PhD in
Environmental
Modelling | 2023

Hybrid Machine Learning Methods for the Reconstruction and Disaggregation of Gridded Climatic and Hydrological Variables

Ujjwal Singh

Department of Water Resources and Environmental Modelling
Faculty of Environmental Sciences
Czech University of Life Sciences Prague

Department of Water Resources and Environmental Modeling
Faculty of Environmental Sciences
Czech University of Life Sciences Prague

Hybrid Machine Learning Methods for the Reconstruction and Disaggregation of Gridded Climatic and Hydrological Variables

Presented by Ujjwal Singh

PhD thesis in Landscape Engineering Environmental Modelling

December 2023

Petr Maca   Associate Professor   Supervisor
Rama Rao Nidamanuri   Professor   Consultant

Acknowledgements

Kant Mishra, Reshma Kumari and Anis Rai, who supported my decision to return to India and pursue two years of a role as a town planner. Your steadfast belief and encouragement have been my guiding light. Thank you.

ABSTRACT

Ongoing climate change examining scenarios, accurate precipitation esti-
mation and understanding runoff dynamics are vital in managing hydrol-
ogy, meteorology, and water resources. While different in focus, these two
domains contribute to an intricate understanding of water cycle compo-
nents, particularly in regions experiencing water resource degradation. In
both studies presented within this doctoral thesis, sophisticated machine
learning techniques deliver in-depth insights. Specifically, this research
delves into two distinct facets of hybrid modelling: the disaggregation of
satellite-based precipitation data and the reconstruction of historical runoff
reconstruction.

The first study focuses on the methodological hybrid machine learn-
ing modelling framework for achieving accurate precipitation estimation
at high spatial and temporal resolution. Downscaling or disaggregation
is a key method for enhancing the spatial-temporal resolution of satellite
data. The authors propose a non-parametric method involving a hybrid
Extreme Gradient Boosting (XGBoost) approach combined with multivari-
ate spatial-temporal Fuzzy clustering. This clustering relies on Integrated
Multi-satellite Retrievals for GPM (IMERG) precipitation and Shuttle Radar
Topography Mission (SRTM) Digital Elevation Data.

Experimentation was conducted to downscale 255 months of IMERG
satellite data over the Czech Republic. The utilization of eight stations for
training and validation, and 19 additional stations for validation, showed
a strong agreement with ground-observed precipitation. The proposed
methodology substantially enhances the accuracy of IMERG precipitation
data and holds promise for applications in other regions with remotely
sensed data, particularly where ground-measured station data is sparse.
Apart from precipitation, runoff plays a significant role in the water cycle.

The second study emphasizes the reconstruction of historical runoff
provides vital information for climate change adaptation. However, the
conventional hydrological models can exhibit bias, leading to the necessity
of machine learning-based computing models.

A novel Hybrid Ensemble Multi-Model Framework (HEMMF) was pro-
posed to reconstruct the gridded runoff of Europe over 500 years. The
HEMMF combines non-parametric extended data pattern recognition and
data-driven methods, employing nine different machine learning algorithms
and four ensembles of ML. Using different datasets, including Moran's spa-

5

tial autocorrelation (SPA) index and various climate-reconstructed data, the study found that the HEMMF offers the best reconstruction performance with a merged dataset.

The reconstructed runoff, made possible by this advanced methodology, helps explain runoff trends, drought propagation, and relationships with climate variables. The proposed methodology has wide-ranging applications, including potential usage for past hydroclimatic data and related analyses across diverse temporal periods, climate scenarios, and geographical scales.

Both methodologies play a crucial role in aligning and refining hydrological and meteorological observational records. Utilizing hybrid machine learning techniques for disaggregating precipitation and reconstructing historical runoff becomes particularly important when considering climate extremes. Notably, discerning trends on land becomes increasingly challenging in areas with limited data coverage. Addressing and bridging these data gaps is paramount. This ensures that such inconsistencies don't influence future projections, guiding our mitigation and adaptation strategies more effectively.

CONTENTS

List of Figures

vi

LIST OF TABLES

Part I

Background

CHAPTER 1

INTRODUCTION

Policymakers, scientists, and various stakeholders express a growing need for precise and dependable weather and climate-gridded data. These data are crucial for multiple applications such as managing water resources, as highlighted by Barnett et al. in 2005 [1], planning infrastructure based on insights from Brody et al. in 2007 [2], and constructing ecosystem models as shown by the IPCC in 2013. High-resolution precipitation data and historical gridded runoff records are crucial for hydrology, meteorology, policymakers and related fields, especially considering the ongoing climate change scenario driven by global warming [3]. Hydrologists and climatologists are highly interested in predicting watershed runoff caused by rainfall events [4].

