real world application. 2009 4th Int. Conf. Innov. Comput. Inf. Control. ICICIC 2009 144–147. https://doi.org/10.1109/ICICIC.2009.195

Mnatzaganian, J., Fokoué, E., Kudithipudi, D., 2016. A Mathematical Formalization of Hierarchical Temporal Memory's Spatial Pooler 3, 1–14. https://doi.org/10.3389/frobt.2016.00081

Namuli, R., Jaumard, B., Awasthi, A., Pillay, P., 2013. Optimisation of biomass waste to energy conversion systems for rural grid-connected applications. Appl. Energy. https://doi.org/10.1016/j.apenergy.2012.06.011

Oh, E., Son, S.Y., 2018. Energy-storage system sizing and operation strategies based on discrete Fourier transform for reliable wind-power generation. Renew. Energy. https://doi.org/10.1016/j.renene.2017.10.028

Perrin, N., Sibly, R.M., 2003. Dynamic Models of Energy Allocation and Investment. Annu. Rev. Ecol. Syst. 24, 379–410. https://doi.org/10.1146/annurev.es.24.110193.002115

Putna, O., Janošťák, F., Šomplák, R., Pavlas, M., 2018. Demand modelling in district heating systems within the conceptual design of a waste-to-energy plant. Energy. https://doi.org/10.1016/j.energy.2018.08.059

Qing, X., Niu, Y., 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy. https://doi.org/10.1016/j.energy.2018.01.177

Rozado, D., Rodriguez, F.B., Varona, P., 2010. Optimizing hierarchical temporal memory for multivariable time series. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 6353 LNCS, 506–518. https://doi.org/10.1007/978-3-642-15822-3_62

Sen, P., Roy, M., Pal, P., 2016. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. Energy. https://doi.org/10.1016/j.energy.2016.10.068

Shi, Y., Eberhart, R., 2002. A modified particle swarm optimizer 69–73. https://doi.org/10.1109/icec.1998.699146

Shoreh, M.H., Siano, P., Shafie-khah, M., Loia, V., Catalão, J.P.S., 2016. A survey of industrial applications of Demand Response. Electr. Power Syst. Res. 141, 31–49. https://doi.org/10.1016/j.epsr.2016.07.008

Šomplák, R., Ferdan, T., Pavlas, M., Popela, P., 2013. Waste-to-energy facility planning under uncertain circumstances. Appl. Therm. Eng. https://doi.org/10.1016/j.applthermaleng.2013.04.003

Šomplák, R., Pavlas, M., Ucekaj, V., Popela, P., 2012. Waste-to-energy facility planning supported by stochasting programming - Part I introduction, in: Chemical Engineering Transactions. https://doi.org/10.3303/CET1229109

Srinivasan, D., 2008. Energy demand prediction using GMDH networks. Neurocomputing 72, 625–629. https://doi.org/10.1016/j.neucom.2008.08.006

Stehlík, P., 2016. Up-to-Date Waste-to-Energy Approach, SpringerBriefs in Applied Sciences and Technology. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-15467-1

Stehlík, P., 2009a. Efficient waste processing and waste to energy: Challenge for the future. Clean Technol. Environ. Policy. https://doi.org/10.1007/s10098-008-0185-0

Stehlík, P., 2009b. Contribution to advances in waste-to-energy technologies. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2009.02.011

Suganthi, L., Samuel, A.A., 2012. Energy models for demand forecasting - A review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2011.08.014

Sun, M., Feng, C., Zhang, J., 2019. Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation. Appl. Energy. https://doi.org/10.1016/j.apenergy.2019.113842

Touš, M., Bébar, L., Houdková, L., Pavlas, M., Stehlík, P., 2009. Waste-to-Energy (W2E) software -a support tool for decision making process, in: Chemical Engineering Transactions. https://doi.org/10.3303/CET0918159

Touš, M., Frýba, L., Pavlas, M., 2013. Improving calculation of lower heating value of waste by data reconciliation - Analysis and evaluation, in: Chemical Engineering Transactions. https://doi.org/10.3303/CET1335146

Touš, M., Pavlas, M., Putna, O., Stehlík, P., Crha, L., 2015a. Combined heat and power production planning in a waste-to-energy plant on a short-term basis. Energy 90, 137–147. https://doi.org/10.1016/j.energy.2015.05.077

Touš, M., Pavlas, M., Putna, O., Stehlík, P., Crha, L., 2015b. Combined heat and power production planning in a waste-to-energy plant on a short-term basis. Energy. https://doi.org/10.1016/j.energy.2015.05.077

Tsai, C.L., Chen, W.T., Chang, C.S., 2016. Polynomial-Fourier series model for analyzing and predicting electricity consumption in buildings. Energy Build. 127, 301–312. https://doi.org/10.1016/j.enbuild.2016.05.083

Wang, S., Ng, T.S., Wong, M., 2016. Expansion planning for waste-to-energy systems using waste forecast prediction sets. Nav. Res. Logist. https://doi.org/10.1002/nav.21676

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., Kang, C., 2019. Probabilistic individual load forecasting using pinball loss guided LSTM. Appl. Energy. https://doi.org/10.1016/j.apenergy.2018.10.078

Winzer, C., 2012. Conceptualizing energy security. Energy Policy 46, 36–48. https://doi.org/10.1016/j.enpol.2012.02.067

Xenos, D.P., Cicciotti, M., Kopanos, G.M., Bouaswaig, A.E.F., Kahrs, O., Martinez-Botas, R., Thornhill, N.F., 2015. Optimization of a network of compressors in parallel: Real Time Optimization (RTO) of compressors in chemical plants - An industrial case study. Appl. Energy. https://doi.org/10.1016/j.apenergy.2015.01.010

Yao, C., Yang, C., Xiong, Z., 2016. Energy-Saving Predictive Resource Planning and Allocation. IEEE Trans. Commun. 64, 5078–5095. https://doi.org/10.1109/TCOMM.2016.2608822

Younes, M.K., Nopiah, Z.M., Basri, N.E.A., Basri, H., 2014. Medium term municipal solid waste generation prediction by autoregressive integrated moving average, in: AIP Conference Proceedings. https://doi.org/10.1063/1.4894366

Yu, M., Hong, S.H., 2017. Incentive-based demand response considering hierarchical electricity market: A Stackelberg game approach. Appl. Energy 203, 267–279. https://doi.org/10.1016/j.apenergy.2017.06.010

Yu, M., Lu, R., Hong, S.H., 2016. A real-time decision model for industrial load management in a smart grid. Appl. Energy 183, 1488–1497. https://doi.org/10.1016/j.apenergy.2016.09.021

Yuan, C., Liu, S., Fang, Z., 2016. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. Energy. https://doi.org/10.1016/j.energy.2016.02.001

Zobaa, A.F., Bansal, R.C., 2011. Handbook of Renewable Energy Technology. World Scientific.

Appendix

Figure A11.1: Model Parameters for Hierarchical Temporal Memory (HTM)

Aspect	Parameter Name	Value/ Settings
Spatial Pooler	Column Count	2048
	Global Inhibition	1
	Boost Strength	2.0
	Numbers of Activated Columns per Inhibition Area	40
	Seed	1956
	Synapse Permanence Active Incrementation	0.05
	Synapse Permanence Connect Threshold	0.1
	Synapse Permanence Inactive Decrement	0.08568
Temporal Memory	Activation Threshold	12
	Cells per Column	32
	Column Count	2048
	Initial Permanence	0.21
	Input Width	2048
	Maximum Segments per Cell	128
	Maximum Synapses per Segment	32
	Minimum Threshold	10
	New Synapse Count	20
	Pay Attention Mode (PAM) Length	1
	Permanence Decrement	0.1
	Permanence Increment	0.1
	Seed	1960

CHAPTER 12 ON THE IMPLEMENTATION OF A NEURAL NETWORK IN THE PROCESS GRAPH FRAMEWORK

This work has been published in Chemical Engineering Transaction.

How B.S., Teng S.Y., Leong W.D., Ng W.P.Q., Lim C.H., Ngan S.L., Lam H.L., 2019, Non-linear Programming via P-graph Framework, Chemical Engineering Transactions, 76, 499-504.

Abstract: P-graph (Process Graphs) is a graph theoretic method which is designed to solve process network synthesis (PNS) problem using combinatorial and optimisation algorithms. Due to its visual interface for data encoding and results display; and its capability of generating multiple solutions (optimal and sub-optimal) simultaneously, the utility of P-graph has expanded into a broad range of studies recently. However, this powerful graph-theoretic method still falls short of dealing non-linear problems. The problem can be found from the cost estimation provided by P-graph software. Despite it allows users to input the sizing cost (noted as "proportional cost" in P-graph software), the capacity and the cost are assumed to be linearly correlated. This inaccurate and unreliable cost estimation has increased the difficulty of making wise decision and therefore lead to undesirable profit loss. This chapter proposes to solve the fundamental linearity problem by implementing trained artificial neural networks (ANN) into P-graph. To achieve this, an ANN model which utilised *thresholded rectified linear unit* (ReLU) activation function is developed in a segregated computational tool. The identified neurons are then modelled in P-graph to convert the input into the nonlinear output. To demonstrate the effectiveness of the proposed method, an illustrative case study of biomass transportation is used. With the use of the trained neurons, the non-linear estimation of transportation cost which considered fuel consumption cost, vehicle maintenance cost and labour cost is successfully modelled in P-graph. This work is expected to pave ways for P-graph users to expand the utility of P-graph in solving other more complex non-linear problems.

12.1 Introduction

P-graph framework has gained loads of attention in the field of process system engineering (PSE) nowadays. P-graph framework, in short, is a graph-theoretic approach which was developed in the early 1990s by Friedler et al. (1992). Its visual appealing nature has been granted as an "attractive and user-friendly optimisation tool" by researchers and students (Lam et al., 2016). A more recent work conducted by Promentilla et al. (2017) further highlighted its pedagogical advantages in Problem-Based Learning (PBL). Over the past two decades, P-graph

methodology had been applied to solve a wide range of PNS problems, including conventional chemical process design synthesis (e.g., reaction pathway identification (Fan et al., 2001), separation process synthesis (Kovacs et al., 1999), azeotropic distillation system (Feng et al., 2000), etc.); biological process synthesis (Fan et al., 2019); biomass network optimisation (How et al., 2018); optimisation of Total Site utility system (Walmsley et al., 2018); and carbon management networks (Aviso et al., 2019). Aside from journal articles, P-graph has been successfully penetrated the current chemical engineering community. Evidently, P-graph framework has been introduced in undergraduate textbooks (Peters et al., 2003), reference books (Klemeš et al., 2011) and chemical engineering magazine (Cabezas et al., 2015).

However, the main drawback of the graph-theoretic method is its incapability in solving non-linear problems. Nevertheless, few researchers have attempted to address this critical limitation. For instance, Aviso and Tan (2018) had implemented the fuzzy P-graph approach (developed by Tick (2009)) in synthesising a poly-generation system. Despite it is still merely a fuzzy linear programming, it shows the possibility of extending P-graph toward non-linear programming. Earlier, Ong et al. (2016) had modelled the non-linear cost function using a piece-wise linear approach. The concept of using virtual operating units to represent the segmented sub-cost-functions is admirable, but the key limitations of this proposed method include (i) the non-linear functions must be known; and (ii) ineffective for problems with more than two dimensions.

To further extend the capability of P-graph in solving non-linear problems, this work therefore proposes a hybrid approach which utilise P-graph framework in conjunction with the use of artificial neural networks (ANN). The conception idea is to utilise the trained neural network to convert linear-input-neurons into nonlinear P-graph model.

12.2 Methods

12.2.1 Trained Neural Network in P-graph Structure

Artificial neural network was introduced by McCulloch and Pitts (1943) and initially aimed to model nervous activity in the human brain. Recently, it serves as a universal black box model which can be applied in a wide range of research (Hornik et al., 1989). In addition, its capability of modelling non-linear problem had been further confirmed by Lek et al. (1996) in their ecology study. As an illustrative example, a multi-layer perceptron type artificial network is implemented in this study. Generally, all neurons in the former layer are directly linked to the

neurons in the next layer. They receive the signals from the neurons located in the previous layer directly before it; and transmit signals to all the neurons located in the subsequent layer directly after it. The general mathematical expression of the signal propagation is expressed as:

$$A_j = f_{ac} \left(\sum_{i=1}^{n} A_i w_i + B_j \right) \quad (12.1)$$

Where A_j denotes the activation of the j^{th} neuron in one layer, while the activation of i^{th} neuron in the layer directly before the layer of j^{th} neuron is referred to as A_i. The assigned weight for each connection between the connecting neurons is indicated as w_i; while B_j refers to the "bias" which improves the modelling capabilities of the neural network. Note that the thresholded rectified linear units (ReLU) is used as the activation function (Konda et al., 2014), f_{ac} in this work (see Eqn. 12.3, where $z \in \mathbb{R}$) since it is found more efficient compared to the conventional sigmoid units (Rynkiewicz, 2019). These parameters can be determined by minimising the mean square error (MSE) between the actual values (X^{ACT}) and estimated values (X^{EST}), where N indicates the total number of sample sizes.

$$\min \quad MSE = \frac{\sum (X^{EST} - X^{ACT})}{N} \quad (12.2)$$

$$f_{ac}(z) = \begin{cases} z, & x > a \\ 0, & otherwise \end{cases} \quad (12.3)$$

Figure 1 presents the illustrative examples of how a multi-layer perceptron type of neural network can be represented in P-graph model. Note that the construction of hidden layers might be different due to the neuron condition. This is to ensure the activation function in Eq(3) is fully satisfied (i.e., negative summed treated as "0").

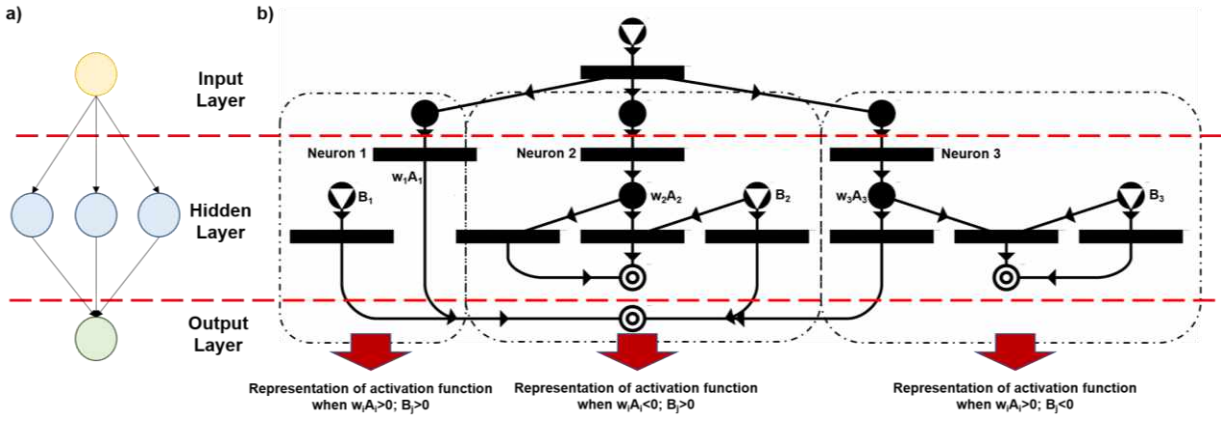


Figure 12.1: Illustration of (a) typical multi-layer perceptron type neural network; (b) neural network in P-graph model

12.2.2 Transportation Cost Estimation

A biomass transportation example is used to demonstrate the proposed method. The non-linear transportation cost function which was developed by How et al. (2016) is adopted in this work. Note that the total transportation cost, C^{TR} encompasses three components, i.e., fuel consumption cost, C^{FUEL}, vehicle maintenance cost C^{MAINT} and labour cost C^{LAB}. These components are expressed in Eqn. 12.4 to 12.5:

$$\min \ C^{TR} = C^{FUEL} + C^{MAINT} + C^{LAB} \tag{12.3}$$

$$C^{FUEL} = N^{TRIP} \times 2 \times d \times r^{FUEL} \times C^{DIESEL} \tag{12.4}$$

$$C^{MAINT} = N^{TRIP} \times 2 \times d \times C^{REPAIR} \tag{12.5}$$

$$C^{LAB} = N^{TRIP} \times OH^{TRIP} \times C^{WAGES} \tag{12.6}$$

where N^{TRIP} refers to the total number of trips required (a function of the transferred load and vehicle capacity constraint); d refers to the distance between the two locations; r^{FUEL} denotes the fuel consumption rate; OH^{TRIP} indicates the delivery time required per trip; while C^{DIESEL}, C^{REPAIR} and C^{WAGES} refer to the diesel price, estimated repair cost per unit of distance travelled and the hourly wages for the workers. For more details, please refer to How et al. (2016).

12.3 Case Study

In this case study, a given amount of biomass is transported to a biorefinery plant via a set of transportation modes (i.e., M1, M2, M3). Table 12.1 summarises the related vehicle data; while other parameters used in this case study are tabulated in Table 12.2.

Table 12.1: Operating specification of transportation modes (How et al., 2016)

Mode	Load Capacity (t)	r^{FUEL} (L/km)	Travel velocity (km/h)	C^{REPAIR} (RM/km)	Loading Delay (h)
M1	15	0.261*	80	0.22	0.50
M2	25	0.278*	75	0.34	0.67
M3	40	0.294*	65	0.45	0.83

*Fuel consumption rate when the vehicle is empty (i.e., r_0^{FUEL}); it is assumed that $r^{FUEL} = 0.001 \times x^2 \times r_0^{FUEL}$, where x refers to the weight of loaded biomass.

Table 12.2: Other parameters used in the case study

Parameter	Value
C^{DIESEL} (RM/L)	2.18
C^{WAGES} (RM/h)	10.00
Daily transferred biomass load (t/d)	0 to 20
d (km)	20
Annual operating days (d/y)	355
Maximum operating hour per day (h/d)	40

12.4 Results and Discussion

The optimal number of neurons required for each transportation mode is summarised in Table 12.3.

Table 12.3: Mean Squared Error obtained

Number of neurons	MSE for M1	MSE for M2	MSE for M3
1	0.0174	0.0032	0.0025
2	0.0012	0.0002	0.0002
3	0.0007	0.0002	0.0002
4	0.0070	0.0008	0.0004

The results show that two neurons are required for the M2 and M3 (having three neurons no longer provide significant improvement); while M1 require three neurons. This is due to the step changes in the outputs for M2-case. All the determined parameters for each neuron are tabulated in Table 12.4.

Table 12.4: Parameters for each neuron

Neuron for M1	w_i	B_j	Neuron for M2	w_i	B_j	Neuron for M3	w_i	B_j
1	0.1111	-0.3312	1	1.000	-0.0168	1	0.5048	-0.5026
2	0.2979	-0.4652	2	0.5303	-0.5482	2	0.9623	-0.0146
3	0.4503	0.9999						

The formulated neural network is then constructed in the form of P-graph (see Figure 12.2). The red-coloured section represents the neural network of M1; green-coloured section represents the neural network of M2, while the blue section refers to the neural network for M3. The optimised results are illustrated as Figure 12.3 and Figure 12.4. The results show that M1 is the optimal transportation mode when the daily delivered biomass load is less than 15 t/d (illustrated as Figure 12.3); while M2 is selected when the daily delivered biomass load exceeds 15 t/d (illustrated as Figure 12.4). This is due to the vehicle capacity constraint of M1, as the biomass loads cannot be fully transferred in a single trip. The model performance is illustrated in Figure 12.5. It shows that the proposed P-graph model is now capable to deal with non-linear systems.