Accurate precipitation data is indispensable for diverse sectors ranging from flood forecasting and water resource management to urban plan-

ning and ecosystem health [5, 6, 7, 8, 9]. Specifically, such data enables the quantification of rainfall's spatial and temporal characteristics, which is crucial for predicting flood events [5]. In water-dependent regions, this information informs the management of reservoirs, lakes, and rivers, effectively influencing water availability [6]. Concurrently, it supports agricultural decision-making processes, such as planting, irrigation, and harvest scheduling [10]. Urban infrastructure, particularly drainage systems and stormwater management, also relies on high-resolution precipitation figures [7, 8]. Moreover, a nuanced understanding of precipitation patterns is essential for maintaining ecosystem health and predicting habitat changes [9]. As climate change intensifies, the role of accurate precipitation data in monitoring long-term weather and runoff trends becomes increasingly critical [11].

The significance of historical runoff data in engineering and environmental assessments is undeniable. Engineers utilize this data to design infrastructural elements such as bridges, culverts, and dams, ensuring their resilience against past extremes and adaptability for anticipated environmental shifts [12]. These records are pivotal for flood risk evaluations, shaping floodplain management strategies and influencing insurance rates [13]. Furthermore, understanding past runoff trends is instrumental in forecasting groundwater availability—integral for regions grappling with water scarcity and renewable energy implications [14]. Such historical records furnish critical context to contemporary environmental occurrences, enabling researchers to discern between anomalous events and established long-term trends. The pressing need for precise rainfall and river flow data cannot be overstated in a world where environmental patterns profoundly impact human societies and their interactions with nature and surroundings are intrinsically intertwined. The demand for meticulous gridded data on precipitation and runoff escalates as global warming intensifies, yielding more pronounced climate alterations. This ensures safety, effective water management, and food security. This thesis, therefore, underscores the importance of disaggregating satellite precipitation, reconstructing historical runoff, and exploring their innovative hybrid methodologies. Detailed discussions of these facets are presented in sections 1.1, 1.2, and 1.3, which focus on addressing the existing research voids.

1.1 Disaggregation of precipitation

Understanding precipitation is essential to studying extreme hydro- meteorological events, irrigation, water resources, and runoff modelling. While rain gauges or automatic weather station networks are commonly used for precipitation measurements, their sparse distribution poses challenges in remote areas with extensive spatial coverage. Since rain gauge measurements are point-specific, they do not adequately represent the spatial variability of precipitation across larger regions. Hence, satellite-based estimates serve as a compelling alternative.

Numerous satellites data, including the Tropical Rainfall Measuring Mission (TRMM), Climate Prediction Center Morphing (CMORPH), Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS), and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), offer precipitation estimation. Among them, the Integrated Multi-satellite Retrievals for GPM (IMERG) stands out for its reliability in measuring precipitation [15, 16]. While IMERG provides precipitation data at a spatial resolution of $0.1°$x $0.1°$ (around 11 x 11 km at the equator), there's a need for higher resolution data to gain a more nuanced understanding of spatial rainfall variability and local-scale land-atmosphere interactions. The technique of spatial disaggregation addresses this by downscaling precipitation, providing more precise data for hydrological applications. This process helps to bridge the gap between resolution discrepancies and uncertainties, facilitating more localized climate change impact assessments.

Numerous studies have explored downscaling methods to enhance spatial resolution and diminish the uncertainty of satellite-based precipitation. They have done this by using vegetation indices such as the Normalized Difference Vegetation Index (NDVI) [17, 18], the Enhanced Vegetation Index (EVI) [18], and environmental factors like the Land Surface Temperature (LST) [17]. These studies often employ exponential regression (ER) and multiple linear regression (MLR) techniques. Yet, introducing ER and MLR can add further uncertainty and error to the downscaled precipitation [19]. To address the limitations of MLR, there is a shift towards the Geographically Weighted Regression (GWR) method [20]. Notably, NDVI and LST have been employed to downscale precipitation using GWR and

machine learning, complemented with ground truth data [17]. Chen et al. (2019) [20] noted that beyond a certain threshold, NDVI fails to capture precipitation patterns, leading to the adoption of EVI to mitigate this problem. However, vegetation indices prove problematic for monthly scale precipitation disaggregation, especially as leaves drop in seasons like autumn and winter [21]. Methods combining elevation and NDVI functions have proven effective in downscaling TRMM precipitation [22].

The local climate system, elevation, and atmospheric circulation significantly influence the spatial-temporal patterns of monthly precipitation [23, 24, 25, 26]. By using clustering techniques, climatic and atmospheric conditions are assessed based on specific and separate groups, aiding in prediction and numerical model simulations [27, 28]. Clustering of different time series patterns is based on intra-group similarities and differences between group patterns.. The hierarchical clustering approach has been employed for IMERG precipitation bias correction by grouping stations with similar climatic and geographical characteristics using precipitation quantiles [29]. Due to the limited availability of high-resolution precipitation data, the K-means spatial clustering method aids in the temporal disaggregation of extreme rainfall from daily to hourly scales in selected Indian metro cities [30]. Various studies have leveraged the Poisson cluster for spatial-temporal precipitation disaggregation [31, 32, 33]. However, combining clustering with regression techniques enhances the reliability of precipitation disaggregation [32].