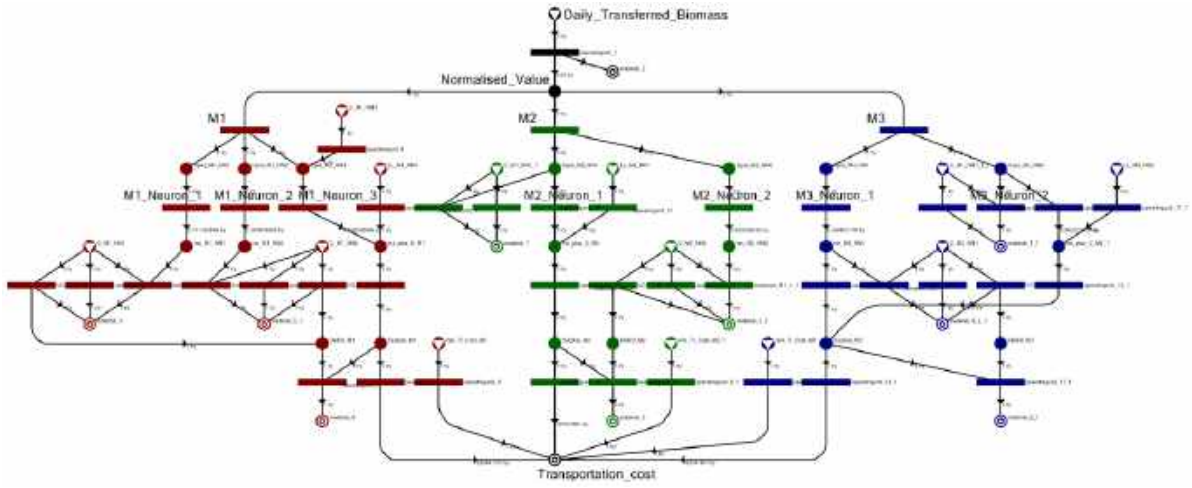


Figure 12.2: P-graph model with the formulated neural networks

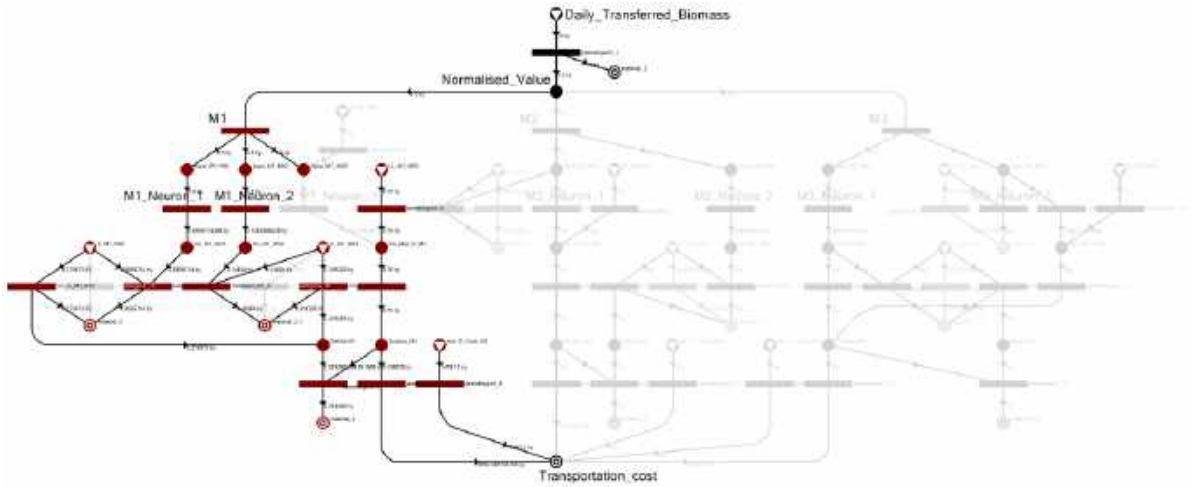


Figure 12.3: Optimal structure when daily biomass transferred load is less than 15 t/d

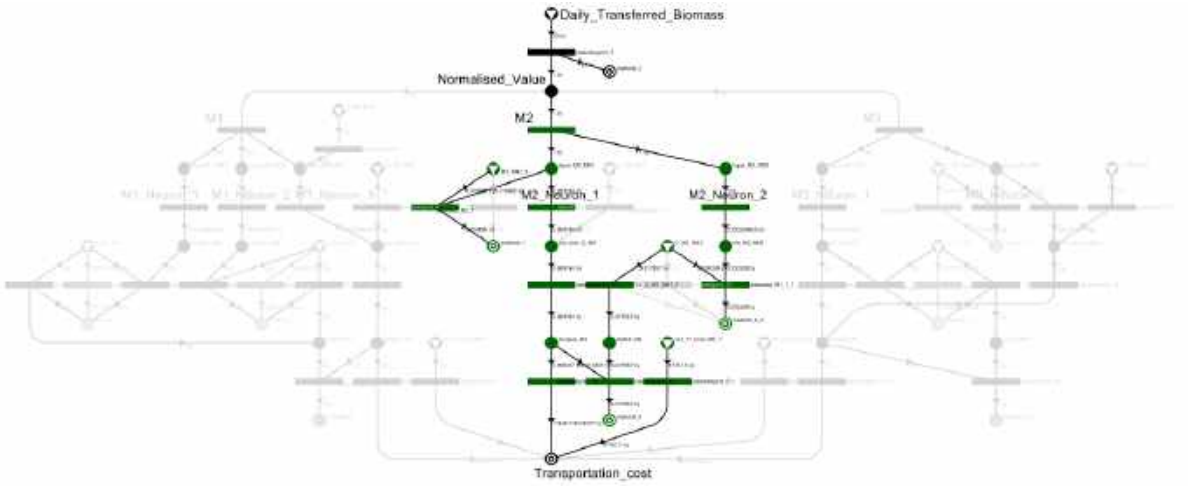


Figure 12.4: Optimal structure when daily biomass transferred load is more than 15 t/d

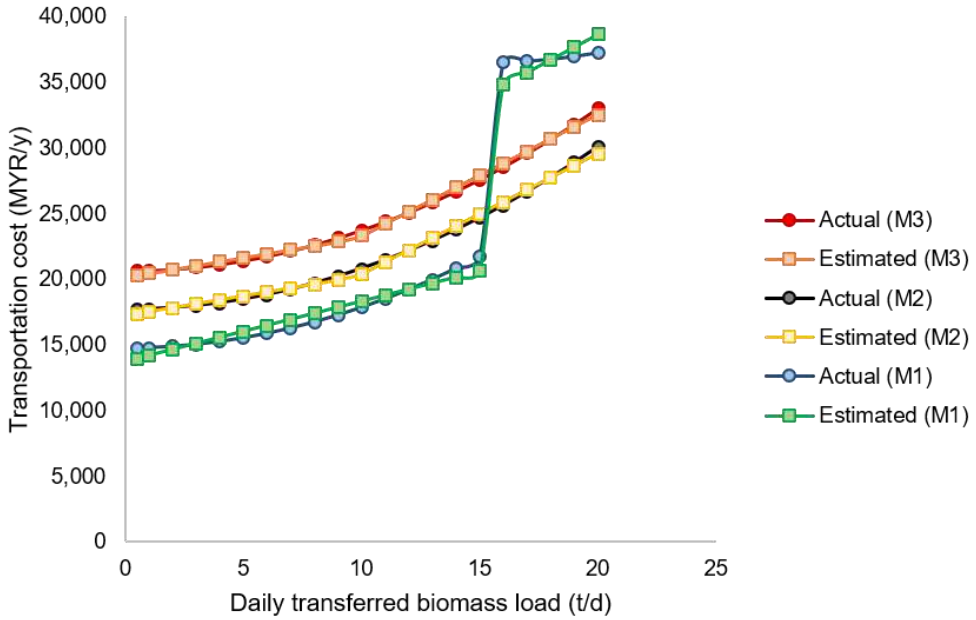


Figure 12.5: Model performance

12.5 Conclusion

This work further consolidates the capability of P-graph framework in non-linear programming by incorporating the use of artificial neural network into the model. The potential of the proposed method can be clearly seen from the model performance chart where the non-linear transportation cost can be determined by using P-graph. With the use of multi-layer perceptron, P-graph is no longer restricted in conventional process system engineering but can also be served as a predictive model without the need to determine the exact non-linear functions. Note that massive amount of data sample is required to ensure the reliability of the attained results. This work can be further extended with the consideration of multiple inputs and outputs in the

neural network, in order to solve non-linear problems, including heat exchanger network optimisation, vapour liquid equilibrium, transportation design (with the consideration of both volume and weight constraint), etc.

References

Aviso K.B., Belmonte B.A., Benjamin M.F.D., Arogo J.I.A., Coronel A.L.O., Janairo C.M.J., Foo D.C.Y., Tan R.R., 2019, Synthesis of optimal and near-optimal biochar-based Carbon Management Networks with P-graph. Journal of Cleaner Production, 214, 893-901.

Aviso K.B., Raymond R.R., 2018. Fuzzy P-graph for optimal synthesis of cogeneration and trigeneration systems, Energy, 154, 258-268.

Cabezas H., Heckl I., Bertok B., Friedler F., 2015. Use the P-graph framework to design supply chains for sustainability, Chemical Engineering Progress, 111, 41-47.

Fan L.T., Bertok B., Friedler F., Shafie S., 2001. Mechanisms of ammonia-synthesis reaction revisited with the aid of a novel graph-theoretic method for determining candidate mechanisms in deriving the rate law of a catalytic reaction, Hungarian Journal of Industry and Chemistry, 29, 71-80.

Fan Y.V., Klemeš, J.J., Perry S., Lee C.T., 2019. Anaerobic digestion of lignocellulosic waste: Environmental impact and economic assessment, Journal of Environmental Management, 231, 352-363.

Feng G., Fan L.T., Seib P.A., Bertok B., Kalotai L., Friedler F., 2003. Graph-Theoretic Method for the Algorithmic Synthesis of Azeotropic-Distillation Systems, Industrial & Engineering Chemistry Research, 42(15), 3602-3611.

Friedler F., Tarjan K., Huang Y.W., Fan L.T., 1992. Combinatorial algorithms for process synthesis, Computers & Chemical Engineering, 16, S313-S320.

Hornik K., Stinchcombe M., White, H., 1989. Multilayer feedforward networks are universal approximators, Neural Networks, 2(5), 359-366.

How B.S., Tan K.Y., Lam H.L., 2016. Transportation decision tool for optimisation of integrated biomass flow with vehicle capacity constraints, Journal of Cleaner Production, 136, 197-223

How B.S., Yeoh T.T., Tan T.K., Chong K.H., Lam H.L., 2018. Debottlenecking of sustainability performance for integrated biomass supply chain: P-graph approach, Journal of Cleaner Production, 193, 720-733.

Klemeš J.J., Friedler F., Bulatov I., Varbanov P., 2011. Sustainability in the Process Industry: Integration and Optimization. McGraw-Hill, New York, U.S.

Konda, K., Memisevic, R., Krueger, D., 2014. Zero-bias autoencoders and the benefits of co-adapting features. *arXiv preprint arXiv:1402.3337*.

Kovacs Z., Ercsey Z., Friedler F., Fan L.T., 1999. Exact super-structure for the synthesis of separation-networks with multiple feed-streams and sharp separators, Computers & Chemical Engineering, 23, S1007-S1010.

Lam H.L., Tan R.R., Aviso K.B., 2016. Implementation of P-graph modules in undergraduate chemical engineering degree programs: experiences in Malaysia and the Philippines, Journal of Cleaner Production, 136(Part B), 254-265.

Lek S., Delacoste M., Baran P., Dimopoulos I., Lauga J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology, Ecological Modelling, 90(1), 39-52.

McCulloch W., Pitts W., 1943. A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics, 5(4), 115-133.

Ong B.H.Y., Walmsley T.G., Atkins M.J., Walmsley M.R.W., Neale J.R., 2016. Approximation of non-linear cost functions in P-graph structures, Chemical Engineering Transactions, 52, 1093-1098.

Peters M.S., Timmerhaus K., West R.E., 2003. Plant Design and Economics for Chemical Engineers (Fifth ed.). McGraw-Hill, New York, U.S.

Promentilla M.A.B., Lucas R.I.G., Aviso K.B., Tan R.R., 2017. Problem-Based Learning of Process Systems Engineering and Process Integration concepts with metacognitive strategies: The case of P-graphs for polygeneration systems, Applied Thermal Engineering, 127, 1317-1325.

Rynkiewicz, J., 2019. Asymptotic statistics for multilayer perceptron with ReLU hidden units, Neurocomputing, (In Press), DOI: doi.org/10.1016/j.neucom.2018.11.097

Tick J., 2009. Fuzzy Extension to P-Graph Based Workflow Models, IEEE 7th International Conference on Computational Cybernetics 2009, 109-112.

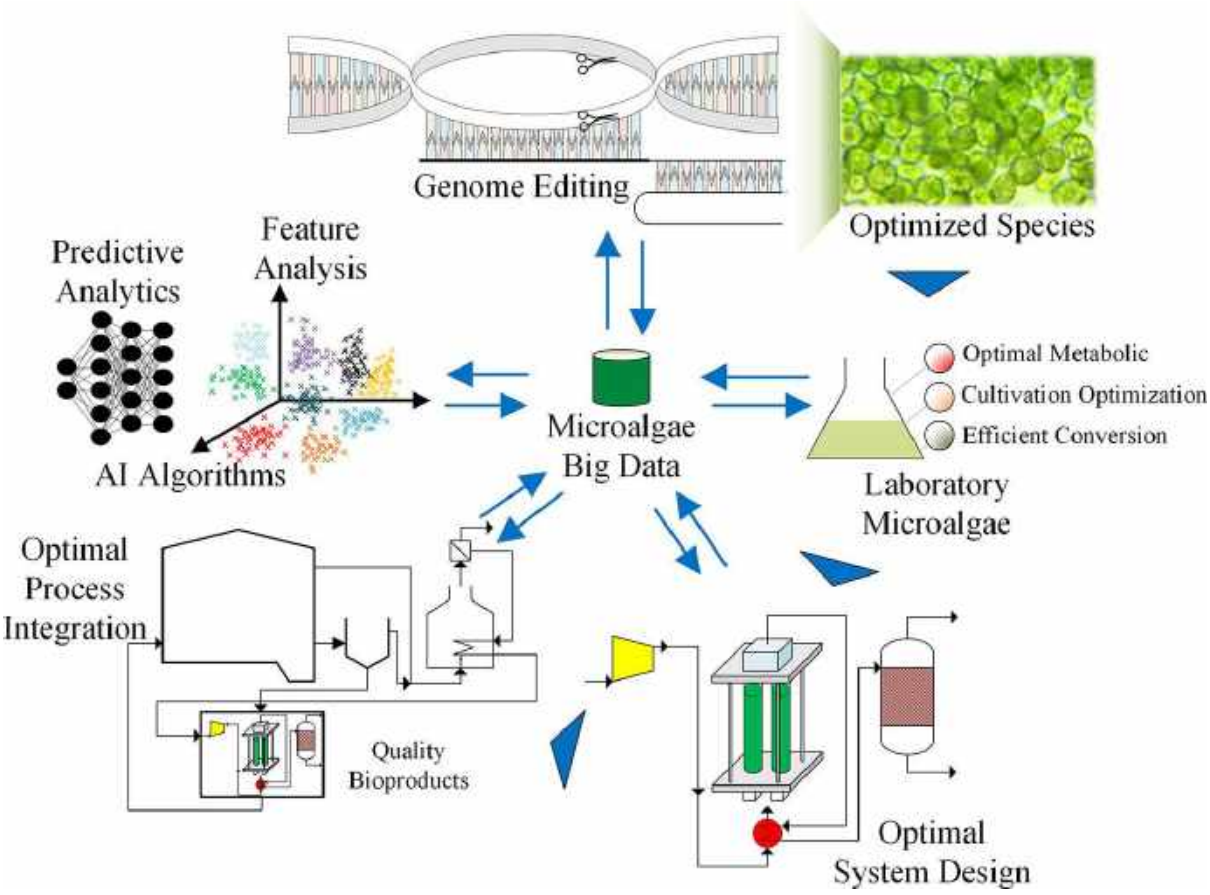
Walmsley T.G., Varbanov P.S., Philipp M., Klemeš, J.J., 2018. Total Site Utility Systems Structural Design Considering Electricity Price Fluctuations, Computer Aided Chemical Engineering, 44, 1159-1164.

CHAPTER 13 FURTHER PERSPECTIVES ON ARTIFICIAL INTELLIGENCE AND BIOSYSTEMS: THE CASE OF MICROALGAE

This work was peer-reviewed and published in Biotechnology Advances.

Teng, S.Y., Yew, G.Y., Sukačová, K., Show, P.L., Máša, V. and Chang, J.S., 2020. Microalgae with artificial intelligence: A digitalized perspective on genetics, systems and products. Biotechnology Advances, p.107631.

Graphical Abstract



Abstract: With recent advances in novel gene-editing tools such as RNAi, ZFNs, TALENs, and CRISPR-Cas9, the possibility of altering microalgae towards designed properties for various application is becoming a reality. Alteration of microalgae genomes can modify metabolic pathways to give elevated yields in lipids, biomass, and other components. The potential of such genetically optimized microalgae can give a "domino effect" in further providing optimization leverages down the supply chain, in aspects such as cultivation, processing, system design, process integration, and revolutionary products. However, the current level of understanding the functional information of various microalgae gene sequences

is still primitive and insufficient as microalgae genome sequences are long and complex. From this perspective, this work proposes to link up this knowledge gap between microalgae genetic information and optimized bioproducts using Artificial Intelligence (AI). With the recent acceleration of AI research, large and complex data from microalgae research can be properly analyzed by combining the cutting-edge of both fields. In this work, the most suitable class of AI algorithms (such as active learning, semi-supervised learning, and meta-learning) are discussed for different cases of microalgae applications. This work concisely reviews the current state of the research milestones and highlight some of the state-of-art that has been carried out, providing insightful future pathways. The utilization of AI algorithms in microalgae cultivation, system optimization, and other aspects of the supply chain is also discussed. This work opens the pathway to a digitalized future for microalgae research and applications.

Keywords: Microalgae, Artificial Intelligence, Genetic Engineering, Process Optimization, System Design, Process Integration

13.1 Introduction

Microalgae are industrially important unicellular organisms with photosynthesis potential to convert carbon dioxide, sunlight, and nutrients into carbohydrates, lipid, and potentially many industrially important compounds (Chew et al., 2017). As a source of modern bioenergy, microalgae are regarded as the third generation of biofuels due to its theoretical carbon-neutral lifecycle (Dragone et al., 2010), non-competitiveness with agricultural or food crops, and possibilities for vertical and high-density cultivation system designs. In this field, microalgae can produce a wide range of fuel such as biodiesel (Chisti, 2008), jet-fuel (Bwapwa et al., 2018a, 2017), hydrogen (Benemann, 2000), syngas (Beneroso et al., 2013), etc. Moreover, microalgae can be harvested and processed to act as raw material for protein-rich food, natural dyes, pharmaceutical products (Khan et al., 2018), etc. Due to a large plethora of different microalgae species and strains (Araujo et al., 2011), there are many possibilities of microalgae property that can be cultivated, giving different functionality.

In the current microalgae research-to-commercialization value chain, the research procedure by identifying the functionality and purpose of microalgae application. Subsequently, high-throughput screening of microalgae strains is carried out by batch experimentation (Taleb et al., 2015). Although previous experimentation and literature data can be used for screening, it

is common that the operating conditions (due to different industrial applications) are different, and thus experimentation is still required. After preliminary screening of microalgae strains, the suitable design of a specific scaled-up bioreactor (Chen et al., 2011) is required to provide an efficient growing environment for the microalgae. In a parallel manner, if further processing of raw products is required, further methods such as thermogravimetric analysis (Figueira et al., 2015), liquid biphasic extraction (Yew et al., 2019), fermentation (Harun et al., 2010), and their optimization (Teng et al., 2019) are carried out to provide an experimental basis for process system design. With this, a tailored-made microalga-based process can be designed to accommodate the industrial objectives. Lastly, with consideration of existing infrastructure within the existing facility, process integration can be carried out to efficiently recover energy and environmental footprint for the overall process system (Chowdhury et al., 2012). A common utilization for this is on the efficient drying of microalgae (Aziz et al., 2014) by waste heat recovery, thus making microalgae much economically feasible.

However, with consideration of the genome length of common microalgae, the aforementioned workflow for microalgae research can provide useful applications but cannot guarantee even near-optimal results within a sensible budget constraint. As an example, for common microalgae (*Chlorella vulgaris* UTEX 395), the genome length is 3600 (Guarnieri et al., 2018), thereby obtaining the genetically optimal microalgae requires a search on $\sim 2.6 \times 10^{2167}$ sequences (as a comparison, atoms in the observable universe is $\sim 10^{82}$), each repeating the same workflow. Additionally, this search space has not accounted for process pathways, integration, and combinations of operational conditions, making this an understatement. With current advances in gene editing for microalgae (with tools such as RNAi, ZFNs, TALENs, and CRISPR-Cas9) and next-generation sequencing (NGS), the experimental exploration of microalgae genome has become much quicker, accurate and popular (Lin et al., 2019). Nevertheless, such exploration requires much assistance from the perspective of bioinformatics (Reijnders et al., 2014), data-driven analytics, and AI to effectively provide a reduced search space for genome functionality exploration, experimental conditions, process design, and integration.

In AI, this field of study has the purpose of imitating intelligence from nature (Russell and Norvig, 2010). AI is a rather broad computational field with various definitions for its categorization and definition (Wang, 2008). Nevertheless, this work considers AI from the perspective of 3 distinct directions being machine learning, metaheuristics, and expert systems.

As a driving-perspective for the fields of microalgae research, the contribution of this paper is it highlights the importance of a data-driven AI-based workflow from genome bioinformatics to process system design and integration. The objective of this work is to concisely undercover the applications of data-driven analytics and AI in microalgae development and elucidate the advantages of doing so. Furthermore, this paper proposes a multi-omics paradigm for data-driven microalgae workflow which can accelerate development time, improve performance efficiencies, and microalgae product design accuracy. Concise review of data-driven methods, techniques, significant research directions, and advantages of such technologies for the purpose of microalgae research has also been presented. Aspects regarding using AI as the core of microalgae-to-commercialization will be discussed in subsequent chapters as (i) genetic engineering in microalgae, (ii) microalgae cultivation and conversion, (iii) system design and control , and (iv) optimal process design and integration.

13.2 Microalgae Genetic Engineering

A gene is a basic physical unit of congenital characteristics that are made up of DNA, whereby they provide a recipe to make molecules of proteins to form a microalgae cell. A pathway in transcriptome sequencing for *Dunaliella tertiolecta* species discovered as important information of DNA where they able to build up and maintain an organism for yield optimum lipid production for microalgae (Rismani-Yazdi et al., 2011). The method for transcriptome is through link starch metabolism to obtain a higher yield of ethanol fermentation by the glycolysis pathway in the microalgae cell. There are 2 crucial types of nucleic acids to form a complete set of genomes, RNA and DNA, where current research is focused on the identified genome network and modified the gene toward the accumulation of lipids, polysaccharides, hydrocarbon in the microalgae cell (Radakovits et al., 2010; Ryan Georgianna and Mayfield, 2012). The genome sequencing is the key where allow scientist to understand the working mechanism of the gene related to the superior growth, besides, it increased the nutrient compounds and revolute to adapt different environmental condition within the cytoplasm. The project worth to be highlighted such as involving genome sequencing on microalgae obtained the chloroplast expression from *Chlamydomonas reinhardtii* (Shrager et al., 2003), identified novel genes for silicic acid transport (Armbrust et al., 2004), electrotransformation on plasmid gene plG121-Hm *Chlorella vulgaris* (Chow and Tung, 1999), and transient expression of luciferase in protoplasts for *Chlorella ellipsoidea* (Jarvis and Brown, 1991). The sequencer in which generate the order for proceeding the short range of genome fragments exclusive found

in The Genome On Line Database (GOLD) information on microalgae species genome and metagenome project based in worldwide. The multiply or overlapping of all the fragments have formed in the genome, such process known as *de novo* genome set-up (Giani et al., 2020; Pagani et al., 2012). Most genetic modification using knockout and knockdown method for constructing cell with a quality physiological performance which is currently rapidly developed for diatoms and other algae for industrial applications such as biofuel (Bogen et al., 2013; Bwapwa et al., 2019, 2018b; Radakovits et al., 2010). The genome sequence is progressing toward utilizing AI analytics which can assist in optimization, detection, and alteration of gene expression.