1.2 Historical runoff reconstruction

Runoff denotes the volume of water originating from a specific land region (represented as a set of grid cells) that ultimately enters river systems. Such runoff is pivotal for evaluating water resources [34]. Presently, water resource scarcity, attributed to droughts [35] and influenced by climate change [36], impacts 66% of the worldwide population. The availability of water resources, influencing potable water, agriculture, industry, and other societal facets, is determined by runoff [37]. As a fundamental natural process on Earth, runoff is a crucial element of the hydrological cycle.

The runoff process influences the intricate and ever-changing water distribution patterns across the Earth. Runoff represents the result of the

inherent hydrological mechanisms associated with a river basin. This basin collects the cumulative runoff data from smaller regions nested within its boundaries. Typically, runoff information is accessible in the form of gridded fields within that river basin area. This data can be represented as spatial averages of these gridded fields corresponding to specific river basin territories.

To illustrate the long-term fluctuations in runoff fields, the Budyko models, as detailed in [38], have been established and adopted as a primary method for the description of long-term runoff variability [39, 40, 41, 42]. These models elucidate the connection between long-term inputs and outputs in many hydrological systems, offering hydrological time series with time frames exceeding a year. Their primary function is to delineate how water input, signified by precipitation, is divided between the two principal components of a hydrological system: actual evapotranspiration (E) and long-term runoff. The comprehensive portrayal of runoff, precipitation, and E relies on average yearly data collected over a standard three-decade observation period [43].

Budyko models primarily rely on long-term precipitation data and potential evapotranspiration (E_0) as inputs. The E_0 denotes the amount of atmospheric water demand when there's ample water available for evapotranspiration. Within Budyko models, these inputs are utilized to estimate both E and runoff, where E stands for actual evapotranspiration [42, 43].

Budyko models have been utilized to evaluate the implications of climate change on future predictions of runoff and evapotranspiration fields [43]. Furthermore, they have been employed to recreate past runoff from the 1970s to 2000s over twelve significant basins in China [44]. Various attempts to reconstruct historical climate data have predominantly relied on palaeoclimatic proxies. For instance, there are reconstructed datasets for precipitation [45] and temperature [46]. These reconstructions offer advanced insights into past spatial and temporal trends that are vital for Budyko models.

The patterns in space and time are strongly correlated with other significant environmental variables. Among these variables is data derived from the Palmer Drought Severity Index (PDSI), a metric grounded in a straightforward water balance model [47]. PDSI reconstructions have been

7

employed in multiple research projects to characterize historical hydroclimatic conditions [48, 49]. Additionally, the spatial correlation frameworks across all spatio-temporal fields offer another critical source of interconnected data.

The Budyko models frequently exhibit systematic errors due to various sources of uncertainties, such as parametric, structural, and uncertainties in spatial-temporal fields, among others [50]. Machine learning (ML) methods could offer viable ways to enhance long-term runoff reconstruction. These methods enable the creation of non-linear models with multiple inputs but a single output, factoring in data from diverse sources like PDSI, precipitation, and other environmental factors. A crucial consideration is how input data spatial autocorrelation (SPA) patterns are represented. Incorporating the SPA of the spatial fields of input data into ML models might boost their accuracy [51]. Yet, ML models fail to replicate the intricate hydrological processes, especially when the training data doesn't sufficiently capture climate variable extremes [52]. Merging hydrological models with ML techniques could be synergistic, paving the way for a hybrid ML strategy [52].

1.3   Hybrid modelling

Hybrid modelling combines predictions from physics-based dynamical weather or climate models with data-driven models to improve hydroclimatic predictability. While dynamical models numerically solve for system property changes over time, data-driven models range from simple linear regression to advanced AI methods, such as machine learning and deep neural networks. Hybrid prediction strategically merges the unique advantages of both dynamic and AI models. Hybrid modelling is gaining popularity in hydroclimatic areas due to its capability to form a robust predictive model [53].

Disaggregating satellite precipitation data and reconstructing historical runoff present a significant challenge using predictive machine learning or physically-based models due to uncertainty propagation in satellite or input data [28,

54]. This uncertainty also generates a bias in the output of the physical-based hydrological model [28]. However, this can be minimised using machine learning (ML) models, but in case the training data is extreme range unavailable, it leads to generating the bias in prediction output [52, 28]. The coupling of the physical hydrological model or statistical information with machine learning (ML) or ensemble of ML leads to the development of the hybrid model, which is a very robust model for runoff prediction or reconstruction as well precipitation disaggregation [52, 28, 54]. This coupling leads to complement each other (hydrological and ML models) to minimise the bias and is currently widely applied in the field of hydrology and climate data reconstruction [52, 54, 28].

1.4 The thesis aim

In today's world of rapid computational progress and increasing climate variability, it's crucial to have precise weather forecasting and resilient hydrological models. The present-day challenges brought about by climate changes emphasize the necessity for novel, holistic strategies to decode complex meteorological trends and retrace past climate records. Yet, despite the proven efficiency of hybrid modeling, only a handful of studies have ventured into reconstructing the 500-year historical runoff and disaggregating precipitation using such techniques. This doctoral research aspires to bridge this gap, harnessing advanced machine learning (ML) methods and hybrid ensemble models to craft frameworks tailored for precipitation disaggregation and retracing historical runoff trajectories.

To succinctly encapsulate the objectives of this thesis, this work turns focus to two pivotal case studies, as detailed below:

Development of a Hybrid Learning Model for Precipitation Disaggregation:

- To explore the potential of a hybrid Extreme Gradient Boosting (XGBoost) model combined with fuzzy multivariate spatial-temporal clustering for disaggregating IMERG satellite precipitation.

- To scrutinize the efficacy of XGBoost when coupled with multivariate clustering techniques and assess its potential in capturing com-

plex precipitation patterns.

- To augment the spatial resolution of the IMERG precipitation data, enhancing it from approximately 11 to 1 kilometer through integrated modelling.

Development of the Hybrid Ensemble Multi-Model Framework (HEMMF) for Historical Runoff Reconstruction:

- To develop the HEMMF, integrating spatial patterns correlation, hydrological Budyko models, and ensemble ML techniques for reconstructing historical spatio-temporal patterns of runoff fields over Europe for the period 1500–1999.

- To interconnect spatial patterns for palaeoclimatic reconstructions (like drought, precipitation, and temperature) with the estimation of historical runoff, utilizing Budyko models.

- To investigate the incorporation of Spatial Auto-correlation (SPA) patterns into ML models and its resultant influence on runoff prediction accuracy.

- To analyze different individual ML techniques' potentials and compare them against ensemble ML models to mitigate uncertainties inherent in the modelling process.

- To introduce and test a two-stage validation procedure for reconstructing palaeoclimatic patterns of runoff time series.

1.5 The structure of thesis

Chapters 2 and 3 delve into the intricacies of hybrid modelling and advanced clustering. Chapter 4 describes the evaluation metrics relevant to these models. Chapter 5 offers comprehensive details regarding the model implementation for the study. Chapters 6 and 7 outline the required data and present the results of the case studies. From Chapters 8 to 10, the outcomes of each case study are detailed, followed by an in-depth discussion and conclusion. Lastly, Chapter 10 shows the list of publications on this thesis work.

Part II

Hybrid modelling and advanced clustering

DISAGGREGATION OF PRECIPITATION

As this thesis work discusses above, "hybrid learning" refers to the amalgamation of multiple components, such as statistical or physics-based models, which often yield more realistic predictive outputs [28, 55]. However, as far as this work is aware, no existing studies delve into the disaggregation of precipitation using a hybrid method that combines the Extreme Gradient Boosting (XGBoost) model with fuzzy multivariate spatial-temporal clustering specifically for disaggregating IMERG satellite precipitation data. While previous studies have identified the effectiveness of XGBoost in disaggregating time-series precipitation data [56, 57], the model has not been paired with multivariate clustering techniques to maximize its benefits. The use of multivariate clustering provides a richer perspective on

intricate weather patterns, accounting for a web of interrelated meteorological and climatic elements. This approach enhances the accuracy of prediction models, leading to more reliable weather forecasts [58]. This chapter introduces a technique for disaggregating IMERG precipitation data by effectively combining two powerful tools: XGBoost and fuzzy spatial-temporal clustering, addressing the existing research gap. This pairing offers mutual benefits, enabling us to refine the spatial resolution of IMERG data from roughly 11 kilometers down to 1 kilometer.

2.1   Related work

Current disaggregation methods primarily utilize statistical techniques, drawing from high-resolution explanatory variables such as soil moisture, vegetation, evapotranspiration, and temperature to generate detailed precipitation information [59]. In this context, methods like partial least squares regression (PLSR), Artificial Neural Networks (ANN), and Random Forest (RF) have gained popularity [60, 61, 62]. Machine learning models, notably ANN, RF, and mixed geographically weighted regression (MGWR), have shown considerable promise [59, 63]. The MGWR model stands out for adeptly addressing both stationary and non-stationary characteristics of explanatory variables [64]. However, the patterns of these variables are influenced by local environments or climatic dynamics and often interrelate with each other [65]. The organization of these variables is represented by clusters of principal components [66]. Merging such components, like clusters, with regression techniques often leads to enhanced predictive accuracy [67].