13.2.1 Next-Generation Sequencing and AI in Microalgae

Recently, integrated next-generation sequencing (NGS) instruments that perform amplification, genomic sequencing, and data analysis have been applied in microalgae research (An et al., 2018). NGS is mainly classified as short read and long reads with their working principles, such as (i) sequencing by ligation (SBL), (ii) sequencing by synthesis (SBS) (ii) single-molecule real-time long reads, and (iii) synthetic long reads (Goodwin et al., 2016). In recent years, many NGS technologies are available for microalgae applications such as the 454 Genome Sequencer, Illumina technologies, ligation-based SOliD Genome Sequencer, semiconductor-based Ion Personal Genome Machine (PGM), and Ion Proton (Kim et al., 2014). Nevertheless, from the authors' best knowledge, the most popular and well-established NGS for microalgae applications are from Illumina (MiSeq and HiSeq) and Pacific Biosciences (PacBio) technologies (Table 13.1). The Illumina technologies have become the state-of-art for microalgae transcriptomic analysis (Rismani-Yazdi et al., 2012), *de novo* assemblies (Blanc-Mathieu et al., 2014), microbiome analysis (Carney et al., 2014) and various genome sequencing application. Moreover, researchers adapting the single molecules through real-time sequencing data (SMRT) PacBio system can provide continuous observation of DNA synthesis without steric hindrance (Eid et al., 2009). The advances of neglecting the steric hindrance where any of the chemical reaction is not affected by the DNA that has been created in the microalgae cell. Furthermore, the structure of chromosomes can be assisted to conformation image tested through capture Hi-C technology to better clarify the microalgae gene expression. There are reported genome creating single cell molecules sequencing through Hi-C technology using optical mapping to capture the formation of the chromosome (Zhang et al., 2018).

Table 13.1: Application of Next-Generation Sequencing on Microalgae

Species	**NGS Technology**	**Remarks**	**Reference**
Benthic microalgae	Illumina MiSeq	A specific primer targeting the rbcL gene was designed and used with MiSeq for diversity exploration of Benthic diatoms.	(An et al., 2018)
Dunaliella tertiolecta	Illumina MiSeq	Used random insertional mutagenesis to create a *D. tertiolecta* mutant D9 which was identified to give pathways with high lipid yield.	(Yao et al., 2015)
Chlamydomonas reinhardtii	Illumina HiSeq X validated with qRT-PCR	Identified 10,635 differentially expressed unigenes under salt stress in *C. reinhardtii*. Within these unigenes, 5920 and 4715 were up- and down-regulated, respectively.	(N. Wang et al., 2018)
Dunaliella salina, Dunaliellatertiolecta	Illumina MiSeq	Used RNA-seq and machine learning to perform meta-analysis models and identified the most important salt stress-responsive meta-genes.	(Panahi et al., 2019)

Table 13.1 (Cont. 1): Application of Next-Generation Sequencing on Microalgae

Species	**NGS Technology**	**Remarks**	**Reference**
***Dunaliella tertiolecta* LB 999**	Illumina MiSeq v.2	Concluded that the repression of *chlH* and *chlN* resulted in nitrogen depletion reduced photosynthesis. Reduced photorespiration was associated with the repression of *MTHFR*, *GGH*, and *metE*.	(Shin et al., 2015)
Polytomella piriformis, P. parva, P. capuana, P. magna	Illumina HiSeq 2500	Complete recovery of mitochondrial genome sequences.	(Tian and Smith, 2016)
Alexandrium tamarense, Cochlodinium polykrikoides	Oxford Nanopore MinION and Illumina MiSeq	Studied the bacterial associated with harmful microalgae. Concluded that Roseobacter clade was associated with *A. tamarense*, while *Gammaproteobacteria* and *Alphaproteobacteria* were associated with *C. polykrikoides*.	(Shin et al., 2018)
Tetraselmis striata	PacBio RS II	Recovered a high-quality draft genome sequence for the microalgae.	(Steadman Tyler et al., 2019)
Undaria pinnatifida	PacBio Sequel system and Hi-C	Recovered the first genome sequence of brown algae *U. pinnatifida* by assembly.	(Shan et al., 2020)

Alternatively, single-molecule real-time long reads devices were proposed by Oxford Nanopore Technology (ONT) by using nanopore principles for NGS (Clarke et al., 2009). Nanopore technology from ONT enabled USB-like microdevices (MinION) for NGS, which have high sequencing length however lowered accuracy compared to Platforms like BioPac (Mikheyev and Tin, 2014). Nevertheless, the cheap device pricing of Oxford Nanopore MinION gives it a competitive edge upon future improvements. The advancement of AI for the application of genome sequencing technique allows for rapid search of relevant information for microalgae genetics and pave ways for efficient gene editing as an important future development for microalgae. Such improvement can increase the compatibility of the microalgae gene in mutagenesis and recombinant DNA technology for improving microalgae nutrient compounds or physiology (Ryan Georgianna and Mayfield, 2012). As an example, the recombinant DNA technology can be introduced by the addition of acetyl-CoA carboxylase (ACCase) into random sites of the microalgae genome for species such as *Cyclotella cryptica* to simulate the lipid accumulation (Dunahay et al., 1996). The recombinant DNA method would need identified various species data for selecting potential microalgae gene expression such as without inherited disease, composition, or functions, characterizing mutations in which would provide the potential for synthesizing particularly targeted nutrients, and detect the disease epidemic affected by the inhibits cell.

The gene expression incorporated with AI algorithms can be used to assist, improve accuracy and automate the identification process (Chuai et al., 2018; Lin and Wong, 2018). The difference between conventional statistical method and machine learning (ML) methods for genome sequencing is that ML does not require full details about the sequencing measurements, while can extract a large number of features from the sequence (Bzdok et al., 2018). For example, cross-species RNA sequence can be analyzed by multiple ML models to give the meta-genes that affect certain microalga pathway (Panahi et al., 2019). ML methods (i.e. Logistic Regression, Random Forest, Classification Via Regression, LMT, Random Subspace) were also applied for the detection of differentially expressed genes (DEG) (L. Wang et al., 2018), while the use of Logistic Regression (with InfoGain feature selection) demonstrated the best accuracy by comparison. Alternatively, neural networks-based methods were also tested to out-perform conventional ML methods (i.e. PCA, pcaReduce, ICA, NMF, tSNE, SINCERA, SIMLR, and SNN-Cliq) for extracting useful information from RNA-seq via dimension reduction (Lin et al., 2017). A finite state machine was used to assist in classifying effective primers for the Polymerase chain reaction (PCR)-based sequencing reaction (Ashlock et al.,

2002). Also, metaheuristics (such as Genetic algorithm) was used for unique region searching in multiplex PCR primer design (Wu et al., 2007). Furthermore, the webtool "pcrEfficiency" was also proposed to predict the PCR efficiency for a given amplicon and primer pair (Mallona et al., 2011), showing potential for further development of AI-based efficiency prediction.

13.2.2 Genome Editing Enabling AI-based Bioinformatics

AI algorithms can be used to assist the process of genetics modifying tools such as RNAi, ZFNs, TALENs, and CRISPR-Cas9 to optimize the selectivity and yield of the microalgae nutrient compound as biomolecules. Suppression of lipid catabolism is a strategy for enhancing lipid accumulation, such as knocking down technique from RNAi for increased the photosynthetic H_2 production from *Chlamydomonas reinhardtii* (Oey et al., 2013). ZFN-mediated gene editing was used in *C. reinhardtii* for enhanced lipid production (Sizova et al., 2013), while TALENs involved nucleases was used for targeted genome editing on *Phaeodactylum tricornutum* to the nutrient compounds (Weyman et al., 2015). In addition, CRISPR-Cas9 was being reported as an effective and rapid tool to perform stable gene editing for microalgae *P. tricornutum* (Nymark et al., 2016). Other utilizations of genome editing applications on microalgae are listed in Table 13.2.

Table 13.2: Techniques and functions for genetic modifications

Species	Tools	Functions	Reference
Dunaliella salina	RNAi	Performed knockout and cloned sequence from the microalgae species by regulating the metabolite for the host physiology.	(Jia et al., 2009)
Chlamydomonas reinhardtii	RNAi	RNAi mediated for silencing chlorophyllide and oxygenase gene and used to study the deactivation function of genes in the organism cell.	(Perrine et al., 2012)
C. reinhardtii	ZFNs	Applied to target COP3 gene and changed the physical expression.	(Mussgnug, 2015)
Nannochloropsis oceanica IMET1	ZFNs	Assists chloroplast transformation and mutagenized to controls the formation of uric acid.	(Kwon et al., 2018)
Phaeodactylum tricornutum	TALENs	Bind in the right orientation and proximity into DNA targeting PtAureala gene, which is a photoreceptor for accepting blue-light (Aureochrome) and can control the cell colonization.	(Serif et al., 2017)
C. reinhardtii	TALENs	Double strand break dTALE-induced activation targeting ARS1 / ARS2 with protein activity being elevated. Besides, may enrichment the nutrient compound for the host cells.	(Gao et al., 2014)
C. reinhardtii **CC-124**	CRISPR-Cas9	CRISPR-Cas 9 causes strain sequence with more tolerance and consistency growth compared to RNAi. CRISPR is more specific on a few amino acids rather than causing the whole frame of sequence shift.	(Shin et al., 2016)
Nannochloropsis oceanica **CCMP1779**	CRISPR-Cas9	The organism is a model for high lipid metabolism which utilized both Cas9 and ribozyme sgRNA. The system may auto-generate targeted mutations.	(Poliner et al., 2018)

The ribonucleic acids interfere (RNAi) as tools for genetics modifying through gene silencing as mechanism development (Figure 13.1(a)) for functional genomics, therapeutic purposes, agriculture, and medications. Microalgae species of *Nannochloropsis* strain using RNAi as reverse genetics tools using knockdown methods for constructs reverse repeat sequences from the selected gene (Wei et al., 2017). The RNAi is mostly used for therapies treatment in targeted tissue for combating inhibits cell (Majumdar et al., 2017). Moreover, *Chlamydomonas sp*. are used for studying unknown phenotypes by suppressing expressions of Rubisco multigene by inverted repeat (IR) gene sequence (Rohr et al., 2004). AI algorithms can assist in performing preselection (during screening) for an optimum combination of different species fragment for the purpose of alteration in physiological and enhancing the natural immunity by altering the metabolism and reproductive system (Neve et al., 2020). For example, a radial basis function support vector machine (SVM) was used to carry out genome-wide RNAi screen with time-lapse imaging and allowed for automatic identification and clustering of chromosome phenotypes (Walter et al., 2010). An iterative ML approach utilizing the GentleBoosting algorithm on regression stumps was also used to screen cellular morphologies via RNAi (Jones et al., 2009). AI approach produced high-quality feature extraction, which is desirable for high-content imaging in RNAi screens. Moreover, the efficacy of anti-sense or siRNA oligonucleotides in RNAi was predicted using boosted genetic programming (Sætrom, 2004), showing better sensitivity than conventional SVMs.

Targeted specific DNA cleavage using Zinc-finger Nucleases (ZFNs) are used for double-strand break (DSB) aim for cellular DNA repair procedure in microalgae cell to alter and cope with the bacteria or fungi during pre-growth for the microalgae cell (Figure 13.1(b)). As ZFNs relies on enzyme produced by zinc finger protein structure, the process of obtaining the new design and selection for the specific genome is challenging. There are reported studies from *Chlamydomonas reinhardtii* as a versatile model for modification of the nuclease domain through ZFNs (Sizova et al., 2013). It allows for the study of the nucleotide cassettes independently express a targeted genetic element from the genome side to prevent endogenous chromosomal allelic locus (Palpant and Dudzinski, 2013). The incorporation of AI algorithm can be used to predict the compatibility and perform selection before identifying a novel gene or combination of the gene which produces undesirable cell sequences such as low yield, short-living duration, and unable to resist the inhibits from bacteria or fungi cell. There are also cases for using zinc-finger nucleases ZFNs gene-editing for *Chlamydomonas* strain for encoded with

Staphylococcus aureus and *Streptococcus pyogenes* to characterize the photoreceptors physiology studies (Carroll, 2011; Greiner et al., 2017). An ensemble micro neural network (available online: http://web.iitd.ac.in/~sundar/ZifNN/Home.html) was developed for predicting the activity between ZFN and their target DNA (Dutta et al., 2016). Moreover, position weight matrix, hidden Markov model, SVM (Persikov et al., 2009) (available online: http://compbio.cs.princeton.edu/zf) and ANN are used to predict the DNA-binding specificity for ZFN (Roy et al., 2012).

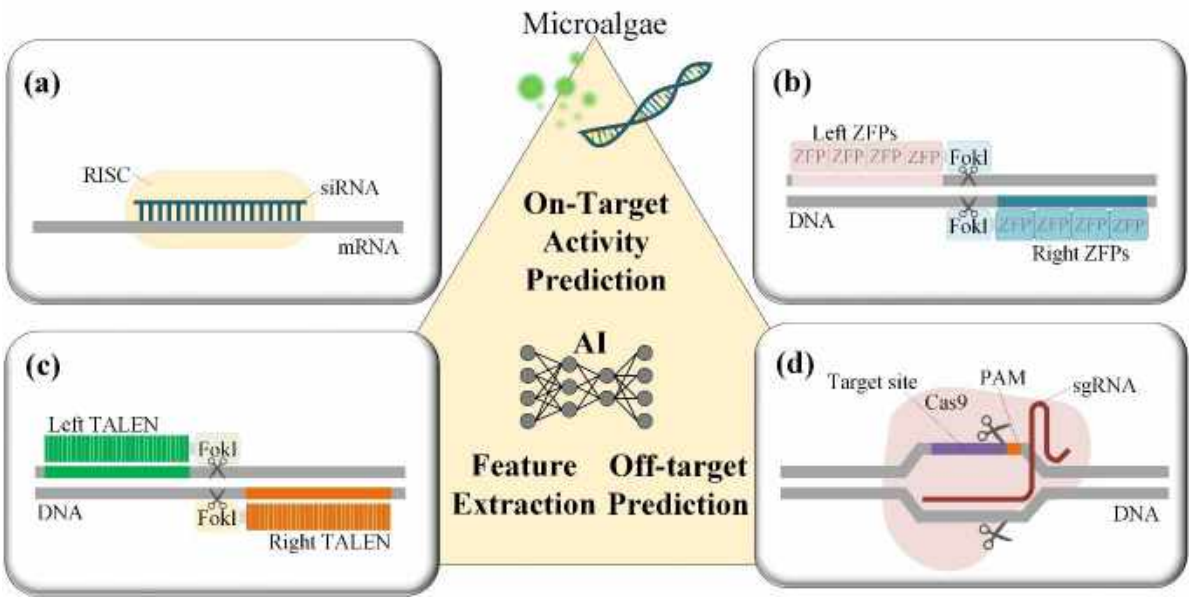


Figure 13.1: AI application and mechanism of gene editing technique (a) RNAi (b) ZFN. PAM: Protospacer adjacent motif. sgRNA: Single guide RNA. RISC: RNA-induced silencing complex. siRNA: Small interfering RNA. mRNA: messenger RNA. ZFP: Zinc finger protein. (c) TALENs (d) CRISPR/CAS9

Transcription activator-like nucleases (TALENs) can act as a faster genome editing substitute (Figure 13.1(c)) compared to ZFN for specifically TALE-derived amino- and carboxy domains design (Joung and Sander, 2013). The transcription activator-like effector nucleases (TALENs) method have more advantage compare to ZFN, it used TALENs by knockout urease protein on *Phaeodactylum tricornutum* gene to reduce the growth of urease in the urea cycle (Joung and Sander, 2013; Weyman et al., 2015). Besides, TALENs are associated with a longer sequence and endonuclease by reactions of dimerization for cleavage. The technique of TALENs has been used in *Chlamydomonas sp.* and *Chlorella sp.* for encoding the acyl-CoA synthetases. This was able to improve the production of fatty acids within the cell can be further fine-tuned for other biotechnological applications (Ng et al., 2017; Oey et al., 2016). The SVM algorithm

can be used to analyze the cellular mechanism via TALENs and total internal reflection fluorescence (TIRF) microscope (Hong et al., 2015). An ML workflow with TALENs was also carried out using the WEKA software package (which considers SVMs, decision trees, Naïve Bayes, etc.) to improve on-target DNA-cleavage specificity.

The clustered regularly interspaced short palindromic repeats (CRISPR) having a pliable immune ability to facilitate genome engineering (Figure 13.1(d)) in the microalgae genome (Patron, 2020). The CRISPR-Cas9 system provides adaptive resistance to host cell for foreign genetic elements being invaded. The most updated tools provides short duration, low cost, high accuracy, and improve the efficient compare to other existing genome editing tools (Doudna and Charpentier, 2014). The resistance function from CRISPR has been developed and applied to *Coccomyxa sp.* which can genetically improve and successfully isolated the genome-edited strains with advances for lipid productivity (Yoshimitsu et al., 2018). The machine learning assisted gene-editing tools can provide off-target predictions (Lin and Wong, 2018), enhance the double-stranded break (DSB) procedure, and optimize guide RNA in CRISPR techniques (Chuai et al., 2018). This is particularly suitable microalgae, as CRISPR/Cas9 has been subject to extensively use for genetic modification in the field of biofuel production (Jeon et al., 2017). For example, *Phaeodactylum tricornutum* cells acted as a targeted RNA segment for the CRISPR to resist the undesired DNA from invading cells to generate stable targeted gene mutations (Nymark et al., 2016). Moreover, Cas-9 can promote genome editing through stimulating double-stranded break (DSB). CRISPR/Cas9 mediated of CpSRP54 gene has been reported successfully knockout using biolistic bombardment by sgRNA (Feng et al., 2013; Shan et al., 2013). In this aspect, AI algorithm can help to predict the conventional pathway for CRISPR-Cas9 for repairing the DNA damage chains, which is error-prone for non-homologous end joining (NHEJ) or homology-directed repair HDR during microalgae or any plant sequence editing pathway (Ran et al., 2013). The photosynthetic rate of eukaryote based on microalgae cell of *Chlamydomonas reinhardtii* was found to be increased by 54 % through two-gene knockout by editing the *Chl* antenna size of *ΔZEP/ΔCpFTSY* (Baek et al., 2016). In short, AI algorithms act in a technology-agnostic manner (i.e. same for all gene editing technology) and can be transferred across different gene editing platforms. The main uses of AI algorithm for gene editing technologies are on-target prediction, off-target prediction, and feature extraction. Nevertheless, with more genetic data on microalgae, the utilization of AI algorithm can potentially accelerate the discovery of functionality of microalgae from their gene (such as increased lipid, improved biomass yield, etc.)

13.2.3 Tools, Databases, and Intelligent Methods for Microalgae Bioinformatics

Many microalgae genomes have already been sequenced and annotated in the public database such as NCBI database (Wheeler, 2006), UniProt (Universal Protein Resource) database (Bateman et al., 2015), ExPASy database (Artimo et al., 2012), BRENDA (Braunschweig Enzyme Database) (Schomburg et al., 2013). Moreover, AlgaeBase (Guiry et al., 2014) is also an algae specific database providing key references (e.g. Genbank), microscope photos, and other useful information for 158,964 species and infraspecific names of algae. Amino acid sequence alignment tools can then be carried out by one-to-one recognition (e.g. BLAST, HMMER) or one-to-many recognition (e.g. InterProScan, FFPred2, Argot) strategies (Reijnders et al., 2014). One-to-many alignment tools are generally slower but have a high success rate, hence, one-to-one recognitions such as BLAST-based methods are commonly carried out first.

Genome-scale metabolic reconstruction is generally a difficult task either requiring a reference database for metabolism or experimental validation (Francke et al., 2005). There are also comprehensive databases containing metabolites, reactions, and pathway for example, the KEGG database (Ogata et al., 1999) and the MetaCyc family (Karp, 2002). Methods for metabolic modeling are categorized by isotope labeling, kinetic, and constraint-based (Čuperlović-Culf, 2013). For the metabolic modeling of microalgae, flux balance analysis (FBA), dynamic FBA, unsteady-state FBA, and elementary modes (EM) are commonly carried out (Tibocha-Bonilla et al., 2018). Nevertheless, the mapping of metabolic models toward the commercial desirability of microalgae is still a challenge. A good starting point is by defining a biomass objective function (BOF) (Feist and Palsson, 2010). Still, the commercial objective is highly dependent on microalgae species screening, commercial-oriented experiments, industrial applications, system, and process of microalgae (Figure 13.2(a)). For example, the commercially desired objective may be the carbon distribution of the algae-based oil (Yang et al., 2016), which does not have a direct indicator in the metabolic model.