The auto-searched Orographic and Atmospheric effects of De-trended Kriging, termed the ASOADeK regression method, is a current strategy for spatially downscaling precipitation in mountainous areas [68]. However, past efforts in this area haven't incorporated multivariate clustering, opting instead for data from Next Generation Radar (NEXRAD) precipitation fields. A notable limitation of this method is NEXRAD's inability to determine snow precipitation rates, as pointed out in Guan's 2009 research [68]. Additionally, the Microcanonical Multiplicative Random Cascade (MMRC) model, when paired with k-means clustering, is used for the temporal disaggregation of precipitation across specific Indian cities [30]. Regrettably,

Deka, P. et al. (2023) [30] work overlooks sensitive variables impacting precipitation for multivariate clustering. The Tropical Rainfall Measuring Mission (TRMM) 3B43 monthly precipitation satellite data has been broken down spatially using longitude, latitude, and elevation as predictors, and station data as the response variable within a multiplicative random cascade regression framework [69].

2.2 Hybrid modelling
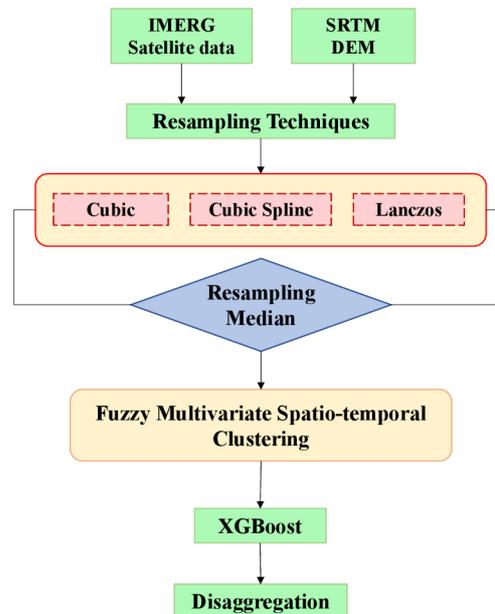
Figure 2.1: Flow chart of the methodology for disaggregation of precipitation.

The methodology's flowchart is illustrated in Figure 2.1. The thesis commences by procuring the IMERG and SRTM satellite data, followed by its resampling. For this, this work employs the 'Cubic', 'Cubic Spline', and

'Lanczos' resampling techniques on the IMERG and SRTM satellite data using the *resample()* function available in the 'terra' R package [70]. Although the nearest-neighbour method is typically favoured for resampling extensive raster data [29, 71], it has shortcomings. This method can yield images with a patchy appearance, potentially leading to the omission or duplication of pixel values [72]. Resampling serves as a tool to recalibrate the satellite data's resolution to my desired scale, specifically, one kilometer in this context. Through resampling, the research work enhances the quality of my satellite data, paving the way for effective disaggregation. Resampling in satellite imagery, disaggregation upholds local data while ensuring a balance in similarity and diversity among pixel values. Thus, this work chooses these specific resampling methods in the research.

2.2.1   Resampling methods

Cubic: This method involves cubic resampling. It determines the weighted average of the closest 16 pixels, yielding a more refined pixel output than the nearest-neighbour technique [73].

Cubic Spline: Utilizing cubic spline interpolation, this method smoothly integrates cubic polynomials to ensure seamless continuity between adjacent raster data pixels. Recognized for its efficiency, the cubic spline method maintains image clarity and often outperforms the nearest-neighbour and bilinear interpolation strategies [72, 74].

Lanczos: With the primary objective of retaining image intricacy during resizing, Lanczos resampling leverages a windowed $sinc$ function, termed the Lanczos kernel, for convolution. This approach notably minimizes aliasing disturbances while amplifying edge definition [75].

The aforementioned resampling strategies each present unique strengths. Consequently, research work computes the ensemble median of all these methods to harmonise and tap into their collective benefits. This ensures heightened precision in the enhanced resolution of IMERG and SRTM raster data. Post-resampling, this thesis work employs the resulting medians to formulate a multivariate cluster, the subsequent step. This clustering process is based on the resampled medians of IMERG and SRTM data originating from the Czech Republic.

2.2.2 Fuzzy multivariate spatio-temporal clustering

This work utilizes the Fuzzy spatial-temporal multivariate clustering approach to categorize similar patterns of time series precipitation and SRTM elevation data. The clustering leverages the Fuzzy C-Means algorithm, a concept pioneered by Bezdek JC [76]. Each cluster's centroid is ascertained using the Fuzzy C-Means (FCM) technique within this algorithm. Additionally, to gauge the 'distance' or disparity between time series data points, the algorithm incorporates the foundational Dynamic Time Warping (DTW) approach. This process is executed with the help of the "dtwclust" R package [77]. The C-Means clustering has showcased significant robustness when handling time series data [78].