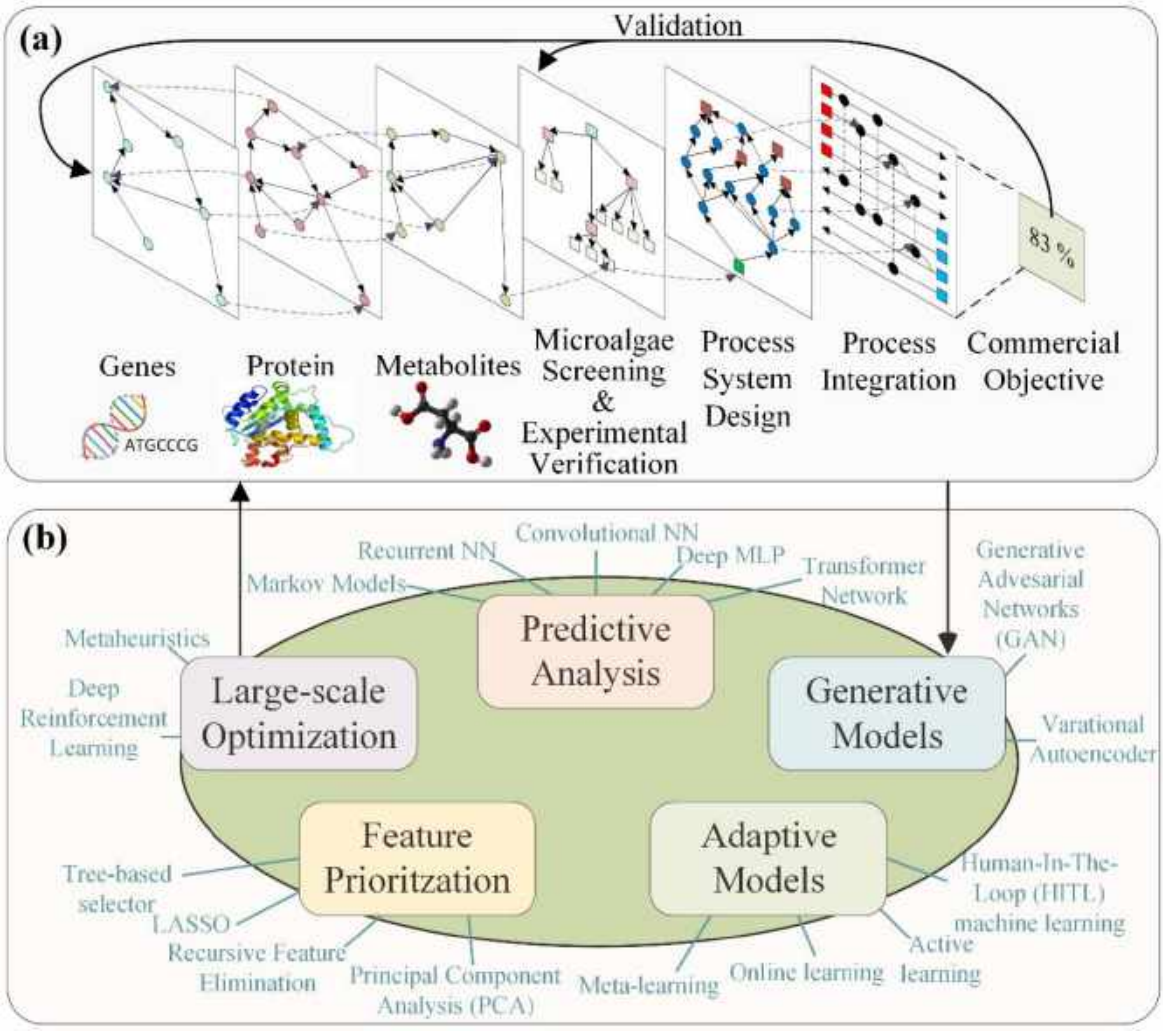


Figure 13.2: Schematic diagram for (a) multi-omics approach for microalgae considering process and commercial objectives (b) application of AI algorithms in the microalgae research pipeline

For biochemical networks studies, Yugi et al. (2016) proposed the concept of using a "trans-omic" network to consider integrative effects of genome, transcriptome, proteome, and metabolome. This opened the doors to a paradigm of machine learning workflow (Camacho et al., 2018), which allows for data-driven exploratory analysis, prediction, classification, and validation (Figure 13.2(b)). For example, DeepBind was developed to predict general sequence specificities in proteins using a hyper-parameter optimized convolutional neural network (Alipanahi et al., 2015). DeepBind can be directly used for general sequence specificities prediction in microalgae and can be potentially fine-tuned on more microalgae data in the future. Protein functionality was also predicted using a multi-task neural network (Sureyya Rifaioglu et al., 2019). NetSurfP 2.0 was also developed and openly accessible on the web for local structural feature prediction in protein (Klausen et al., 2019), which uses a bi-directional

LSTM network. The prediction of protein-protein interaction was also carried out using a stacked autoencoder (Sun et al., 2017) achieving an average accuracy of 97.19 %. For metabolic pathway prediction, machine learning (ML) algorithms (which includes decision tree, Naïve Bayes, and random forest) were studied and gave a slight increase in prediction with respect to PathoLogic prediction (Dale et al., 2010). The work also concluded that ML methods were able to dissect and display the evidence for the presence of a pathway. More recently, A meta-learning approach deployed using the tree-based pipeline optimization (TPOT) framework to predict dynamic metabolic pathways from time-series multi-omics data (Costello and Martin, 2018). The model obtained better prediction than the traditional kinetic modeling approach while providing faster development of pathway.

13.3 AI in Experimental Analysis of Microalgae

For the experimental analysis of microalgae, this paper focuses on the experiments for the purpose of developing commercial products. Therefore, the subchapters below will focus on the application of AI algorithms on (i) strain-species identification, (ii) screening of microalgae species and strains, (iii) microalgae cultivation, and (iv) microalgae conversion.

13.3.1 Strain-species Identification by Computer Vision and AI

The AI approaches were applied in three areas. The first area coupled AI with microscope image processing resulted to a fast species identification method. The second area was the combination of AI tools with fluorescence measurement-based methods aimed to quantify cell concentration in microalgal culture (Liu et al., 2020). While the third area focused on AI and matrix-assisted laser desorption/ionization (MALDI)-based spectrometer to characterize microalgae from their mass spectra (Andrade et al., 2015; Barbano et al., 2015). The species identification was developed mainly for ecological monitoring of water ecosystems such as monitoring of the cyanobacteria blooms or taxonomic classification of phytoplankton (Coltelli et al., 2017; Luo et al., 2011; Shin et al., 2000; Sosik and Olson, 2007). The identification of microalgae in biotechnology research is focused mainly on the identification of morphotypes. Giraldo-Zuluaga et al. (2018) developed a methodology for automatic identification of *Scenedesmus* coenobia based on microscope image processing using the ANN and SVM as a classification technique for the final step of pattern recognition. *Scenedesmus* coenobia was recognized with accuracies of 98.63 % and 97.32 %. Similarly, Hayashi et al. (2018) presented a convolutional neural network (CNN)-based automatic classifier of morphotypes of microalga

Cyanidioschyzon merolae for the measurement of division rate as a parameter of quality controlling in cultivation. The division rate was expressed as the number of cells in the interphase and mitotic phase in the algal culture. The number of cells in the interphase and mitotic phase was analyzed based on phase-contrast images processing using CNN. Compared to experimental results the two-layer function model achieved an accuracy of 92 %. Recently, an instance segmentation for algae identification in water samples was proposed by using a Mask-RCNN (Region-based Convolutional Neural Network) model (Ruiz-Santaquiteria et al., 2020). From microscopic photos of various mixed species of algae in water, Mask-RCNN achieved 85 % average precision, 86 % sensitivity, and 91 % specificity. Furthermore, the combination of matrix-assisted laser desorption/ionization (MALDI)-based mass spectrometers and self-organizing maps (SOM) was also combined to provide a latent (mosaic image) profile for various species of algae microorganisms (Wirth et al., 2012). The SOM profile showed that the meta-features of each microalgae strains were distinct and could identify infectious algae using a random forest classifier. In general, AI-based algorithms have a large potential to speed up and automate the task of strain-species identification of microalgae not only via microscopic images but also using spectroscopic analysis.

13.3.2 AI-assisted Screening of Microalgae Species and Strains

Species and strains selection is a crucial step for biotechnological production using microalgae (Barclay and Apt, 2013; Borowitzka, 2013; Gnouma et al., 2018). Strains are selected based on cultivation aims especially for commercial applications in feed, nutrition, or cosmetics. The strains are evaluated in terms of growth rate, biomass production or content of high-value products or pollutants removal (Gnouma et al., 2018; Griffiths and Harrison, 2009). The species-appropriate for commercial applications should be characterized by high biomass production to achieve sustainable and cost-effective biotechnology production (Christenson and Sims, 2011; Rodolfi et al., 2009).

Diverse measurement methods were used for biomass determination included simple conventional methods like dry cell weight or manual cell count (Sarrafzadeh et al., 2015). The development of advanced measurement methods is important for the application of new approaches in rapidly evolving biotechnologies. The fluorescence measurement-based methods allow easy and fast measurements of many samples resulted in the generation of large datasets suitable for applications of AI techniques. However, works focus on the biomass estimation for cultivation purposes using AI methods are sporadic (Franco et al., 2019; Liu et

al., 2020). Liu et al. (2020) tested the feasibility of the application of artificial neural networks (ANN) for precise monitoring of algal density using a single fluorescence emission spectrum measurement combined with back-propagation ANN Optimized by Genetic Algorithms (GA-BP model). The developed model was used to predict the cell concentration because there is a nonlinear relationship between the measured fluorescence emission spectra used as input and the algal cell concentration used as output. Results of the study has shown high potential of the usage of the in-situ fluorescence spectrometry in combination with an ANN for estimation of algal cell concentration. However, it is necessary the further reduce of prediction error. Other usage of ANN was carried out (Franco et al., 2019) for differentiation between monoalgal and mixed algal cultures to detect the presence of contaminants in monoalgal cultures. The studies were based on the measurement of light absorption spectra of the monoalgal cultures of green algae (*Chlorella vulgaris* and *Scenedesmus almeriensis*) and cyanobacteria (*Nostoc sp.* and *Spirulina platensis*) and mixed algal cultures combined mentioned species. Each monoalgal and mixed algal culture was measured 25 times. A total of 550 measurements created a large dataset for the development of ANNs. The study concluded that the methods based on measurement of light absorption spectra can be used for the development of ANN model that distinguished monoalgal culture from a mixed algal culture with the identification of prevailing strains. Hence the cultures were necessary to maintain under similar cultivation conditions. Alternatively, selective breeding can increase the yield of the cell nutrient compound and suit with specific cultivation environments such as brackish water, high salinities water, and municipals wastewater (Takouridis et al., 2015). Selective breeding shuffling strategy is to employ the initial strain toward growth in stress conditions or specific medium (Takouridis et al., 2015). The strain for *Chlamydomonas reinhardtii* has a high interest in genetic engineering due to having established reliable genetic tools (Rasala et al., 2014).

13.3.3 Data-driven Acceleration of Microalgae Cultivation

Microalgal cultivation represents the transfer of microalgae from the natural environment to an artificial environment with uncontrolled or controlled conditions. The search for the appropriate cultivation conditions is the necessary step for the successful cultivation of microalgae. The essential environmental parameters affected the growth of microalgae are light, temperature, nutrients included macronutrient, micronutrients, and availability of CO_2 through the aeration of microalgal culture (Breuer et al., 2013; Krzemińska et al., 2015). The light is an essential factor for photosynthesis. The relationship between algae growth and light intensity

is expressed by a "light response curve" which is the response of the photosynthetic growth cells to increasing light intensity (Leverenz et al., 1990; Torzillo and Vonshak, 2013). The temperature of the environment affects biochemical reactions in the cells resulted in changes in the contents of chemical compounds in the cells (Hu, 2013). A temperature lower than physiological optimum causes the increase of unsaturated lipid content in membranes and increasing of enzymes production as an adaptation for maintenance of photosynthetic processes (Nishida and Murata, 1996; Thompson et al., 1992). Higher temperatures can induce the production of carotenoids such as astaxanthin or β-carotene (Liu and Lee, 2000). Microalgal growth is also affected by nutrients contained in cultivation mediums. The essential macronutrients are mainly carbon (organic or inorganic), nitrogen and phosphorus which are limited for microalgal growth (Hu, 2013; Pedersen and Borum, 1996). The crucial micronutrient is iron which is essential for photosynthetic processes, respiration, and DNA synthesis. Iron limitation resulted in a decrease in photosynthetic activity (Bottin and Lagoutte, 1992; McKay et al., 1999).

The efforts for the maximization of biotechnology production lead to the development of techniques for the optimization of cultivation technologies. Optimization of cultivation processes with multiple variables can be performed by statistical methods for example response surface methodology (Cao et al., 2020). However, the employing of the methods of AI such as the ANN and genetic algorithms brings strong tools for modeling nonlinear systems without detailed kinetic knowledge of the system (Nayak et al., 2018; Wu and Shi, 2007).

The ANNs were employed for studies of culture medium composition and its effect on microalgal growth (García-Camacho et al., 2016; López-Rosales et al., 2013; Morowvat and Ghasemi, 2016; Wu and Shi, 2007). Wu and Shi (2007) proposed a simple hybrid neural network model for fed-batch heterotrophic cultivation (i.e. without light and utilizing organic carbon source) of *Chlorella* sp. The model was based on the glucose concentration as the input parameter and the specific growth rate as the output. The hybrid neural network model with one input (glucose concentration) estimated the experimental results well and the model was evaluated as a sufficient tool for optimization of heterotrophic cultivation. The multilayered feed-forward model was designed for the optimization of the cultivation medium to increase lipid content in *Chlorella vulgaris* biomass (Morowvat and Ghasemi, 2016). The input parameters were the glucose, nitrate, and phosphate concentrations in the cultivation medium. The minimum amount of required hidden layers for the final model was 10 neurons. The

created model for the determination of optimum conditions leads to the production of lipid-rich biomass produced during *Chlorella vulgaris* cultivation.

Multi-nutrient interactions and the effect on microalgal growth using modeling were studied by López-Rosales et al. (2013) and García-Camacho et al. (2016). In both works, the feed-forward back-propagation neural network (FBN) was used for modeling of the growth dynamic variations as the response of different concentrations of macronutrients and micronutrients in cultivation medium for dinoflagellate. Raw data from 500 respectively 426 batch experiments was used as the input. Resulting models sufficiently predicted algal growth for an extremely broad range of nutrients concentrations in cultivation media with many components (García-Camacho et al., 2016; López-Rosales et al., 2013). Moreover, the model used a wide range of micronutrients, presenting valuable information on micronutrient influence on microalgal growth (García-Camacho et al., 2016).

While the above-mentioned works dealt exclusively with the prediction of the influence of nutrients in cultivation media, Hu et al. (2008) have been proposed the ANN model for control of light intensity at the photobioreactor. The controlled photobioreactor system was regarded as a single-input (light intensity) and a single-output (biomass growth) system. The study demonstrated that ANN-model predictive control can predict microalgal biomass growth under different light intensities with sufficient accuracy. An ANN model for prediction of biomass production in outdoor culture was described in work of Pappu et al. (2013) and (Supriyanto et al., 2019) that used a wider range of inputs (mainly biomass concentration, nitrates concentration, pH, light intensity, etc.) compared to Hu et al. (2008). Pappu et al. (2013) found that datasets of 12 days (one measurement per day) were sufficient to develop a predicting model for *Spirulina* production in outdoor culture. The more complex model was developed by Supriyanto et al. (2019) that used a total of 553 datasets for the design of the BP-ANN model consisted of eight input neurons, eleven hidden neurons, and one output neuron. The proposed ANN model showed precise prediction with a correlation coefficient higher than 0.9 using all cultivation parameters. The correlation of the output was still high (R^2=0.9382) even five cultivation parameters were used.

A predictive model for optimization of lutein production was proposed by Del Rio-Chanona et al. (2019) used data-driven modeling techniques for the study of the dynamic effect of light intensity and nitrate concentration on lutein production. The work was aimed to compare the

efficiency of physics-based modeling and data-driven modeling for increased microalgal lutein production. The data-driven model-based cultivation strategies have shown an increase of 40-50 % in lutein production, while only one of the physical-based model strategies leads to higher lutein production (increase 35 %).

Only sporadic works are deal with the optimization of harvesting strategy. Zenooz et al. (2017) compared different neural network architecture for the modeling of microalgae flocculation. The best prediction of microalgal flocculation was achieved by neural network architectures of a multilayered perceptron. Susanna et al. (2019) proposed the nonlinear autoregressive multilayer perception model for prediction of microalgal growth resulted in maximizing the productivity of *Spirulina platensis* in tubular photobioreactor by finding the balance between growth and harvest cycle. The input parameters for the developing model were time (days), temperature, pH, dissolved oxygen, nitrogen concentration, and initial biomass concentration. The resulting model predicted the growth of *Spirulina* up to three days with a coefficient of determination > 0.94. Range of cell density between 0.3 and 0.6 OD was determined as optimum for biomass productivity which increased by about 14 %.

Microalgal biotechnologies cover several areas. Above all, it is commercial production for nutrition, cosmetics and feed (Becker, 2017). However, the algal biotechnology can be used for environmental applications such as wastewater treatment or CO_2 sequestration (Sukačová et al., 2017). In this area, the ANN model has been developed for the prediction of water decontamination (Khataee et al., 2010, 2009). Khataee et al. (Khataee et al., 2010, 2009) applied the three-layered FBN for modeling the removal of the textile dye using microalgae (*Chlamydomonas* sp., *Chlorella* sp., *Cosmarium* sp., *Euglena* sp.). The results showed that the model can describe the process of dye removal. Moreover, it was found that the biological decontamination efficiency was dependent on the reaction time, initial dye concentration, algal concentration, pH, and temperature. Nayak et al. (2018) tested different neural network architecture for improving microalgal biomass production with integrated flue gas CO_2 sequestration and wastewater utilization as a source of nutrients. Optimized parameters resulted to increase of CO_2 sequestration about 57 %. Modeling using ANN was used sporadically in works focus on wastewater treatment and CO_2 sequestration, however, presented works have shown high potential for prediction of processes connected with wastewater treatment and CO_2 sequestration.

13.3.4 Fine-Tuning Microalgae Conversion Technologies using AI Algorithms

The experimental investigation of the conversion of microalgae into useful products requires a large amount of time and effort from researchers. For the conversion of microalgae, many technologies are studied such as thermochemical methods (i.e. gasification, pyrolysis, liquefaction, hydrogenation), biochemical method (i.e. transesterification, fermentation) and extraction methods using physical or chemical approaches (Suali and Sarbatly, 2012). As microalgae are the organism with multiple chemical composition, their conversion mechanism is not easily predicted (Teng et al., 2019). The most common application for AI in the fields of experimental conversion is for performance prediction and optimal condition determination (Table 13.3).

Table 13.3: Recent application of AI algorithms for improving microalgae conversion technology

Feed/Species	**Conversion Technology**	**Algorithm**	**Remarks**
Algal Mat	Pyrolysis	Single-layer ANN	Prediction of pyrolysis behavior (Mayol et al., 2019).
***C. vulgaris, N. oceanica, Chlamydomonas* sp.**	Pyrolysis	Particle swarm optimization and independent parallel reaction model	Prediction of microalgae pyrolysis kinetics with consideration of carbohydrates, protein, and lipids (Chen et al., 2018).
C. vulgaris	Pyrolysis and Gasification	Neuro-evolution with deep neural networks	Prediction of thermal conversion and searching for optimal conditions with minimal energy consumptions (Teng et al., 2019).
***Spirulina* sp.**	**Combustion**	Single-layer ANN and numerical methods	Predict the combustion performance, combustion parameters, and emission from inlet conditions and blends of microalgae-based diesel (Salam and Verma, 2019).

Table 13.3 (Cont.): Recent application of AI algorithms for improving microalgae conversion technology

Feed/Species	**Conversion Technology**	**Algorithm**	**Remarks**
Nannochloropsis oculata	Hydrothermal liquefaction (HTL)	Multiple linear component additivity model	Predict HTL conversion of product yield and quality (Leow et al., 2015).
***Chlorella* CG12**	Transesterification (supercritical methanol)	Response surface methodology (RSM), ANN and genetic algorithm	Prediction and optimization of process condition in supercritical methanol transesterification of microalgae oil to biodiesel (Srivastava et al., 2018).
Jatropha-algae	Transesterification (KOH-catalyzed)	Neuro-fuzzy inference system and RSM	Prediction of transesterification performance from catalyst loading, temperature, time, and oil mix of jatropha-algae oil (Kumar et al., 2018).
***Chlorella* sp.**	Ultrasonic-assisted Transesterification	Single-layered ANN with RSM.	Prediction of ultrasonic power, reaction time, methanol ratio, and chloroma ratio to optimize FAME content and exergy efficiency (Karimi, 2017).
Mixed microalgal biomass	Enzymatic Hydrolysis	Single-layered ANN	Prediction of reducing sugar concentration from substrate concentration, temperature, pH, and retention time (Shokrkar et al., 2017).

At the current state, AI algorithms have allowed experimental conversion to produce high accuracy prediction and optimal conditions under uncertainty (Chen et al., 2018; Teng et al., 2019). However, the improvement of highly time-dependent conversion technology such as fermentation (Hong et al., 1996) and anaerobic digestion via AI algorithms should be revisited. With advanced neural network architectures that can effectively model temporal effects (such as long-short term memory (LSTM) networks, a CNN for temporal modeling (McCormick and Villa, 2019), etc.), dynamic microalgae conversion technologies can be explored optimally. Moreover, the consideration of microalgae "signatures" (i.e. strains, genomes, microstructure, etc.) should be included in AI algorithms since these aspects can greatly affect the performance of microalgae (Lin et al., 2019; Wolf et al., 2015). These aspect can potentially allow the AI algorithm to unravel the basic working principle (such as pathway and evolution) for microalgae (Mutwil, 2020) in terms of their performance.

13.4 Intelligent System Design and Control for Microalgae

The design of effective cultivation systems and photobioreactor is an important topic as the system serves as the housing for microalgae growth. The modes of microalgae cultivation include phototrophic cultivation, heterotrophic cultivation, mixotrophic cultivation, and photoheterotrophic cultivation, requiring different photobioreactor design (Chen et al., 2011). In general, open and closed systems are preliminarily chosen and the consideration will depend on contamination risk, CO_2 losses, evaporation losses, light use efficiency, area-to-volume ratio, total area, process control, biomass productivities, investment costs, harvesting costs and scale-up difficulty (Xu et al., 2009). Optimization of microalgae cultivation is a non-direct task, as microalgae growth is transient, and thus commonly dynamic optimization approaches are required (He et al., 2012). For the optimization of microalgae systems, important factors such as light intensity, temperature, pH, mixing, light penetration, gas injection, and fluid mechanics must be evaluated (Bitog et al., 2011). Conventionally, hydrodynamics and optimization of the photobioreactor geometry can be carried out using computational fluid dynamics (CFD). The effects of temperature (Pires et al., 2017) and light intensity can be considered using correlations and physical models (Slegers et al., 2011). Mass growth and transfer in microalgae systems are usually modeled using kinetic equations, such as Monod-based models (He et al., 2012). Nevertheless, it was demonstrated that the combination of agent-based modeling and Photosynthesis Factory (PSF) has the potential to carry out 3D growth kinetics (Husselmann and Hawick, 2013), sped up by graphical processing units (GPUs).

Different microalgae strains lead to a varied growth rate and optimal conditions for design and control (Figure 13.3). For example, *Scenedesmus almeriensis* has a high growth rate and optimal temperature of 35°C and 8 pH, requiring the use of Sequential Quadratic Programming (SQP) for optimization of the dynamic model of its raceway reactor. Rather than exact algorithms, nature-inspired metaheuristic algorithms (such as genetic algorithm, particle swarm optimization, differential evolution, etc.) can also provide faster optimization speed for the optimization of microalgae reactors (Hernández-Pérez et al., 2019). Multi-objective optimization of *Dunaliella tertiolecta* was also carried out using an elitist non-dominated sorting genetic algorithm (NSGA-II) and found that there are three important objectives (ie. consideration of biomass, lipids, and cultivation time/cost). Bioreactors can also be optimized using many other metaheuristic algorithms such as the backtracking search algorithm (Mohd Zain et al., 2018). More recently, real-time optimization of batch bioprocess can also be carried out by reinforcement learning algorithms such as policy gradients (Petsagkourakis et al., 2020). Furthermore, the evaluation of reactor performance (such as chemical oxygen demand (COD) in wastewater) can be carried out using ANN, thus validating the optimality of design (Yazdani et al., 2020).

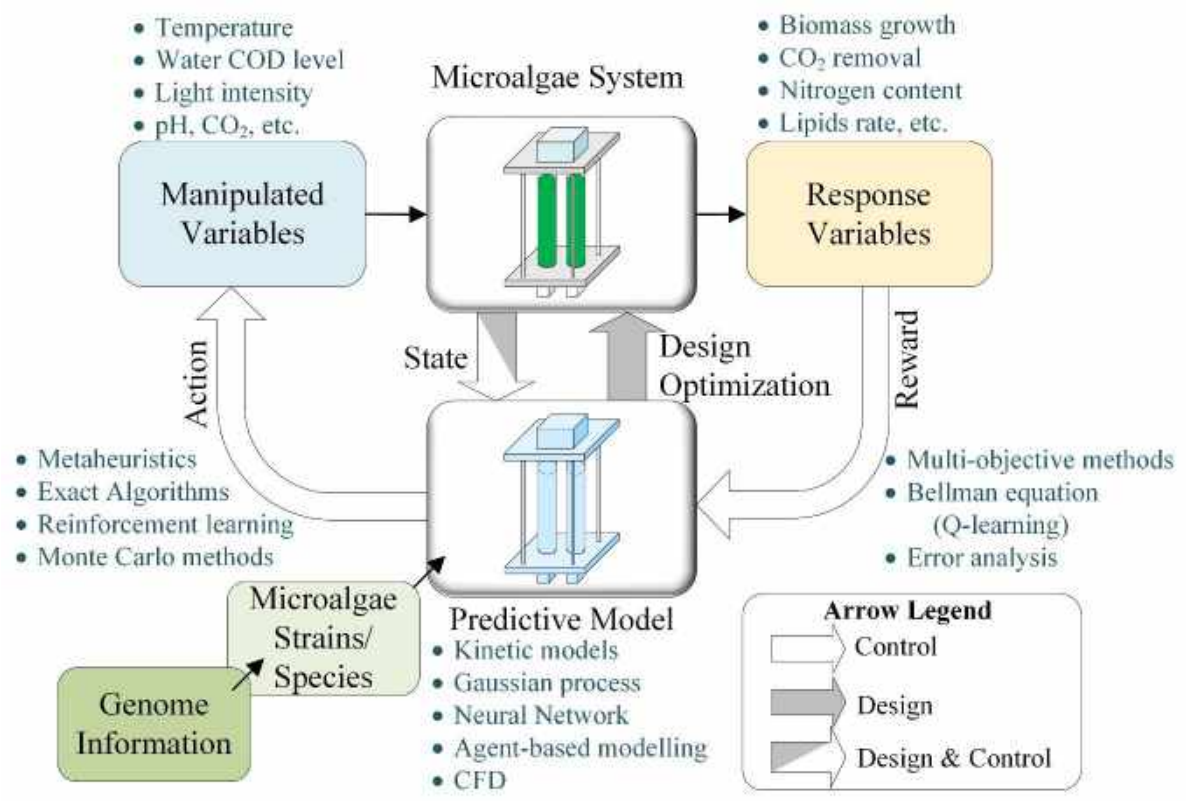


Figure 13.3: Methods and techniques for control and design of microalgae systems with consideration of microalgae genome and strains

Another useful application for AI in the microalgae system is for predictive control of the system. The main challenge for microalgae model predictive control (MPC) is in constructing the dynamic model of microalgae. In the direction of microalgae dynamic modeling, there are various models that include the Droop model, photo-acclimation model, pigment dynamics model, Beer-Lambert-based light model, and other kinetic models (Bernard, 2011). These kinetic models have provided useful applications of the control and optimization of cultivation systems (Bernard et al., 2016). A moving horizon estimator (MHE) and MPC was used to optimize the lipid production rate of *Auxenochlorella protothecoides*, achieving a 36 % increase in lipid concentration. Nevertheless, kinetic models for biochemical systems will require parameter tuning to be validated in applications (Wilkinson et al., 2007). Alternatively, the dynamics of microalgae systems can be learned by an ANN and used for MPC. For this, light-algae bioreactor with a high reaction rate, biomass growth, CO_2 removal was achieved using an ANN-MPC model (Hu et al., 2008). More recently, the control of bioreactor can be achieved using reinforcement learning (RL) algorithms (such as Q-learning, partially supervised RL, etc.) to achieve low-error precision control (Pandian and Noel, 2018). Gaussian process-based nonlinear model predictive control (GP-NMPC) were also proposed for bioprocess to shift the computation complexity towards a Monte Carlo offline constraint tightening using back-offs (Bradford et al., 2020). This strategy gives robustness to system uncertainty and allows for more efficient online learning. Deep reinforcement learning (DRL) with a Deep Q-learning (DQN) model was also deployed to directly learn and optimize the output of co-culture bioreactors (Treloar et al., 2020). This DRL paradigm can also be used for the co-cultivation of microalgae species to improve yields under mixotrophic conditions (Rashid et al., 2019).

13.5 AI-enhanced Process Integration for Microalgae-based Biorefineries

Process integration is the study of process units using a holistic approach to minimize resource usages and waste (Smith, 2005). Microalgae technology is a suitable process debottlenecking unit (Figure 13.4(a) and 13.4(b)) within the fields of process integration as it is effective in providing industrial waste (i.e. CO_2 and wastewater) remediation while producing various valuable products (McGinn et al., 2011). Rizwan et al. (2015) optimized a conventional microalgae biorefinery with the consideration of residual microalgae for a single strain of *C. vulgaris*. The study found that the gross operating margin (GOM) was below the breakeven point, making the pure cultivation of *C. vulgaris* for the sole purpose of producing biofuel economic infeasible. An optimal process network for biodiesel production from integrated

cooking oil and algae feed (Martín and Grossmann, 2012) giving an excellent algae-based biodiesel production cost of 0.42$/gal and energy consumption of 1.94MJ/gal. The results suggested that process optimization, heat integration, and the consideration of multi-feed production is critical for microalgae-based process design. For the design and synthesis of *C. vulgaris* algae processing network to mitigate emission from natural gas combustion, mixed-integer nonlinear programming (MINLP) approach was deployed to obtain algae-based biodiesel with $7.02/gasoline gallon equivalent (GGE) and a global warming potential at 26.491 kg CO_2-eq/GGE (Gong and You, 2014). This showed that microalgae technology had to be simultaneously used for waste reduction and chemical production in an integrative manner to be economically attractive. A differential evolution-based metaheuristic algorithm (Hernández-Pérez et al., 2019) was used to optimize a microalgae-to-biodiesel process which reduced carbon emission by 47.23 % Looking forward, with more consideration of more possible technologies, microalgae strains, utility integration (Ng et al., 2012), microalgae system dynamics (García-Camacho et al., 2016) the optimization of microalgae process will require extra large-scale optimization and data analysis. For process integration in microalgae-based biorefineries, AI technologies can speed up optimization time (Toimil and Gómez, 2017), provide predictive analytics, uncover system dynamics and uncertainty (Ochoa-Estopier et al., 2013).

A widely available process with high potential for microalgae-system integration is the biogas anaerobic digestion (AD) process (Figure 13.4(c)). Within the biogas (AD) process, microalgae can be directly fed as biomass input to the AD process while utilizing the CO_2 gas which is either separated from the biogas (Sialve et al., 2009) or the combusted flue gas (Zamalloa et al., 2011). Due to its direct application and wide availability, anaerobic digestion of microalgae has been studied for more than 60 years where many strains of microalgae have been studied to improve biogas yield (Zabed et al., 2020). More recently, the use of microalgal to treat liquid digestate (LD) from AD system has grown much interest (Xia and Murphy, 2016). LD was found to provide a high amount of nitrogen and phosphorus nutrient, while pH (due to uptake of CO_2) became the limiting factor (Uggetti et al., 2014). This implies the importance of pH regulation for this purpose. The integrated microalgae-based biodiesel and biogas production from recycling LD nutrients has also become an interesting paradigm for process integration studies.

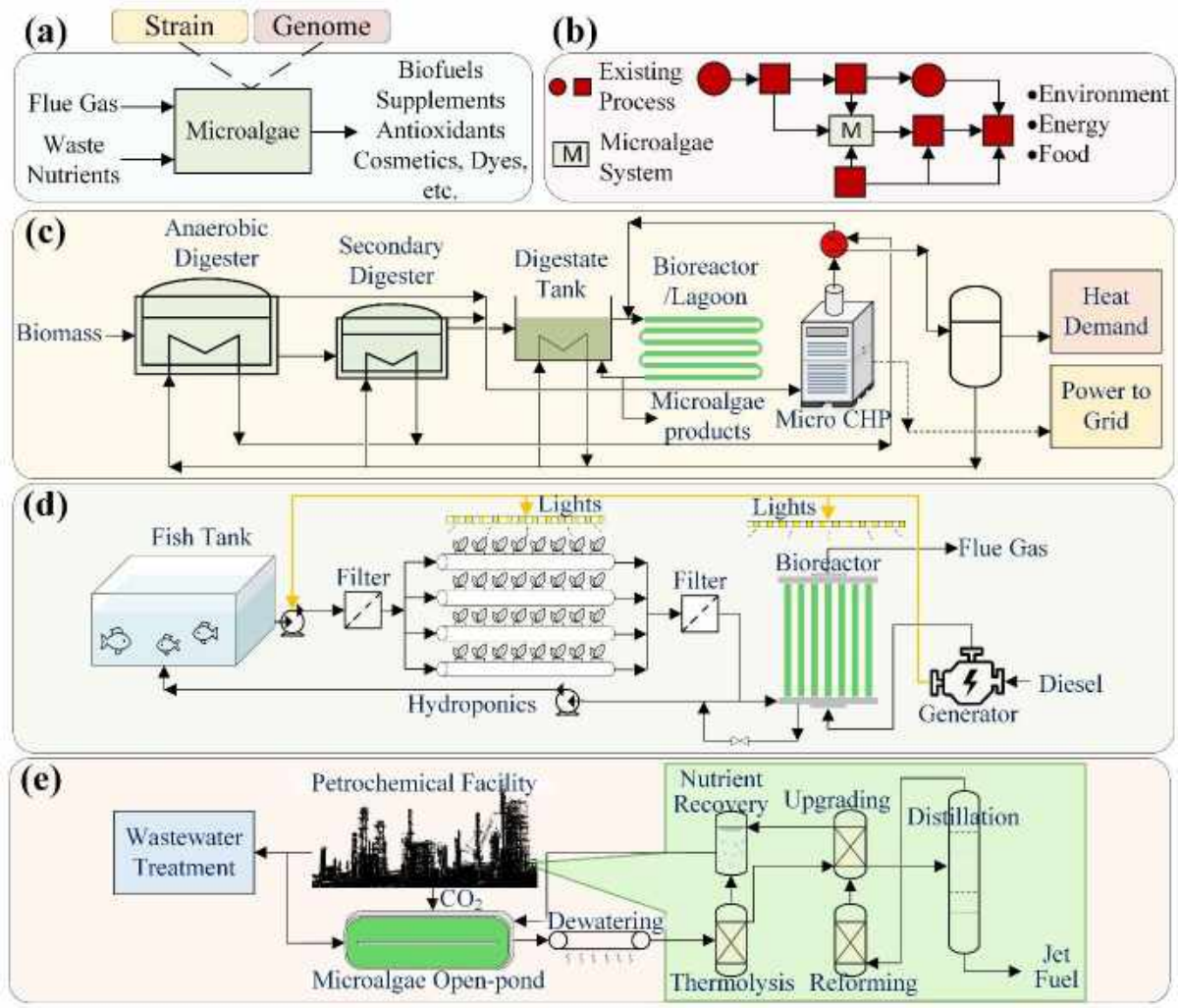


Figure 13.4: Schematic diagram of (a) Strains and genome effect in microalgae (b) microalgae as a process debottlenecking block (c) integration of microalgae in biogas facility (d) integration of microalgae in aquaponic facility (e) integration of microalgae in petrochemical industries.

Apart from AD process, microalgae systems can also be integrated into aquaponic systems (Figure 13.4(d)) to control the ammonia level within the water network (Addy et al., 2017). In this aspect, algal-bacterial consortia can be introduced to significantly improve nitrogen utilization efficiency by 13.79 % compared to media-based aquaponics (Fang et al., 2017). Additionally, microalgae can be processed into high-quality fish food (Andrade, 2018) and be re-integrated into the aquaponic system. Microalgae is also commonly integrated with combined heat and power (CHP) plants (Le et al., 2019). The commercial capacity of CHP plants can easily provide a large amount of CO_2, which opens the possibility of an upscaled microalgae system. Nevertheless, heat integration and process optimization remain a large challenge for CHP-based processing system (Lira-Barragán et al., 2015). The integration of

microalgae systems has also gained some attention in petrochemical industries (Figure 13.4(e)), providing bioenergy and carbon mitigation in the industry (Gonçalves et al., 2016). Microalgae systems can reduce contaminants in petrochemical wastewater (i.e. nitrogen, phosphorus, suspended solids, chemical oxygen demands (COD), diesel range organics, gasoline organics) and utilize them for growth (Hodges et al., 2017). Further on, the microalgae can be converted within the already available refinery into petrochemical products such as jet fuel (Wang et al., 2014). Integration of microalgae technologies was also possible for processes such as ethanol fermentation, sugarcane mills (Colling Klein et al., 2018), simultaneous recovery of lutein and biodiesel (Dineshkumar et al., 2015) and many other processes.

13.5 An Outlook for AI and Microalgae Biotechnologies

The use of AI is gaining in popularity in the research field of microalgae as AI algorithms can provide useful informatics on high uncertainty biosystems. In a broad scope of AI, the algorithms can be separated into 3 distinct fields, being machine learning, metaheuristics, and expert systems (Figure 13.5(a)). While the most common applications of ML in microalgae are in supervised learning and unsupervised learning (see Chapter 13.2-13.4), advanced ML framework like active learning can be used to give efficient sampling techniques in microalgae experiments and thus reducing the number of experiments to achieve optimal conditions (Figure 13.5(b)). This active learning approach has been applied in microalgae classification (Drews- et al., 2013), however, has yet to gain popularity in other parts of microalgae research fields. Recently, another subcategory of ML has also proven its effectiveness in microalgae research, being meta-learning (or sometimes known as “learning to learn”) (Figure 13.5 (c)). The concept of meta-learning has been well-described (Vilalta and Drissi, 2002) and the approach mainly uses metadata to speed up or automate ML tasks. A recent work (Panahi et al., 2019) shown that meta-learning had a high potential for RNA-seq meta-analysis in microalgae, which uncovered important biological pathways. Meta-learning using neuro-evolution was also carried out to predict the optimal conversion of microalgae via thermal methods (Teng et al., 2019). Furthermore, meta-learning techniques were also recently deployed for the prediction of algae growth (Serry et al., 2018), potentially giving more application to microalgae in the future.

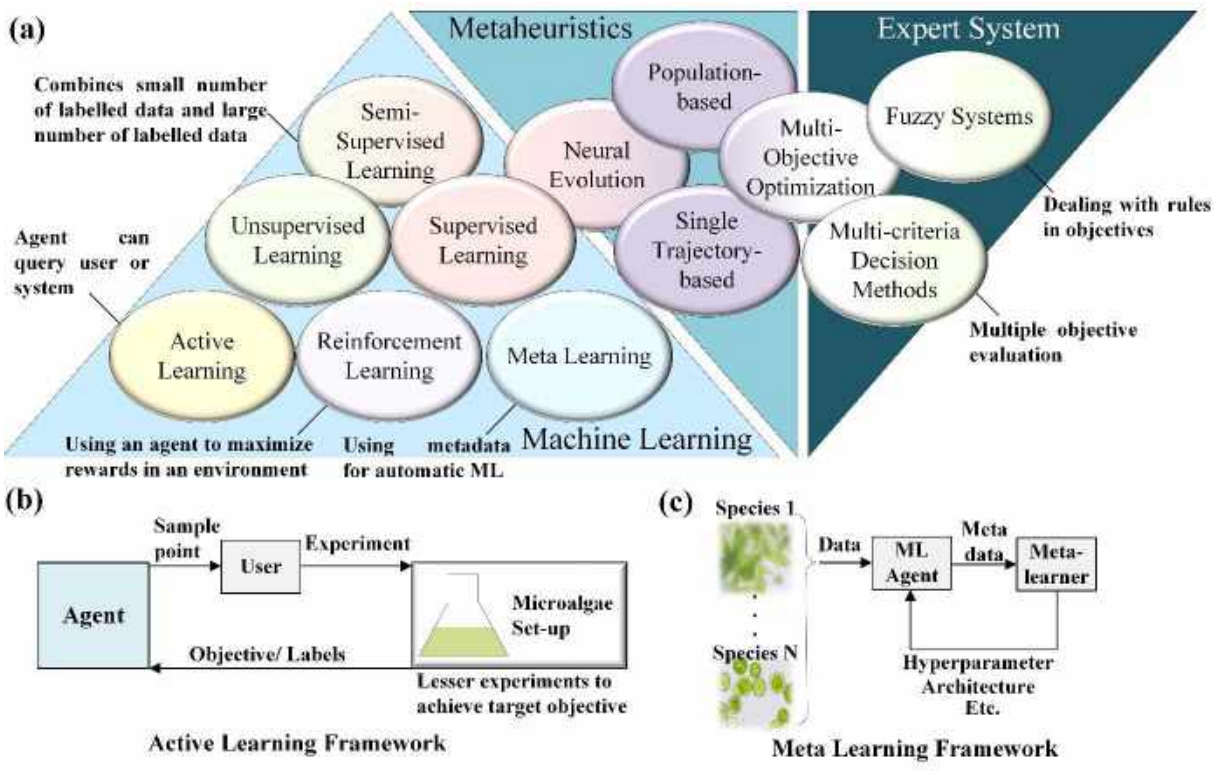


Figure 13.5: (a) Classification of a broad definition of AI and popular subcategories for the application of microalgae (b) Active learning framework for microalgae experiment (c) Meta-learning framework to deal with microalgae of different strains

Alternatively, semi-supervised learning (i.e. generative models) has also started seeing more applications in sequence design and genome editing (Linder et al., 2020) to automatically generate desirable genome sequence. Some of the potential applications of AI algorithm in microalgae research are compiled in Table 13.4. Nevertheless, more effort in compiling microalgae experimental data in various stages of research should be carried out. At the current stage, there are repository and databases for genome sequences and basic information of algae (Chapter 13.2.3). Databases from experimental growth, conversion, system design can also be compiled and standardized to accelerate the application of microalgae biotechnologies.

Table 13.4: Classes of AI algorithm for microalgae research tasks

Research Stage	Potential Class of Algorithms	Main Purpose	Examples of algorithms with application
Genome Sequencing	Supervised learning	• Primer selection. • Detection of gene expression.	Logistic regression, random forest, logistic model tree, random subspace (L. Wang et al., 2018)
	Unsupervised learning	• Genome feature extraction • Feature ranking	ANN, PCA, NMF, tSNE (Lin et al., 2017)
Gene Editing	Supervised learning	• Genome feature extraction	Ensemble ANN (Dutta et al., 2016), SVM, Hidden Markov Model (Persikov et al., 2009)
	Semi-supervised learning/ Generative models	• Sequence design	Deep exploration networks (Linder et al., 2020)
Strain-species identification	Supervised learning	• Image classification • Property classification	Mask R-CNN (Ruiz-Santaquiteria et al., 2020), ANN, SVM (Giraldo-Zuluaga et al., 2018)
	Unsupervised learning	• Featurization of microalgae signature.	SOM (Wirth et al., 2012)
Cultivation and Conversion	Supervised learning	• Optimal condition prediction	Radial basis function ANN, hyperbolic tangent ANN,(Zenooz et al., 2017), evolving neural network (Teng et al., 2019)
	Metaheuristics	• Search for optimal condition • Optimal parameter tuning	Particle swarm optimization (Chen et al., 2018), simulated annealing (Teng et al., 2019)

Table 13.4 (Cont.): Classes of AI algorithm for microalgae research tasks

Research Stage	**Potential Class of Algorithms**	• **Main Purpose**	**Examples of algorithms with application**
Cultivation and Conversion	Active learning	• Optimal experiment design under uncertainty	Markov chain Monte Carlo (Marchello et al., 2018)
	Expert systems	• Multi-criteria analysis for optimization	Analytical hierarchical process, TOPSIS (Nwokoagbara et al., 2015)
System design and control	Supervised learning	• Kinetic prediction for system design • System dynamics prediction	ANN-based model predictive control (Hu et al., 2008)
	Metaheuristics	• Optimization of system design and control	NSGA-II (Ansoni and Seleghim, 2016; Sinha et al., 2017), backtrack search algorithm (Mohd Zain et al., 2018)
	Reinforcement/ deep reinforcement learning	• Predictive control of systems • Simultaneous optimization and control	Online fitted deep Q-learning (Treloar et al., 2020), policy gradients (Petsagkourakis et al., 2020), Gaussian process with Monte Carlo (Bradford et al., 2020).
Process design and integration	Supervised learning	• System dynamics prediction	ANN (Arranz et al., 2008), gaussian process (Ažman and Kocijan, 2007)
	Metaheuristics	• High-speed optimization for large process networks	Differential evolution (Hernández-Pérez et al., 2019)
	Expert Systems	• Multi-criteria analysis for optimization	Analytical hierarchical process (Katooli et al., 2019)

13.6 Conclusion

In conclusion, microalgae are highly diverse microorganisms that require a lot of data and information from various fields of genome analysis, strain-species selection, experimental cultivation, conversion, process design, control, and integration. This work provides a concise review of using AI algorithms and microalgae informatics to improve current procedures of materializing desirable microalgae products from their genetic information. In terms of gene sequencing and editing, AI algorithms can extract critical knowledge and predict molecular interactions. Recent computer vision and ML algorithms have allowed for accurate strain species screening and classification, making high-quality microalgae images useful for further analysis. Various AI algorithm can also provide improvement in the cultivation and conversion of microalgae by reducing the number of experiments and condition optimization. Metaheuristics are particularly useful for system design and process integration due to the problem usually being large scale, while reinforcement learning, and model predictive control is the state-of-art for system control. With a growing compilation of microalgae data throughout the research field, a commercial-oriented "multi-omics" pipeline can be established for microalgae research. Future works in microalgae development should focus on data-driven analytics and more advanced ML methods (such as reinforcement learning, active learning, meta-learning etc.) for the purpose of genome sequencing, gene editing, unravelling genetic information, strain-species identification, microalgae cultivation and conversion, process system design, control and integration. Research efforts should also be allocated in unifying microalgae databases with the consideration of the "multi-omics" pipeline with commercial considerations. This opens the doors of possibility towards a digitalized and efficient future for microalgae biotechnologies and products.