This work has designated eight clusters based on the thesis work's comprehensive understanding of the Czech Republic's precipitation dynamics. These clusters collectively encapsulate the diverse spatial-temporal precipitation data as visualized in Figure 7.2 a). This classification strategy guarantees that each cluster contains one or more observation stations. Out of 27 ground observation stations aiding in training and validating the disaggregated precipitation, the most precise station corresponding to each cluster—contributing 80% of the data—is allocated for training. Notably, 20% of the data from the eight training stations, which remain outside the purview of model training, are designated for validation. Furthermore, comprehensive datasets from 19 other stations play a pivotal role in validating both the IMERG and the disaggregated precipitation via the XGBoost machine learning framework.

2.2.3 XGBoost ML model

Chen and Guestrin introduced the XGBoost model [79], drawing inspiration from the principle of boosting – the technique of morphing weak learners into potent ones. The methodology hinges on ensemble learning, where a succession of weak models (typically decision trees) are trained iteratively. Their collective predictions amalgamate to craft a final, potent model. The essence of boosting revolves around honing in on data points preceding models struggled to classify or predict correctly, subsequently ameliorating the aggregate performance through each iteration. XGBoost

stands out as a formidable machine learning technique for forecasting time series, even eclipsing the deep learning-based Long Short-Term Memory (LSTM) model under certain circumstances [80, 81].

CHAPTER 3

HISTORICAL RUNOFF RECONSTRUC-TION

Recent advancements in hybrid frameworks have demonstrated notable effectiveness in runoff estimation, as evidenced by various studies [82, 83, 84, 52]. My research introduces a comprehensive Hybrid Ensemble Multi-Model Framework (HEMMF), detailed in subsequent sections, dedicated to predicting and historically reconstructing continental-scale runoff. This innovative method seamlessly merges the strengths of process-driven hydrological models and machine learning (ML) techniques [84, 52]. Breakthroughs in hybrid ML and pattern recognition [85, 52] further pave the way for my ensemble ML model, which synthesizes insights from a range

of ML methodologies [86, 87].

The essence of my hybrid approach lies in its amalgamation of ML components, hydrological modelling [52], and ensemble ML techniques [88]. One can envision enhancing this model further by integrating an ensemble of ML models, SPA, and process-based hydrological models, a synergy seldom explored in the existing literature. The proposed architecture leverages ML, ensemble ML models, Budyko models, and SPA to heighten the precision of annual runoff reconstructions collaboratively.

In this research, my primary accomplishment has been formulating and evaluating HEMMF, aimed at recreating the historical spatio-temporal dynamics of runoff across Europe between 1500 and 1999. HEMMF synergistically merges spatial pattern correlations, data-driven SPA calculations, hydrological model estimations, and singular and ensemble ML techniques. Building upon the pioneering work of Konapala et al. (2020) [52], which advocated for the inclusion of Budyko model outputs into ML algorithms, this research has established a link between spatial patterns of past climate phenomena, such as droughts [48], precipitation [45], and temperature [46], with historic runoff estimations using Budyko models.

Additionally, my study broadens the data palette by incorporating SPA representations across all input time series. While the influence of SPA on ML-based time series predictions is undeniable [51], its inclusion in runoff forecasting through ML remains sparse. My runoff reconstruction utilizes a grid-based climate dataset, encompassing precipitation [45], temperature [46], and the Palmer Drought Severity Index (PDSI) [48], aggregated annually over Europe. The HEMMF strategy allowed us to gauge the potential of individual ML models, juxtaposing them with ensemble ML. To minimize model-related uncertainties, this work trialled a variety of ML ensemble configurations. This research introduces a dual-phase validation method for recreating runoff time series paleoclimatic dynamics.

3.1   Related work

Over recent decades, various ML models have been effectively utilized for pattern recognition in time series data and reconstructing runoff [89, 90, 91, 92]. For instance, the quantile regression forest (QRF) was employed

across 511 U.S. basins to predict daily streamflow [93]. In this research, a QRF model corrected the streamflow simulated by a physics-driven hydrological model using precipitation and E_0 as inputs. Another study by Zhang, Yuhang, et al. [94] employed QRF for streamflow estimation in 522 sub-basins of the Yangtze River. These studies' results affirmed the QRF model's enhanced accuracy in daily streamflow prediction [94, 93].

S. Gangopadhyay et al. [95] implemented the K-nearest neighbors (KNN) for annual streamflow reconstruction on the upper Colorado River, using a tree ring proxy for the period from 1400–2005. Their findings showed that the reconstructed streamflow closely aligned with observed data during the validation phase [96, 95]. Additionally, KNN was integrated into a hybrid ML model to predict water flow reliably in the Indus river basin [97].

The extreme gradient boosting (XGBoost) model was employed to forecast a 10-day streamflow for China's Three Gorges Dam [98]. This model utilized 10-day inflow data and showed efficacy in forecasting from 1990 to 2015. The XGBoost model also delivered satisfactory results for daily streamflow simulations [99].