Nomenclature

Abbreviation	Definition/Meaning
RNAi	Ribonucleic Acids Interfere
ZFN	Zinc-finger Nucleases
TALEN	Transcription Activator-like Nucleases
CRISPR-Cas9	Clustered Regularly Interspaced Short Palindromic Repeats-associated protein 9
AI	Artificial Intelligence
NGS	Next-generation Sequencing

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
SBL	Sequencing by Ligation
SBS	Sequencing by Synthesis
PGM	Personal Genome Machine
PacBio	Pacific Biosciences
SMRT	Single Molecules through Real-time Sequencing
ONT	Oxford Nanopore Technology
USB	Universal Serial Bus
ML	Machine Learning
DEG	Differentially expressed genes
PCA	Principal Component Analysis
ICA	Independent Component Analysis
tSNE	t-distributed Stochastic Neighbor Embedding
NMF	Non-negative Matrix Factorization
SINCERA	Single Cell RNA-seq Profiling Analysis (A computational pipeline)
SIMLR	Single-cell Interpretation via Multi-kernel Learning
SNN-Cliq	A novel clustering method for identification of cell types from single-cell transcriptomes
RNA-seq	RNA sequencing
PCR	Polymerase Chain Reaction
siRNA	Small Interfering RNA
ANN	Artificial Neural Network
SVM	Support Vector Machine
DSB	Double-strand Break
TIRF	Total Internal Reflection Fluorescence
WEKA	Waikato Environment for Knowledge Analysis Software
BRENDA	Braunschweig Enzyme Database
NCBI	National Center for Biotechnology Information
FBA	Flux Balance Analysis

EM	Elementary Modes
BOF	Biomass Objective Function
LSTM	Long Short-term Memory (Neural Network)
TPOT	Tree-based Pipeline Optimization
CNN	Convolutional Neural Network
Mask-RCNN	Region-based Convolutional Neural Network
SOM	Self-organizing Maps
MALDI	Matrix-assisted Laser Desorption/Ionization
GA-BP	Back Propagation Neural Network Optimized by Genetic Algorithm
FBN	Feed-forward Back-propagation Neural Network
RSM	Response Surface Methodology
CFD	Computational Fluid Dynamic
GPU	Graphical Processing Units
SQP	Sequential Quadratic Programming
NSGA-II	Non-dominated Sorting Genetic Algorithm II
MPC	Model Predictive Control
MHE	Moving Horizon Estimator
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
DQN	Deep Q-Learning
ANN-MPC	Artificial Neural Network-based Model Predictive Control
GOM	Gross Operating Margin
MINLP	Mixed-integer nonlinear programming
GGE	Gasoline Gallon Equivalent
AD	Anaerobic Digestion
LD	Liquid Digestate
CHP	Combined Heat and Power Plants

References

Addy, M.M., Kabir, F., Zhang, R., Lu, Q., Deng, X., Current, D., Griffith, R., Ma, Y., Zhou, W., Chen, P., Ruan, R., 2017. Co-cultivation of microalgae in aquaponic systems. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2017.08.151

Alipanahi, B., Delong, A., Weirauch, M.T., Frey, B.J., 2015. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat. Biotechnol. https://doi.org/10.1038/nbt.3300

An, S.M., Choi, D.H., Lee, H., Lee, J.H., Noh, J.H., 2018. Next-generation sequencing reveals the diversity of benthic diatoms in tidal flats. Algae. https://doi.org/10.4490/algae.2018.33.4.3

Andrade, L.M., 2018. Chlorella and Spirulina Microalgae as Sources of Functional Foods, Nutraceuticals, and Food Supplements; an Overview. MOJ Food Process. Technol. https://doi.org/10.15406/mojfpt.2018.06.00144

Andrade, L.M., Mendes, M.A., Kowalski, P., Nascimento, C.A.O., 2015. Comparative study of different matrix/solvent systems for the analysis of crude lyophilized microalgal preparations using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. https://doi.org/10.1002/rcm.7110

Ansoni, J.L., Seleghim, P., 2016. Optimal industrial reactor design: Development of a multiobjective optimization method based on a posteriori performance parameters calculated from CFD flow solutions. Adv. Eng. Softw. https://doi.org/10.1016/j.advengsoft.2015.08.008

Araujo, G.S., Matos, L.J.B.L., Gonçalves, L.R.B., Fernandes, F.A.N., Farias, W.R.L., 2011. Bioprospecting for oil producing microalgal strains: Evaluation of oil and biomass production for ten microalgal strains. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2011.01.089

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M., Brzezinski, M.A., Chaal, B.K., Chiovitti, A., Davis, A.K., Demarest, M.S., Detter, J.C., Glavina, T., Goodstein, D., Hadi, M.Z., Hellsten, U., Hildebrand, M., Jenkins, B.D., Jurka, J., Kapitonov, V. V., Kröger, N., Lau, W.W.Y., Lane, T.W., Larimer, F.W., Lippmeier, J.C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M.S., Palenik, B., Pazour, G.J., Richardson, P.M., Rynearson, T.A., Saito, M.A., Schwartz, D.C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F.P., Rokhsar, D.S., 2004. The genome of the diatom Thalassiosira Pseudonana: Ecology, evolution, and metabolism. Science (80-.). https://doi.org/10.1126/science.1101156

Arranz, A., Bordel, S., Villaverde, S., Zamarreño, J.M., Guieysse, B., Muñoz, R., 2008. Modeling photosynthetically oxygenated biodegradation processes using artificial neural networks. J. Hazard. Mater. https://doi.org/10.1016/j.jhazmat.2007.11.027

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. https://doi.org/10.1093/nar/gks400

Ashlock, D., Wittrock, A., Wen, T.J., 2002. Training finite state machines to improve PCR primer design, in: Proceedings of the 2002 Congress on Evolutionary Computation, CEC 2002. https://doi.org/10.1109/CEC.2002.1006202

Aziz, M., Oda, T., Kashiwagi, T., 2014. Integration of energy-efficient drying in microalgae utilization based on enhanced process integration. Energy. https://doi.org/10.1016/j.energy.2014.03.126

Ažman, K., Kocijan, J., 2007. Application of Gaussian processes for black-box modelling of biosystems. ISA Trans. https://doi.org/10.1016/j.isatra.2007.04.001

Baek, K., Kim, D.H., Jeong, J., Sim, S.J., Melis, A., Kim, J.S., Jin, E., Bae, S., 2016. DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. Sci. Rep. https://doi.org/10.1038/srep30620

Barbano, D., Diaz, R., Zhang, L., Sandrin, T., Gerken, H., Dempster, T., 2015. Rapid characterization of microalgae and microalgae mixtures using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS). PLoS One. https://doi.org/10.1371/journal.pone.0135337

Barclay, W., Apt, K., 2013. Strategies for Bioprospecting Microalgae for Potential Commercial Applications, in: Handbook of Microalgal Culture: Applied Phycology and Biotechnology: Second Edition. https://doi.org/10.1002/9781118567166.ch4

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Gane, P., Castro, L.G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W., Luo, J., Macdougall, A., Mutowo, P., Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Cowley, A., Figueira, L., Li, W., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant,

S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Suzek, B.E., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Yerramalla, M.S., Zhang, J., 2015. UniProt: A hub for protein information. Nucleic Acids Res. https://doi.org/10.1093/nar/gku989

Becker, E.W., 2017. Nutritional properties of microalgae: Potentials and constraints, in: CRC Handbook of Microalgal Mass Culture. https://doi.org/10.1201/9780203712405

Benemann, J.R., 2000. Hydrogen production by microalgae, in: Journal of Applied Phycology. https://doi.org/10.1002/9781118659892.ch8

Beneroso, D., Bermúdez, J.M., Arenillas, A., Menéndez, J.A., 2013. Microwave pyrolysis of microalgae for high syngas production. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2013.06.102

Bernard, O., 2011. Hurdles and challenges for modelling and control of microalgae for CO 2 mitigation and biofuel production, in: Journal of Process Control. https://doi.org/10.1016/j.jprocont.2011.07.012

Bernard, O., Mairet, F., Chachuat, B., 2016. Modelling of microalgae culture systems with applications to control and optimization. Adv. Biochem. Eng. Biotechnol. https://doi.org/10.1007/10_2014_287

Bitog, J.P., Lee, I.B., Lee, C.G., Kim, K.S., Hwang, H.S., Hong, S.W., Seo, I.H., Kwon, K.S., Mostafa, E., 2011. Application of computational fluid dynamics for modeling and designing photobioreactors for microalgae production: A review. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2011.01.015

Blanc-Mathieu, R., Verhelst, B., Derelle, E., Rombauts, S., Bouget, F.Y., Carré, I., Château, A., Eyre-Walker, A., Grimsley, N., Moreau, H., Piégu, B., Rivals, E., Schackwitz, W., Van de Peer, Y., Piganeau, G., 2014. An improved genome of the model marine alga Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. BMC Genomics. https://doi.org/10.1186/1471-2164-15-1103

Bogen, C., Al-Dilaimi, A., Albersmeier, A., Wichmann, J., Grundmann, M., Rupp, O., Lauersen, K.J., Blifernez-Klassen, O., Kalinowski, J., Goesmann, A., Mussgnug, J.H., Kruse, O., 2013. Reconstruction of the lipid metabolism for the microalga Monoraphidium neglectum

from its genome sequence reveals characteristics suitable for biofuel production. BMC Genomics. https://doi.org/10.1186/1471-2164-14-926

Borowitzka, M.A., 2013. Dunaliella: Biology, Production, and Markets, in: Handbook of Microalgal Culture: Applied Phycology and Biotechnology: Second Edition. https://doi.org/10.1002/9781118567166.ch18

Bottin, H., Lagoutte, B., 1992. Ferrodoxin and flavodoxin from the cyanobacterium Synechocystis sp PCC 6803. Biochim. Biophys. Acta (BBA)/Protein Struct. Mol. https://doi.org/10.1016/0167-4838(92)90465-P

Bradford, E., Imsland, L., Zhang, D., del Rio Chanona, E.A., 2020. Stochastic data-driven model predictive control using gaussian processes. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2020.106844

Breuer, G., Lamers, P.P., Martens, D.E., Draaisma, R.B., Wijffels, R.H., 2013. Effect of light intensity, pH, and temperature on triacylglycerol (TAG) accumulation induced by nitrogen starvation in Scenedesmus obliquus. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2013.05.105

Bwapwa, J.K., Anandraj, A., Trois, C., 2018a. Microalgae processing for jet fuel production. Biofuels, Bioprod. Biorefining. https://doi.org/10.1002/bbb.1878

Bwapwa, J.K., Anandraj, A., Trois, C., 2018b. Conceptual process design and -simulation of microalgae oil -conversion to aviation fuel. Biofuels, Bioprod. Biorefining. https://doi.org/10.1002/bbb.1890

Bwapwa, J.K., Anandraj, A., Trois, C., 2017. Possibilities for conversion of microalgae oil into aviation fuel: A review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2017.05.224

Bwapwa, J.K., Mutanda, T., Anandraj, A., 2019. A Sustainable Approach for Bioenergy and Biofuel Production from Microalgae, in: Gayen, K., Bhowmick, T.K., Maity, S.K. (Eds.), Sustainable Downstream Processing of Microalgae for Industrial Application. CRC Press, p. 137. https://doi.org/10.1201/9780429027970-6

Bzdok, D., Altman, N., Krzywinski, M., 2018. Statistics versus machine learning. Nat. Methods. https://doi.org/10.1038/nmeth.4642

Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., Collins, J.J., 2018. Next-Generation Machine Learning for Biological Networks. Cell. https://doi.org/10.1016/j.cell.2018.05.015

Cao, M., Kang, J., Gao, Y., Wang, X., Pan, X., Liu, P., 2020. Optimization of cultivation conditions for enhancing biomass, polysaccharide and protein yields of Chlorella sorokiniana by response surface methodology. Aquac. Res. https://doi.org/10.1111/are.14589

Carney, L.T., Reinsch, S.S., Lane, P.D., Solberg, O.D., Jansen, L.S., Williams, K.P., Trent, J.D., Lane, T.W., 2014. Microbiome analysis of a microalgal mass culture growing in municipal wastewater in a prototype OMEGA photobioreactor. Algal Res. https://doi.org/10.1016/j.algal.2013.11.006

Carroll, D., 2011. Genome engineering with zinc-finger nucleases. Genetics. https://doi.org/10.1534/genetics.111.131433

Chen, C.Y., Yeh, K.L., Aisyah, R., Lee, D.J., Chang, J.S., 2011. Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: A critical review. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2010.06.159

Chen, W.H., Chu, Y.S., Liu, J.L., Chang, J.S., 2018. Thermal degradation of carbohydrates, proteins and lipids in microalgae analyzed by evolutionary computation. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2018.01.036

Chew, K.W., Yap, J.Y., Show, P.L., Suan, N.H., Juan, J.C., Ling, T.C., Lee, D.J., Chang, J.S., 2017. Microalgae biorefinery: High value products perspectives. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2017.01.006

Chisti, Y., 2008. Biodiesel from microalgae beats bioethanol. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2007.12.002

Chow, K.C., Tung, W.L., 1999. Electrotransformation of Chlorella vulgaris. Plant Cell Rep. https://doi.org/10.1007/s002990050660

Chowdhury, R., Viamajala, S., Gerlach, R., 2012. Reduction of environmental and energy footprint of microalgal biodiesel production through material and energy integration. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2011.12.099

Christenson, L., Sims, R., 2011. Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2011.05.015

Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., Gu, F., Qu, S., Huang, D., Wei, J., Liu, Q., 2018. DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. Genome Biol. https://doi.org/10.1186/s13059-018-1459-4

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., Bayley, H., 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. https://doi.org/10.1038/nnano.2009.12

Colling Klein, B., Bonomi, A., Maciel Filho, R., 2018. Integration of microalgae production with industrial biofuel facilities: A critical review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2017.04.063

Coltelli, P., Barsanti, L., Evangelista, V., Gualtieri, P., 2017. Algae through the looking glass. Microsc. Res. Tech. https://doi.org/10.1002/jemt.22820

Costello, Z., Martin, H.G., 2018. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. npj Syst. Biol. Appl. https://doi.org/10.1038/s41540-018-0054-3

Čuperlović-Culf, M., 2013. Dynamic metabolic profiling and metabolite network and pathways modeling, in: NMR Metabolomics in Cancer Research. https://doi.org/10.1533/9781908818263.335

Dale, J.M., Popescu, L., Karp, P.D., 2010. Machine learning methods for metabolic pathway prediction. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-11-15

Del Rio-Chanona, E.A., Ahmed, N.R., Wagner, J., Lu, Y., Zhang, D., Jing, K., 2019. Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. Biotechnol. Bioeng. https://doi.org/10.1002/bit.27131

Dineshkumar, R., Dash, S.K., Sen, R., 2015. Process integration for microalgal lutein and biodiesel production with concomitant flue gas CO2 sequestration: a biorefinery model for healthcare, energy and environment. RSC Adv. https://doi.org/10.1039/c5ra09306f

Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). https://doi.org/10.1126/science.1258096

Dragone, G., Fernandes, B., Vicente, A., Teixeira, J., 2010. Third generation biofuels from microalgae. Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol. https://doi.org/10.1016/j.apenergy.2011.03.012

Drews-, P., Colares, R.G., Machado, P., de Faria, M., Detoni, A., Tavano, V., 2013. Microalgae classification using semi-supervised and active learning based on Gaussian mixture models. J. Brazilian Comput. Soc. https://doi.org/10.1007/s13173-013-0121-y

Dunahay, T.G., Jarvis, E.E., Dais, S.S., Roessler, P.G., 1996. Manipulation of microalgal lipid production using genetic engineering. Appl. Biochem. Biotechnol. - Part A Enzym. Eng. Biotechnol. https://doi.org/10.1007/BF02941703

Dutta, S., Madan, S., Parikh, H., Sundar, D., 2016. An ensemble micro neural network approach for elucidating interactions between zinc finger proteins and their target DNA. BMC Genomics. https://doi.org/10.1186/s12864-016-3323-9

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science (80-.). https://doi.org/10.1126/science.1162986

Fang, Y., Hu, Z., Zou, Y., Zhang, Jian, Zhu, Z., Zhang, Jianda, Nie, L., 2017. Improving nitrogen utilization efficiency of aquaponics by introducing algal-bacterial consortia. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2017.08.116

Feist, A.M., Palsson, B.O., 2010. The biomass objective function. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2010.03.003

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., Zhu, J.K., 2013. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. https://doi.org/10.1038/cr.2013.114

Figueira, C.E., Moreira, P.F., Giudici, R., 2015. Thermogravimetric analysis of the gasification of microalgae Chlorella vulgaris. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.09.059

Francke, C., Siezen, R.J., Teusink, B., 2005. Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol. https://doi.org/10.1016/j.tim.2005.09.001

Franco, B.M., Navas, L.M., Gómez, C., Sepúlveda, C., Acién, F.G., 2019. Monoalgal and mixed algal cultures discrimination by using an artificial neural network. Algal Res. https://doi.org/10.1016/j.algal.2019.101419

Gao, H., Wright, D.A., Li, T., Wang, Y., Horken, K., Weeks, D.P., Yang, B., Spalding, M.H., 2014. TALE activation of endogenous genes in Chlamydomonas reinhardtii. Algal Res. https://doi.org/10.1016/j.algal.2014.05.003

García-Camacho, F., López-Rosales, L., Sánchez-Mirón, A., Belarbi, E.H., Chisti, Y., Molina-Grima, E., 2016. Artificial neural network modeling for predicting the growth of the microalga Karlodinium veneficum. Algal Res. https://doi.org/10.1016/j.algal.2016.01.002

Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G., 2020. Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput. Struct. Biotechnol. J. https://doi.org/10.1016/j.csbj.2019.11.002

Giraldo-Zuluaga, J.H., Salazar, A., Diez, G., Gomez, A., Martínez, T., Vargas, J.F., Peñuela, M., 2018. Automatic identification of Scenedesmus polymorphic microalgae from microscopic images. Pattern Anal. Appl. https://doi.org/10.1007/s10044-017-0662-3

Gnouma, A., Sehli, E., Medhioub, W., Ben Dhieb, R., Masri, M., Mehlmer, N., Slimani, W., Sebai, K., Zouari, A., Brück, T., Medhioub, A., 2018. Strain selection of microalgae isolated from Tunisian coast: characterization of the lipid profile for potential biodiesel production. Bioprocess Biosyst. Eng. https://doi.org/10.1007/s00449-018-1973-5

Gonçalves, A.L., Alvim-Ferraz, M.C.M., Martins, F.G., Simões, M., Pires, J.C.M., 2016. Integration of microalgae-based bioenergy production into a petrochemical complex: Techno-economic assessment. Energies. https://doi.org/10.3390/en9040224

Gong, J., You, F., 2014. Global optimization for sustainable design and synthesis of algae processing network for CO2 mitigation and biofuel production using life cycle optimization. AIChE J. https://doi.org/10.1002/aic.14504

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: Ten years of next-generation sequencing technologies. Nat. Rev. Genet. https://doi.org/10.1038/nrg.2016.49

Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., Sizova, I., Hegemann, P., 2017. Targeting of photoreceptor genes in Chlamydomonas reinhardtii via zinc-finger nucleases and CRISPR/Cas9. Plant Cell. https://doi.org/10.1105/tpc.17.00659

Griffiths, M.J., Harrison, S.T.L., 2009. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. J. Appl. Phycol. https://doi.org/10.1007/s10811-008-9392-7

Guarnieri, M.T., Levering, J., Henard, C.A., Boore, J.L., Betenbaugh, M.J., Zengler, K., Knoshaug, E.P., 2018. Genome sequence of the oleaginous green alga, Chlorella vulgaris UTEX 395. Front. Bioeng. Biotechnol. https://doi.org/10.3389/fbioe.2018.00037

Guiry, M.D., Guiry, G.M., Morrison, L., Rindi, F., Miranda, S.V., Mathieson, A.C., Parker, B.C., Langangen, A., John, D.M., Bárbara, I., Carter, C.F., Kuipers, P., Garbary, D.J., 2014. AlgaeBase: An On-line Resource for Algae. Cryptogam. Algol. https://doi.org/10.7872/crya.v35.iss2.2014.105

Harun, R., Danquah, M.K., Forde, G.M., 2010. Microalgal biomass as a fermentation feedstock for bioethanol production. J. Chem. Technol. Biotechnol. https://doi.org/10.1002/jctb.2287

Hayashi, K., Kato, S., Matsunaga, S., 2018. Convolutional neural network-based automatic classification for algal morphogenesis. Cytologia (Tokyo). https://doi.org/10.1508/cytologia.83.301

He, L., Subramanian, V.R., Tang, Y.J., 2012. Experimental analysis and model-based optimization of microalgae growth in photo-bioreactors using flue gas. Biomass and Bioenergy. https://doi.org/10.1016/j.biombioe.2012.02.025

Hernández-Pérez, L.G., Sánchez-Tuirán, E., Ojeda, K.A., El-Halwagi, M.M., Ponce-Ortega, J.M., 2019. Optimization of Microalgae-to-Biodiesel Production Process Using a Metaheuristic Technique. ACS Sustain. Chem. Eng. https://doi.org/10.1021/acssuschemeng.9b00274

Hodges, A., Fica, Z., Wanlass, J., VanDarlin, J., Sims, R., 2017. Nutrient and suspended solids removal from petrochemical wastewater via microalgal biofilm cultivation. Chemosphere. https://doi.org/10.1016/j.chemosphere.2017.01.107

Hong, S.H., Cortesio, C.L., Drubin, D.G., 2015. Machine-Learning-Based Analysis in Genome-Edited Cells Reveals the Efficiency of Clathrin-Mediated Endocytosis. Cell Rep. https://doi.org/10.1016/j.celrep.2015.08.048

Hong, T., Zhang, J., Morris, A.J., Martin, E.B., Karim, M.N., 1996. Neural based predictive control of a multivariable microalgae fermentation, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. https://doi.org/10.1109/icsmc.1996.569793

Hu, D., Liu, H., Yang, C., Hu, E., 2008. The design and optimization for light-algae bioreactor controller based on Artificial Neural Network-Model Predictive Control. Acta Astronaut. https://doi.org/10.1016/j.actaastro.2008.02.008

Hu, Q., 2013. Environmental Effects on Cell Composition, in: Handbook of Microalgal Culture: Applied Phycology and Biotechnology: Second Edition. https://doi.org/10.1002/9781118567166.ch7

Husselmann, A. V., Hawick, K.A., 2013. Simulating growth kinetics in a data-parallel 3d lattice photobioreactor. Model. Simul. Eng. https://doi.org/10.1155/2013/153241

Jarvis, E.E., Brown, L.M., 1991. Transient expression of firefly luciferase in protoplasts of the green alga Chlorella ellipsoidea. Curr. Genet. https://doi.org/10.1007/BF00355062

Jeon, S., Lim, J.M., Lee, H.G., Shin, S.E., Kang, N.K., Park, Y. Il, Oh, H.M., Jeong, W.J., Jeong, B.R., Chang, Y.K., 2017. Current status and perspectives of genome editing technology for microalgae. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-017-0957-z

Jia, Y., Xue, L., Liu, H., Li, J., 2009. Characterization of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene from the halotolerant alga dunaliella salina and inhibition of its expression by RNAi. Curr. Microbiol. https://doi.org/10.1007/s00284-008-9333-3

Jones, T.R., Carpenter, A.E., Lamprecht, M.R., Moffat, J., Silver, S.J., Grenier, J.K., Castoreno, A.B., Eggert, U.S., Root, D.E., Golland, P., Sabatini, D.M., 2009. Scoring diverse cellular

morphologies in image-based screens with iterative feedback and machine learning. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.0808843106

Joung, J.K., Sander, J.D., 2013. TALENs: A widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/nrm3486

Karimi, M., 2017. Exergy-based optimization of direct conversion of microalgae biomass to biodiesel. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2016.09.032

Karp, P.D., 2002. The MetaCyc Database. Nucleic Acids Res. https://doi.org/10.1093/nar/30.1.59

Katooli, M.H., Aslani, A., Razi Astaraee, F., Mazzuca Sobczuk, T., Bakhtiar, A., 2019. Multi-criteria analysis of microalgae production in Iran. Biofuels. https://doi.org/10.1080/17597269.2018.1542566

Khan, M.I., Shin, J.H., Kim, J.D., 2018. The promising future of microalgae: Current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. Microb. Cell Fact. https://doi.org/10.1186/s12934-018-0879-x

Khataee, A.R., Zarei, M., Pourhassan, M., 2010. Bioremediation of malachite green from contaminated water by three microalgae: Neural network modeling. Clean - Soil, Air, Water. https://doi.org/10.1002/clen.200900233

Khataee, A.R., Zarei, M., Pourhassan, M., 2009. Application of microalga Chlamydomonas sp. for biosorptive removal of a textile dye from contaminated water: Modelling by a neural network. Environ. Technol. https://doi.org/10.1080/09593330903370018

Kim, K.M., Park, J.H., Bhattacharya, D., Yoon, H.S., 2014. Applications of next-generation sequencing to unravelling the evolutionary history of algae. Int. J. Syst. Evol. Microbiol. https://doi.org/10.1099/ijs.0.054221-0

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Sønderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B., Marcatili, P., 2019. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins Struct. Funct. Bioinforma. https://doi.org/10.1002/prot.25674

Krzemińska, I., Piasecka, A., Nosalewicz, A., Simionato, D., Wawrzykowski, J., 2015. Alterations of the lipid content and fatty acid profile of Chlorella protothecoides under different light intensities. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.07.043

Kumar, S., Jain, S., Kumar, H., 2018. Performance evaluation of adaptive neuro-fuzzy inference system and response surface methodology in modeling biodiesel synthesis from jatropha–algae oil. Energy Sources, Part A Recover. Util. Environ. Eff. https://doi.org/10.1080/15567036.2018.1515277

Kwon, Y.M., Kim, K.W., Choi, T.Y., Kim, S.Y., Kim, J.Y.H., 2018. Manipulation of the microalgal chloroplast by genetic engineering for biotechnological utilization as a green biofactory. World J. Microbiol. Biotechnol. https://doi.org/10.1007/s11274-018-2567-8

Le, T.G., Tran, D.T., van Do, T.C., Van Nguyen, T., 2019. Design considerations of microalgal culture ponds and photobioreactors for wastewater treatment and biomass cogeneration, in: Microalgae Biotechnology for Development of Biofuel and Wastewater Treatment. https://doi.org/10.1007/978-981-13-2264-8_21

Leow, S., Witter, J.R., Vardon, D.R., Sharma, B.K., Guest, J.S., Strathmann, T.J., 2015. Prediction of microalgae hydrothermal liquefaction products from feedstock biochemical composition. Green Chem. https://doi.org/10.1039/c5gc00574d

Leverenz, J.W., Falk, S., Pilström, C.M., Samuelsson, G., 1990. The effects of photoinhibition on the photosynthetic light-response curve of green plant cells (Chlamydomonas reinhardtii). Planta. https://doi.org/10.1007/BF00197105

Lin, C., Jain, S., Kim, H., Bar-Joseph, Z., 2017. Using neural networks for reducing the dimensions of single-cell RNA-Seq data. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx681

Lin, J., Wong, K.C., 2018. Off-target predictions in CRISPR-Cas9 gene editing using deep learning, in: Bioinformatics. https://doi.org/10.1093/bioinformatics/bty554

Lin, W.R., Tan, S.I., Hsiang, C.C., Sung, P.K., Ng, I.S., 2019. Challenges and opportunity of recent genome editing and multi-omics in cyanobacteria and microalgae for biorefinery. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2019.121932

Linder, J., Bogard, N., Rosenberg, A.B., Seelig, G., 2020. A Generative Neural Network for Maximizing Fitness and Diversity of Synthetic DNA and Protein Sequences. Cell Syst. https://doi.org/10.1016/j.cels.2020.05.007

Lira-Barragán, L.F., Gutiérrez-Arriaga, C.G., Bamufleh, H.S., Abdelhady, F., Ponce-Ortega, J.M., Serna-González, M., El-Halwagi, M.M., 2015. Reduction of greenhouse gas emissions from steam power plants through optimal integration with algae and cogeneration systems. Clean Technol. Environ. Policy. https://doi.org/10.1007/s10098-015-0982-1

Liu, B.H., Lee, Y.K., 2000. Secondary carotenoids formation by the green alga Chlorococcum sp, in: Journal of Applied Phycology. https://doi.org/10.1023/a:1008185212724

Liu, J.Y., Zeng, L.H., Ren, Z.H., Du, T.M., Liu, X., 2020. Rapid in situ measurements of algal cell concentrations using an artificial neural network and single-excitation fluorescence spectrometry. Algal Res. https://doi.org/10.1016/j.algal.2019.101739

López-Rosales, L., Gallardo-Rodríguez, J.J., Sánchez-Mirón, A., Contreras-Gómez, A., García-Camacho, F., Molina-Grima, E., 2013. Modelling of multi-nutrient interactions in growth of the dinoflagellate microalga Protoceratium reticulatum using artificial neural networks. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2013.07.141

Luo, Q., Gao, Y., Luo, J., Chen, C., Liang, J., Yang, C., 2011. Automatic identification of diatoms with circular shape using texture analysis. J. Softw. https://doi.org/10.4304/jsw.6.3.428-435

Majumdar, R., Rajasekaran, K., Cary, J.W., 2017. RNA interference (RNAi) as a potential tool for control of mycotoxin contamination in crop plants: Concepts and considerations. Front. Plant Sci. https://doi.org/10.3389/fpls.2017.00200

Mallona, I., Weiss, J., Marcos, E.C., 2011. PcrEfficiency: A Web tool for PCR amplification efficiency prediction. BMC Bioinformatics. https://doi.org/10.1186/1471-2105-12-404

Marchello, A.E., Oliveira, N.L., Lombardi, A.T., Polpo, A., 2018. An investigation onto Cd toxicity to freshwater microalga Chlorella sorokiniana in mixotrophy and photoautotrophy: A Bayesian approach. Chemosphere. https://doi.org/10.1016/j.chemosphere.2018.08.019

Martín, M., Grossmann, I.E., 2012. Simultaneous optimization and heat integration for biodiesel production from cooking oil and algae. Ind. Eng. Chem. Res. https://doi.org/10.1021/ie2024596

Mayol, A.P., Maningo, J.M.Z., Chua-Unsu, A.G.A.Y., Felix, C.B., Rico, P.I., Chua, G.S., Manalili, E. V., Fernandez, D.D., Cuello, J.L., Bandala, A.A., Ubando, A.T., Madrazo, C.F., Dadios, E., Culaba, A.B., 2019. Application of artificial neural networks in prediction of pyrolysis behavior for algal mat (LABLAB) biomass, in: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, HNICEM 2018. https://doi.org/10.1109/HNICEM.2018.8666376

McCormick, M., Villa, A.E.P., 2019. LSTM and 1-D Convolutional Neural Networks for Predictive Monitoring of the Anaerobic Digestion Process, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-30493-5_65

McGinn, P.J., Dickinson, K.E., Bhatti, S., Frigon, J.C., Guiot, S.R., O'Leary, S.J.B., 2011. Integration of microalgae cultivation with industrial waste remediation for biofuel and bioenergy production: Opportunities and limitations, in: Photosynthesis Research. https://doi.org/10.1007/s11120-011-9638-0

McKay, R.M.L., La Roche, J., Yakunin, A.F., Durnford, D.G., Geider, R.J., 1999. ACCUMULATION OF FERREDOXIN AND FLAVODOXIN IN A MARINE DIATOM IN RESPONSE TO FE. J. Phycol. https://doi.org/10.1046/j.1529-8817.1999.3530510.x

Mikheyev, A.S., Tin, M.M.Y., 2014. A first look at the Oxford Nanopore MinION sequencer. Mol. Ecol. Resour. https://doi.org/10.1111/1755-0998.12324

Mohd Zain, M.Z. bin, Kanesan, J., Kendall, G., Chuah, J.H., 2018. Optimization of fed-batch fermentation processes using the Backtracking Search Algorithm. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2017.07.034

Morowvat, M.H., Ghasemi, Y., 2016. Medium optimization by artificial neural networks for maximizing the triglycerides-rich lipids from biomass of Chlorella vulgaris. Int. J. Pharm. Clin. Res.

Mussgnug, J.H., 2015. Genetic tools and techniques for Chlamydomonas reinhardtii. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-015-6698-7

Mutwil, M., 2020. Computational approaches to unravel the pathways and evolution of specialized metabolism. Curr. Opin. Plant Biol. https://doi.org/10.1016/j.pbi.2020.01.007

Nayak, M., Dhanarajan, G., Dineshkumar, R., Sen, R., 2018. Artificial intelligence driven process optimization for cleaner production of biomass with co-valorization of wastewater and flue gas in an algal biorefinery. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.08.048

Neve, I.A.A., Sowa, J.N., Lin, C.C.J., Sivaramakrishnan, P., Herman, C., Ye, Y., Han, L., Wang, M.C., 2020. Escherichia coli metabolite profiling leads to the development of an RNA interference strain for Caenorhabditis elegans. G3 Genes, Genomes, Genet. https://doi.org/10.1534/g3.119.400741

Ng, I.S., Tan, S.I., Kao, P.H., Chang, Y.K., Chang, J.S., 2017. Recent Developments on Genetic Engineering of Microalgae for Biofuels and Bio-Based Chemicals. Biotechnol. J. https://doi.org/10.1002/biot.201600644

Ng, R.T.L., Tay, D.H.S., Ng, D.K.S., 2012. Simultaneous process synthesis, heat and power integration in a sustainable integrated biorefinery. Energy and Fuels. https://doi.org/10.1021/ef301283c

Nishida, I., Murata, N., 1996. Chilling sensitivity in plants and cyanobacteria: The crucial contribution of membrane lipids. Annu. Rev. Plant Physiol. Plant Mol. Biol. https://doi.org/10.1146/annurev.arplant.47.1.541

Nwokoagbara, E., Olaleye, A.K., Wang, M., 2015. Biodiesel from microalgae: The use of multi-criteria decision analysis for strain selection. Fuel. https://doi.org/10.1016/j.fuel.2015.06.074

Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M., Winge, P., 2016. A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci. Rep. https://doi.org/10.1038/srep24951

Ochoa-Estopier, L.M., Jobson, M., Smith, R., 2013. Operational optimization of crude oil distillation systems using artificial neural networks. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2013.05.030

Oey, M., Ross, I.L., Stephens, E., Steinbeck, J., Wolf, J., Radzun, K.A., Kügler, J., Ringsmuth, A.K., Kruse, O., Hankamer, B., 2013. RNAi Knock-Down of LHCBM1, 2 and 3 Increases Photosynthetic H2 Production Efficiency of the Green Alga Chlamydomonas reinhardtii. PLoS One. https://doi.org/10.1371/journal.pone.0061375

Oey, M., Sawyer, A.L., Ross, I.L., Hankamer, B., 2016. Challenges and opportunities for hydrogen production from microalgae. Plant Biotechnol. J. https://doi.org/10.1111/pbi.12516

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. https://doi.org/10.1093/nar/27.1.29

Pagani, I., Liolios, K., Jansson, J., Chen, I.M.A., Smirnova, T., Nosrat, B., Markowitz, V.M., Kyrpides, N.C., 2012. The Genomes OnLine Database (GOLD) v.4: Status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr1100

Palpant, N.J., Dudzinski, D., 2013. Zinc finger nucleases: Looking toward translation. Gene Ther. https://doi.org/10.1038/gt.2012.2

Panahi, B., Frahadian, M., Dums, J.T., Hejazi, M.A., 2019. Integration of cross species RNA-seq meta-analysis and machine-learning models identifies the most important salt stress–responsive pathways in microalga Dunaliella. Front. Genet. https://doi.org/10.3389/fgene.2019.00752

Pandian, B.J., Noel, M.M., 2018. Control of a bioreactor using a new partially supervised reinforcement learning algorithm. J. Process Control. https://doi.org/10.1016/j.jprocont.2018.07.013

Patron, N.J., 2020. Beyond natural: synthetic expansions of botanical form and function. New Phytol. https://doi.org/10.1111/nph.16562

Pedersen, M.F., Borum, J., 1996. Nutrient control of algal growth in estuarine waters. Nutrientlimitation and the importance of nitrogen requirements and nitrogen storageamong phytoplankton and species of macroalgae. Mar. Ecol. Prog. Ser. https://doi.org/10.3354/meps142261

Perrine, Z., Negi, S., Sayre, R.T., 2012. Optimization of photosynthetic light energy utilization by microalgae. Algal Res. https://doi.org/10.1016/j.algal.2012.07.002

Persikov, A. V., Osada, R., Singh, M., 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. Bioinformatics. https://doi.org/10.1093/bioinformatics/btn580

Petsagkourakis, P., Sandoval, I.O., Bradford, E., Zhang, D., del Rio-Chanona, E.A., 2020. Reinforcement learning for batch bioprocess optimization. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2019.106649

Pires, J.C.M., Alvim-Ferraz, M.C.M., Martins, F.G., 2017. Photobioreactor design for microalgae production through computational fluid dynamics: A review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2017.05.064

Poliner, E., Takeuchi, T., Du, Z.Y., Benning, C., Farré, E.M., 2018. Nontransgenic Marker-Free Gene Disruption by an Episomal CRISPR System in the Oleaginous Microalga, Nannochloropsis oceanica CCMP1779. ACS Synth. Biol. https://doi.org/10.1021/acssynbio.7b00362

Radakovits, R., Jinkerson, R.E., Darzins, A., Posewitz, M.C., 2010. Genetic engineering of algae for enhanced biofuel production. Eukaryot. Cell. https://doi.org/10.1128/EC.00364-09

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. https://doi.org/10.1038/nprot.2013.143

Rasala, B.A., Chao, S.S., Pier, M., Barrera, D.J., Mayfield, S.P., 2014. Enhanced genetic tools for engineering multigene traits into green algae. PLoS One. https://doi.org/10.1371/journal.pone.0094028

Rashid, N., Ryu, A.J., Jeong, K.J., Lee, B., Chang, Y.K., 2019. Co-cultivation of two freshwater microalgae species to improve biomass productivity and biodiesel production. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2019.05.106

Reijnders, M.J.M.F., van Heck, R.G.A., Lam, C.M.C., Scaife, M.A., dos Santos, V.A.P.M., Smith, A.G., Schaap, P.J., 2014. Green genes: Bioinformatics and systems-biology innovations drive algal biotechnology. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2014.10.003

Rismani-Yazdi, H., Haznedaroglu, B.Z., Bibby, K., Peccia, J., 2011. Transcriptome sequencing and annotation of the microalgae Dunaliella tertiolecta: Pathway description and gene discovery for production of next-generation biofuels. BMC Genomics. https://doi.org/10.1186/1471-2164-12-148

Rismani-Yazdi, H., Haznedaroglu, B.Z., Hsin, C., Peccia, J., 2012. Transcriptomic analysis of the oleaginous microalga Neochloris oleoabundans reveals metabolic insights into triacylglyceride accumulation. Biotechnol. Biofuels. https://doi.org/10.1186/1754-6834-5-74

Rizwan, M., Lee, J.H., Gani, R., 2015. Optimal design of microalgae-based biorefinery: Economics, opportunities and challenges. Appl. Energy. https://doi.org/10.1016/j.apenergy.2015.04.018

Rodolfi, L., Zittelli, G.C., Bassi, N., Padovani, G., Biondi, N., Bonini, G., Tredici, M.R., 2009. Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. Biotechnol. Bioeng. https://doi.org/10.1002/bit.22033

Rohr, J., Sarkar, N., Balenger, S., Jeong, B.R., Cerutti, H., 2004. Tandem inverted repeat system for selection of effective transgenic RNAi strains in Chlamydomonas. Plant J. https://doi.org/10.1111/j.1365-313X.2004.02227.x

Roy, S., Dutta, S., Khanna, K., Singla, S., Sundar, D., 2012. Prediction of DNA-binding specificity in zinc finger proteins, in: Journal of Biosciences. https://doi.org/10.1007/s12038-012-9213-7

Ruiz-Santaquiteria, J., Bueno, G., Deniz, O., Vallez, N., Cristobal, G., 2020. Semantic versus instance segmentation in microscopic algae detection. Eng. Appl. Artif. Intell. https://doi.org/10.1016/j.engappai.2019.103271

Russell, S., Norvig, P., 2010. Artificial Intelligence A Modern Approach, Third Edit. ed, Pearson. https://doi.org/10.1017/S0269888900007724

Ryan Georgianna, D., Mayfield, S.P., 2012. Exploiting diversity and synthetic biology for the production of algal biofuels. Nature. https://doi.org/10.1038/nature11479

Sætrom, P., 2004. Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. Bioinformatics. https://doi.org/10.1093/bioinformatics/bth364

Salam, S., Verma, T.N., 2019. Appending empirical modelling to numerical solution for behaviour characterisation of microalgae biodiesel. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2018.11.014

Sarrafzadeh, M.H., La, H.J., Seo, S.H., Asgharnejad, H., Oh, H.M., 2015. Evaluation of various techniques for microalgal biomass quantification. J. Biotechnol. https://doi.org/10.1016/j.jbiotec.2015.10.010

Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., Schomburg, D., 2013. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. Nucleic Acids Res. https://doi.org/10.1093/nar/gks1049

Serif, M., Lepetit, B., Weißert, K., Kroth, P.G., Rio Bartulos, C., 2017. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. Algal Res. https://doi.org/10.1016/j.algal.2017.02.005

Serry, H., Hassanien, A.E., Zaghlou, S., Hefny, H.A., 2018. Predicting algae growth in the nile river using meta-learning techniques, in: Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-319-64861-3_70

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L., Gao, C., 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. https://doi.org/10.1038/nbt.2650