Additionally, the multivariate adaptive regression spline (MARS) was utilized for monthly streamflow forecasting in locations like the Tigris River in Baghdad and the mountainous Swat river basin in Pakistan [100, 101]. MARS demonstrated promising potential for streamflow predictions, and its hybrid version further improved accuracy [102]. However, it's notable that limited studies leverage ML models like QRNN, RPART, RLM, CTREE, and GBCSS for streamflow time series pattern reconstruction. Methods like QRF, XGBoost, and MARS haven't been extensively explored for paleoclimatic historical runoff reconstruction.

ML model accuracy for time series runoff simulation has improved, especially with the hybrid ML model that uses outputs from physics-driven hydrological models [103, 52]. For instance, the hybrid approach, which integrates ML and hydrological models using various meteorological inputs, has proven superior in estimating runoff compared to solely relying on physics-based models or data-driven algorithms [104, 105]. Although this hybrid framework hasn't been extensively tested for historical runoff, it has exhibited reduced biases in simulations [52].

Another evolution in the modeling landscape is ensemble ML techniques [106, 107, 108]. These methods combine outputs from multiple ML models to produce a joint output. Many studies have demonstrated that ensemble-focused approaches offer superior prediction accuracy compared to individual ML models [109]. For example, Tyralis et al. [110] showcased the effectiveness of an ensemble approach, integrating models like LM, MARS, XGBoost, and NN, for runoff forecasting. Ensemble methods have also shown promise in reconstructing historical patterns of paleoclimatic data [111, 112, 113]. Collectively, these ensemble models have consistently elevated the accuracy of hydrological time series predictions in multiple studies [114, 115, 116, 117, 118].

3.2 Hybrid ensemble multi-model framework
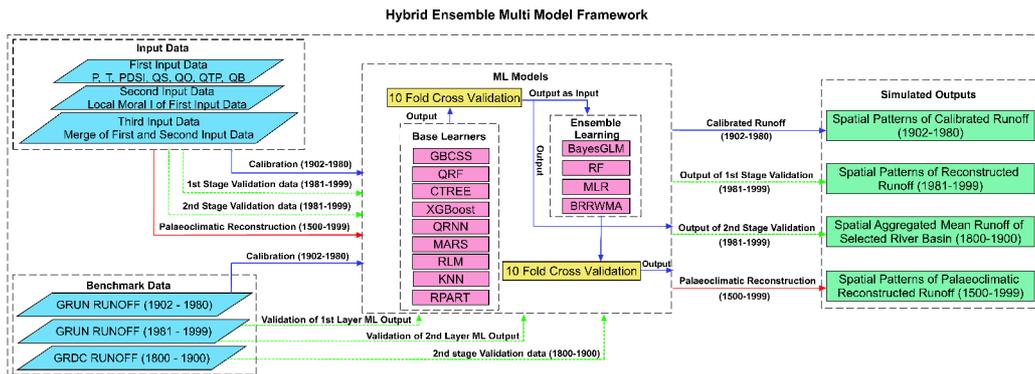
Figure 3.1: Methodological flowchart. The initial dataset incorporates seven-time series inputs: precipitation, temperature, PDSI, along with runoff from using Budyko models (QS, QO, QTB, and QB) that are described in the works of Schreiber [39], Ol'Dekop [40], Turc-Pike [41, 42], and Budyko [119]. The subsequent datasets represent the primary dataset's Spatial Autocorrelation (SPA). The final dataset is an amalgamation of the first two. The Global Runoff Reconstruction (GRUN) and Global Runoff Data Centre (GRDC) datasets serve as reference points for a specific timeframe.

The hybrid framework blends various components for enhanced ef-

ficiency, integrating expanded input data patterns with a set of machine learning algorithms, as depicted in Fig. 3.1. This diagram illustrates the three-stage implementation of my HEMMF proposal.

In the initial stage, HEMMF's extended input data patterns draw upon climatological fields (see Section 6.2), the spatial autocorrelation of input data (detailed in Section 3.4), and outputs from parsimonious hydrological models that are physics-based (refer to Section 3.3). While these input data are in raster forms, they transform vector data, readying them for the subsequent phase of HEMMF's ML models.

During the second phase, the case study tests the HEMMF using extended input data patterns and a duo of ML model types: base learners and ensemble learning, incorporating 10-fold cross validation as depicted in Fig. 3.1. Here, base learners get their input from the extended data patterns, with nine specific ML models processing the vector data to simulate runoff. The base learner's predictions are then fed into ensemble learning to fine-tune the runoff simulation. A deep dive into these nine ML models nested within the base learners, and the ensemble models tested in the HEMMF's ensemble learning can be found in Sections 3.2.1 and 3.2.2 respectively.

In the concluding third phase, simulations are calibrated, cross-checked, and reshaped concerning the chosen ML ensemble and benchmark data, as elaborated in Sections 3.2.1 and 3.2.2.