Shan, T., Yuan, J., Su, L., Li, J., Leng, X., Zhang, Y., Gao, H., Pang, S., 2020. First Genome of the Brown Alga Undaria pinnatifida: Chromosome-Level Assembly Using PacBio and Hi-C Technologies. Front. Genet. https://doi.org/10.3389/fgene.2020.00140

Sharon Mano Pappu, J., Vijayakumar, G.K., Ramamurthy, V., 2013. Artificial neural network model for predicting production of Spirulina platensis in outdoor culture. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2012.12.082

Shin, E., Liobet, J.W., Gardner, E.L., Hines, C.S., Dow, 2000. Classification of the strain and growth phase of cyanobacteria in potable water using an electronic nose system. IEE Proc. Sci. Meas. Technol. https://doi.org/10.1049/ip-smt:20000422

Shin, H., Lee, E., Shin, J., Ko, S.R., Oh, H.S., Ahn, C.Y., Oh, H.M., Cho, B.K., Cho, S., 2018. Elucidation of the bacterial communities associated with the harmful microalgae Alexandrium tamarense and Cochlodinium polykrikoides using nanopore sequencing. Sci. Rep. https://doi.org/10.1038/s41598-018-23634-6

Shin, H.S., Hong, S.J., Kim, H., Yoo, C., Lee, H., Choi, H.K., Lee, C.G., Cho, B.K., 2015. Elucidation of the growth delimitation of Dunaliella tertiolecta under nitrogen stress by integrating transcriptome and peptidome analysis. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.07.002

Shin, S.E., Lim, J.M., Koh, H.G., Kim, E.K., Kang, N.K., Jeon, S., Kwon, S., Shin, W.S., Lee, B., Hwangbo, K., Kim, J., Ye, S.H., Yun, J.Y., Seo, H., Oh, H.M., Kim, K.J., Kim, J.S., Jeong, W.J., Chang, Y.K., Jeong, B.R., 2016. CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. Sci. Rep. https://doi.org/10.1038/srep27810

Shokrkar, H., Ebrahimi, S., Zamani, M., 2017. Extraction of sugars from mixed microalgae culture using enzymatic hydrolysis: Experimental study and modeling. Chem. Eng. Commun. https://doi.org/10.1080/00986445.2017.1356291

Shrager, J., Hauser, C., Chang, C.W., Harris, E.H., Davies, J., McDermott, J., Tamse, R., Zhang, Z., Grossman, A.R., 2003. Chlamydomonas reinhardtii genome project. A guide to the generation and use of the cDNA information. Plant Physiol. https://doi.org/10.1104/pp.016899

Sialve, B., Bernet, N., Bernard, O., 2009. Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2009.03.001

Sinha, S.K., Kumar, M., Guria, C., Kumar, A., Banerjee, C., 2017. Biokinetic model-based multi-objective optimization of Dunaliella tertiolecta cultivation using elitist non-dominated sorting genetic algorithm with inheritance. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2017.03.146

Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., Hegemann, P., 2013. Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. Plant J. https://doi.org/10.1111/tpj.12066

Slegers, P.M., Wijffels, R.H., van Straten, G., van Boxtel, A.J.B., 2011. Design scenarios for flat panel photobioreactors. Appl. Energy. https://doi.org/10.1016/j.apenergy.2010.12.037

Smith, R., 2005. Chemical Process Design and Integration, John Wiley & Sons, Ltd. https://doi.org/10.1529/biophysj.107.124164

Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods. https://doi.org/10.4319/lom.2007.5.204

Srivastava, G., Paul, A.K., Goud, V. V., 2018. Optimization of non-catalytic transesterification of microalgae oil to biodiesel under supercritical methanol condition. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2017.10.093

Steadman Tyler, C.R., Hovde, B.T., Daligault, H.E., Zhang, X.L., Kunde, Y., Marrone, B.L., Twary, S.N., Starkenburg, S.R., 2019. High-Quality Draft Genome Sequence of the Green Alga Tetraselmis striata (Chlorophyta) Generated from PacBio Sequencing. Microbiol. Resour. Announc. https://doi.org/10.1128/mra.00780-19

Suali, E., Sarbatly, R., 2012. Conversion of microalgae to biofuel. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2012.03.047

Sukačová, K., Kočí, R., Žídková, M., Vítěz, T., Trtílek, M., 2017. Novel insight into the process of nutrients removal using an algal biofilm: The evaluation of mechanism and efficiency. Int. J. Phytoremediation. https://doi.org/10.1080/15226514.2017.1303810

Sun, T., Zhou, B., Lai, L., Pei, J., 2017. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. BMC Bioinformatics. https://doi.org/10.1186/s12859-017-1700-2

Supriyanto, Noguchi, R., Ahamed, T., Rani, D.S., Sakurai, K., Nasution, M.A., Wibawa, D.S., Demura, M., Watanabe, M.M., 2019. Artificial neural networks model for estimating growth of polyculture microalgae in an open raceway pond. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2018.10.002

Sureyya Rifaioglu, A., Doğan, T., Jesus Martin, M., Cetin-Atalay, R., Atalay, V., 2019. DEEPred: Automated Protein Function Prediction with Multi-task Feed-forward Deep Neural Networks. Sci. Rep. https://doi.org/10.1038/s41598-019-43708-3

Susanna, D., Dhanapal, R., Mahalingam, R., Ramamurthy, V., 2019. Increasing productivity of Spirulina platensis in photobioreactors using artificial neural network modeling. Biotechnol. Bioeng. https://doi.org/10.1002/bit.27128

Takouridis, S.J., Tribe, D.E., Gras, S.L., Martin, G.J.O., 2015. The selective breeding of the freshwater microalga Chlamydomonas reinhardtii for growth in salinity. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2014.10.120

Taleb, A., Pruvost, J., Legrand, J., Marec, H., Le-Gouic, B., Mirabella, B., Legeret, B., Bouvet, S., Peltier, G., Li-Beisson, Y., Taha, S., Takache, H., 2015. Development and validation of a screening procedure of microalgae for biodiesel production: Application to the genus of marine microalgae Nannochloropsis. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2014.11.068

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F., Máša, V., 2019. Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresour. Technol. 292, 121971. https://doi.org/10.1016/j.biortech.2019.121971

Thompson, P.A., Guo, M. -x, Harrison, P.J., 1992. Effects Of Variation In Temperature. I. On The Biochemical Composition Of Eight Species Of Marine Phytoplankton. J. Phycol. https://doi.org/10.1111/j.0022-3646.1992.00481.x

Tian, Y., Smith, D.R., 2016. Recovering complete mitochondrial genome sequences from RNA-Seq: A case study of Polytomella non-photosynthetic green algae. Mol. Phylogenet. Evol. https://doi.org/10.1016/j.ympev.2016.01.017

Tibocha-Bonilla, J.D., Zuñiga, C., Godoy-Silva, R.D., Zengler, K., 2018. Advances in metabolic modeling of oleaginous microalgae Mike Himmel. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-018-1244-3

Toimil, D., Gómez, A., 2017. Review of metaheuristics applied to heat exchanger network design. Int. Trans. Oper. Res. 24, 7–26. https://doi.org/10.1111/itor.12296

Torzillo, G., Vonshak, A., 2013. Environmental Stress Physiology with Reference to Mass Cultures, in: Handbook of Microalgal Culture: Applied Phycology and Biotechnology: Second Edition. https://doi.org/10.1002/9781118567166.ch6

Treloar, N.J., Fedorec, A.J.H., Ingalls, B., Barnes, C.P., 2020. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1007783

Uggetti, E., Sialve, B., Latrille, E., Steyer, J.P., 2014. Anaerobic digestate as substrate for microalgae culture: The role of ammonium concentration on the microalgae productivity. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2013.11.036

Vilalta, R., Drissi, Y., 2002. A perspective view and survey of meta-learning. Artif. Intell. Rev. https://doi.org/10.1023/A:1019956318069

Walter, T., Held, M., Neumann, B., Hériché, J.K., Conrad, C., Pepperkok, R., Ellenberg, J., 2010. Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. J. Struct. Biol. https://doi.org/10.1016/j.jsb.2009.10.004

Wang, H.Y., Bluck, D., Van Wie, B.J., 2014. Conversion of microalgae to jet fuel: Process design and simulation. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2014.05.092

Wang, L., Xi, Y., Sung, S., Qiao, H., 2018. RNA-seq assistant: Machine learning based methods to identify more transcriptional regulated genes. BMC Genomics. https://doi.org/10.1186/s12864-018-4932-2

Wang, N., Qian, Z., Luo, M., Fan, S., Zhang, X., Zhang, L., 2018. Identification of salt stress responding genes using transcriptome analysis in green alga chlamydomonas reinhardtii. Int. J. Mol. Sci. https://doi.org/10.3390/ijms19113359

Wang, P., 2008. What do you mean by "AI"?, in: Frontiers in Artificial Intelligence and Applications.

Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H., Xu, J., 2017. RNAi-based targeted gene knockdown in the model oleaginous microalgae Nannochloropsis oceanica. Plant J. https://doi.org/10.1111/tpj.13411

Weyman, P.D., Beeri, K., Lefebvre, S.C., Rivera, J., Mccarthy, J.K., Heuberger, A.L., Peers, G., Allen, A.E., Dupont, C.L., 2015. Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. Plant Biotechnol. J. https://doi.org/10.1111/pbi.12254

Wheeler, D.L., 2006. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. https://doi.org/10.1093/nar/gkj158

Wilkinson, S.J., Benson, N., Kell, D.B., 2007. Proximate parameter tuning for biochemical networks with uncertain kinetic parameters. Mol. Biosyst. https://doi.org/10.1039/b707506e

Wirth, H., von Bergen, M., Murugaiyan, J., Rösler, U., Stokowy, T., Binder, H., 2012. MALDI-typing of infectious algae of the genus Prototheca using SOM portraits. J. Microbiol. Methods. https://doi.org/10.1016/j.mimet.2011.10.013

Wolf, J., Ross, I.L., Radzun, K.A., Jakob, G., Stephens, E., Hankamer, B., 2015. High-throughput screen for high performance microalgae strain selection and integrated media design. Algal Res. https://doi.org/10.1016/j.algal.2015.07.005

Wu, L.C., Horng, J.T., Huang, H.Y., Lin, F.M., Huang, H. Da, Tsai, M.F., 2007. Primer design for multiplex PCR using a genetic algorithm. Soft Comput. https://doi.org/10.1007/s00500-006-0137-8

Wu, Z., Shi, X., 2007. Optimization for high-density cultivation of heterotrophic Chlorella based on a hybrid neural network model. Lett. Appl. Microbiol. https://doi.org/10.1111/j.1472-765X.2006.02038.x

Xia, A., Murphy, J.D., 2016. Microalgal Cultivation in Treating Liquid Digestate from Biogas Systems. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2015.12.010

Xu, L., Weathers, P.J., Xiong, X.R., Liu, C.Z., 2009. Microalgal bioreactors: Challenges and opportunities. Eng. Life Sci. https://doi.org/10.1002/elsc.200800111

Yang, X., Guo, F., Xue, S., Wang, X., 2016. Carbon distribution of algae-based alternative aviation fuel obtained by different pathways. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2015.10.045

Yao, L., Tan, T.W., Ng, Y.K., Ban, K.H.K., Shen, H., Lin, H., Lee, Y.K., 2015. RNA-Seq transcriptomic analysis with Bag2D software identifies key pathways enhancing lipid yield in a high lipid-producing mutant of the non-model green alga Dunaliella tertiolecta. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-015-0382-0

Yazdani, H., Khoshhal, A., Mousavi, N.S., 2020. Evaluating the performance of a sequencing batch reactor for sanitary wastewater treatment using artificial neural network. Environ. Prog. Sustain. Energy. https://doi.org/10.1002/ep.13438

Yew, G.Y., Chew, K.W., Malek, M.A., Ho, Y.C., Chen, W.H., Ling, T.C., Show, P.L., 2019. Hybrid liquid biphasic system for cell disruption and simultaneous lipid extraction from microalgae Chlorella sorokiniana CY-1 for biofuel production. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-019-1591-8

Yoshimitsu, Y., Abe, J., Harayama, S., 2018. Cas9-guide RNA ribonucleoprotein-induced genome editing in the industrial green alga Coccomyxa sp. strain KJ 06 Biological Sciences 0604 Genetics. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-018-1327-1

Yugi, K., Kubota, H., Hatano, A., Kuroda, S., 2016. Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple "Omic" Layers. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2015.12.013

Zabed, H.M., Akter, S., Yun, J., Zhang, G., Zhang, Y., Qi, X., 2020. Biogas from microalgae: Technologies, challenges and opportunities. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2019.109503

Zamalloa, C., Vulsteke, E., Albrecht, J., Verstraete, W., 2011. The techno-economic potential of renewable energy through the anaerobic digestion of microalgae. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2010.09.017

Zenooz, A.M., Ashtiani, F.Z., Ranjbar, R., Nikbakht, F., Bolouri, O., 2017. Comparison of different artificial neural network architectures in modeling of Chlorella sp. flocculation. Prep. Biochem. Biotechnol. https://doi.org/10.1080/10826068.2016.1275013

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., Liang, J., Cai, C., Liu, Z., Liu, B., Wang, F., Li, S., Liu, F., Li, X., Cheng, L., Yang, W., Li, M. he, Grossniklaus, U., Zheng, H., Wang, X., 2018. Improved Brassica rapa reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic. Res. https://doi.org/10.1038/s41438-018-0071-9

CHAPTER 12 FINAL CONCLUSIONS AND FUTURE WORKS

In conclusion, this work proposes a novel process improvement framework that is process-agnostic and can be implemented in any generic process systems. Literature review demonstrates that the availability of high quality and useful data within process facility is generally difficult as many small and medium industries do not have good data infrastructure (such as SCADA system). In this case, the proposed framework considers the availability of data, data cleaning, structuring, and pre-processing, process optimization for single unit, process optimization for multiple units, process debottlenecking, multi-dimensional process debottlenecking, decision-making for new technologies, and real-time managements. The elements in the framework was demonstrated in multiple case studies from oil refineries, cogeneration systems, waste-to-energy systems, biogas facilities which was mainly situated in Malaysia and Czech Republic.

An overall guiding framework for process improvement was proposed in Chapter 3, Figure 3.2. For low data availability, Chapter 4 proposed a one-shot learning approach using Siamese networks and transfer learning to deal with low data availability problems. Chapter 5 demonstrates the usefulness of a progressive depth swarm evolution (PDSE) neuro-evolutionary approach to optimize a microalgae thermal reactor in achieving high conversion and low energy consumption. Chapter 6 and 7 demonstrates the usefulness of dimension reduction-based optimization for prioritization of operational variables during multi-unit process optimization. For capacity debottlenecking, Chapter 8 and 9 demonstrates a novel bottleneck tree analysis method (BOTA) that can be used with process simulation or neural network ensembles. Chapter 10 shows the effectiveness of using Monte Carlo simulations, neural network, and decision trees to carry out techno-economic analysis in the case where a new unit is considered for existing plants. Next, Chapter 11 propose the use of a neocortex inspired algorithm, Hierarchical Temporal Memory (HTM) and novel dual-mode optimization algorithm for the purpose of real-time management of a waste-to-energy plant.

In this work, the application of AI and data-driven approaches were proven to be effective for the purpose of industrial process improvement. These techniques can be used seamlessly with traditional model-based approach (such as commercial process simulators), and even the P-

graph framework. For this, Chapter 12 demonstrates the novel implementation of a neural network in the P-graph framework. This work also contributed to the following aspect:

- Improve process modelling and optimization procedure under low data availability and high-uncertainty nonlinear systems.
- Demonstrate the practicability of dimension reduction-based optimization for multi-unit process optimization
- Propose novel and flexible process capacity debottlenecking method that can be used in any process plants.
- Improved decision-making for the consideration of new technologies and real-time management by utilizing data-driven insights.

In such aspects, this work has provided a useful paradigm for data-driven process engineering in industrial application, academic research contribution, and future engineering education. Future works will focus on more novel and efficient algorithms for process improvement. Also, as described in Chapter 13, future works will also focus more on utilizing AI for microalgae technologies in process design and improvement.

Appendix

List of Author's publications and research outputs

This is the list of the author's publication. The list is spitted into two chapters: (i) selected publications (that are inclusive of the work in this thesis), (ii) other related publications, and (iii) software.

Selected Publications

Teng, S.Y., Touš, M., Leong, W.D., How, B.S., Lam, H.L. and Máša, V., Recent advances on industrial data-driven energy savings: Digital twins and infrastructures. Renewable and Sustainable Energy Reviews, 135, p.110208. **(IF=12.11)**

Teng, S.Y., Loy, A.C.M., Leong, W.D., How, B.S., Chin, B.L.F. and Máša, V., 2019. Catalytic thermal degradation of Chlorella vulgaris: Evolving deep neural networks for optimization. Bioresource technology, 292, p.121971. **(IF=6.669)**

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H. and Lam, H.L., 2020. Bottleneck Tree Analysis (BOTA) with green and lean index for process capacity debottlenecking in industrial refineries. Chemical Engineering Science, 214, p.115429. **(IF=3.871)**

Teng, S.Y., How, B.S., Leong, W.D., Teoh, J.H., Cheah, A.C.S., Motavasel, Z. and Lam, H.L., 2019. Principal component analysis-aided statistical process optimisation (PASPO) for process improvement in industrial refineries. Journal of Cleaner Production, 225, pp.359-375. **(IF=6.395)**

Vondra, M., Touš, M. and **Teng, S.Y.**, 2019. Digestate evaporation treatment in biogas plants: A techno-economic assessment by Monte Carlo, neural networks and decision trees. Journal of Cleaner Production, 238, p.117870. **(IF=6.395)**

Teng, S.Y., Yew, G.Y., Sukačová, K., Show, P.L., Máša, V. and Chang, J.S., 2020. Microalgae with artificial intelligence: A digitalized perspective on genetics, systems and products. Biotechnology Advances, p.107631. **(IF=10.744)**

Teng, S.Y., Máša, V., Stehlík, P. and Lam, H.L., 2019. Deep Learning Approach for Industrial Process Improvement. Chemical Engineering Transactions, 76, pp.487-492. **(SJR= 0.32)**

Teng S.Y., Máša V., Lam H.L., Stehlík P., 2020, A One-Shot Learning Framework to Model Process Systems, Chemical Engineering Transactions, 81, 937-942. **(SJR= 0.32)**

How, B.S., **Teng, S.Y.**, Leong, W.D., Ng, W.P.Q., Lim, C.H., Ngan, S.L. and Lam, H.L., 2019. Non-linear Programming via P-graph Framework. Chemical Engineering Transactions, 76, pp.499-504. **(SJR= 0.32)**

Other Related Publications

Ngan, S.L., How, B.S., **Teng, S.Y.**, Promentilla, M.A.B., Yatim, P., Er, A.C. and Lam, H.L., 2019. Prioritization of sustainability indicators for promoting the circular economy: The case of developing countries. Renewable and Sustainable Energy Reviews, 111, pp.314-331. **(IF= 10.556)**

Leong, W.D., **Teng, S.Y.**, How, B.S., Ngan, S.L., Lam, H.L., Tan, C.P. and Ponnambalam, S.G., 2019. Adaptive analytical approach to lean and green operations. Journal of Cleaner Production, 235, pp.190-209. **(IF=6.395)**

Ngan, S.L., How, B.S., **Teng, S.Y.**, Leong, W.D., Loy, A.C.M., Yatim, P., Promentilla, M.A.B. and Lam, H.L., 2020. A hybrid approach to prioritize risk mitigation strategies for biomass polygeneration systems. Renewable and Sustainable Energy Reviews, 121, p.109679. **(IF=12.110)**

Konečná, E., **Teng, S.Y.** and Máša, V., 2020. New insights into the potential of the gas microturbine in microgrids and industrial applications. Renewable and Sustainable Energy Reviews, 134, p.110078. **(IF=12.110)**

Zhang, X., **Teng, S.Y.**, Loy, A.C.M., How, B.S., Leong, W.D. and Tao, X., 2020. Transition Metal Dichalcogenides for the Application of Pollution Reduction: A Review. Nanomaterials, 10(6), p.1012. **(IF=4.324)**

Kong K.G.H., Lo S.L.Y., How B.S., Leong W.D., **Teng S.Y.**, Ng W.P., Sunarso J., 2020, Enhanced Automated Targeting Model for Multi-Period Energy Planning, Chemical Engineering Transactions, 81, 607-612. **(SJR= 0.32)**

Leong, W.D., How, B.S., **Teng, S.Y.**, Ngan, S.L., Ng, W.P.Q., Lim, C.H., Lam, H.L., Tan, C.P. and Ponnambalam, S.G., 2019. Integration of Analytic Network Process in Adaptive Lean and Green Processing. Chemical Engineering Transactions, 76, pp.559-564. **(SJR= 0.32)**

Leong, W.D., **Teng, S.Y.**, How, B.S., Ngan, S.L., Abd Rahman, A., Tan, C.P., Ponnambalam, S.G. and Lam, H.L., 2020. Enhancing the Adaptability: Lean and Green Strategy towards the Industry Revolution 4.0. Journal of Cleaner Production,273, p.122870. **(IF=7.246)**

Yew, G.Y., Puah, B.K., Chew, K.W., **Teng, S.Y.**, Show, P.L. and Nguyen, T.H.P., 2020. Chlorella vulgaris FSP-E cultivation in waste molasses: Photo-to-property estimation by artificial intelligence. Chemical Engineering Journal,402, p.126230. **(IF=10.652)**

Lo, S.L.Y., How, B.S., Leong, W.D., **Teng, S.Y.**, Rahamdhani, M.A., Sunarso, J., 2021. Techno-economic analysis for biomass supply chain: A state-of-the-art review. Renewable and Sustainable Energy Reviews, 135, p.110164. **(IF=12.110)**

Software

Teng, S.Y., 2019. EvoOpt- Evolutionary Optimization in Python. tsyet12/EvoOpt: EvoOpt v0.12 pre-release. doi:10.5281/zenodo.3241951

Teng, S.Y., Michal Touš, 2020. Energy Brains: Automatic Improvement of Energy-Intensive Processes using Data-Driven Analytics. v0.10 alpha. Inhouse software.