The selection of the nine ML models and four ensemble methods was made judiciously from a plethora of modern ML and ensemble techniques. The primary criterion was to ensure a rich diversity of high-performing ML methodologies [120, 121]. The ensemble's composition hinges on selecting base learners that collectively present a complementary performance [122]. The exact number of these learners varies based on factors like data patterns, the problem's nature, and ensuring diversity and accuracy [122, 123]. Odd numbers of base learners have historically proven more effective for ensemble learning [124, 125, 126, 127]. Thus, my choice of nine base and four ensemble learners was deliberate, aimed at optimizing each ensemble model's performance, emphasizing algorithmic diversity and similarity. This approach aids in assessing ensemble output variations and congruence. Among the chosen ensemble models, multi-linear regression and

random forest have previously proven their robustness [128]. However, combining the ML ensemble with the Bayesian generalized linear model and Bayesian ridge regression in a model average is a novel approach. Opting for a limited ensemble structure in my hybrid framework is strategic, aiming to reduce computational overhead [121].

3.2.1 Base learners

HEMMF's ML models employ climatic and hydrological data to produce yearly runoff, with each ML model rendering unique features in reconstructing the runoff pattern. The nine ML models serving as HEMMF's base learners are outlined below:

Quantile Regression Neural Network (QRNN): This model, rooted in the standard multilayer perceptron-based ANN [129], synergizes the advantages of both neural networks and quantile regression [130]. QRNN, a versatile form of quantile regression, was pioneered by Taylor and Cannon [130, 131]. QRNN incorporates a differentiable approximation to its regression cost function, pivotal to enhancing parameter estimations through the finite smoothing algorithm [132]. This approximation also rectifies expectile regression or conventional least squares regression issues.

Gradient Boosting with Component wise Smoothing Splines (GBCSS): Introduced by Buhlmann and Yu [133], GBCSS utilizes a boosting algorithm that centers on component-wise base learners. It employs P-splines with B-spline basis base learners [134], optimizing prediction accuracy while averting overfitting [135].

Recursive Partitioning and Regression Trees (RPART): Crafted by Breiman et al. [136], RPART continuously divides a dataset until a set termination criterion emerges. Each division focuses on the independent variable, aiming to minimize variability in the predicted variable. This method's impurity, a measure of leaf node heterogeneity, is quantified using the Gini index.

Conditional Inference Tree (CTREE): CTREE, also termed unbiased recursive partitioning, was put forth by Hothorn, Zeileis, and Strobl [137, 138, 139]. It estimates the regression relationship via binary recursive par-

titioning within a conditional inference framework. The CTREE process comprises a series of tests and splits, recursively evaluating the relationship between input and response variables.

EXtreme Gradient Boosting (XGBoost): Chen and Guestrin introduced XGBoost in 2016 [79]. It's a conversion technique transforming a weak learner into a strong one, where the former marginally surpasses random guesses and the latter boasts high accuracy. At its core, XGBoost leverages a gradient boosting tree as its booster.

Multivariate Adaptive Regression Spline (MARS): Conceived by Friedman in 1991 [140], MARS automatically formulates a segmented linear model, proficiently capturing the nonlinear characteristics of polynomial regression by interpreting knots as step functions.

Robust Linear Model (RLM): An extended linear regression variant, RLM integrates M-estimation using the Huber method [141]. RLM employs iterated re-weighted least squares (IRLS) for model fitting and stands out for its efficacy in data with outliers.

k-Nearest Neighbors (KNN): Presented by Huber [142], KNN categorizes data based on the features of neighboring data points [143]. Using kernel functions, KNN emphasizes neighbours according to their proximity, making predictions rooted in local approximations.

Quantile Regression Forest (QRF): An augmentation of the random forest, QRF broadens the full conditional spread of a predictive variable, aiding outlier detection [144]. While a random forest furnishes comprehensive details about the mean conditional distribution of the response variable, QRF refines predictions of conditional quantiles for predictor variables.

3.2.2 Ensemble learning

The ensemble learning approach in HEMMF integrates four distinct ensemble ML models. These models use the results from their base learners as inputs, refine these results, and deliver error-adjusted runoff predictions. A description of each ensemble ML model follows.

Multilinear Regression (MLR) is a fundamental model employed to delineate the linear correlation between input and output parameters.

Random Forest (RF) operates in a manner akin to bagging but with specific modifications [145]. Multiple trees are generated within RF using varied bootstrapped samples from the input data. The RF model then enhances the model's robustness by averaging the predictions from all trees, a method that also helps in preventing overfitting. The methodology of Breiman's random forest is employed in this study [146].

Bayesian Generalized Linear Model (BayesGLM) operates similarly to generalized linear regression. It's grounded in a Gaussian family linear model and produces credible interval statements. Additionally, it imposes a prior restriction using a t-distribution [147]. The fitting of the BayesGLM is achieved using the modified expectation-maximization (EM) and the pseudo-data technique [148].

Bayesian Ridge Regression with Model Averaged (BRRWMA) forecasts the regression coefficients based on posterior evaluations. The resultant regression model averages out predictions using posterior estimates, aiming to address model uncertainty [149]. This machine learning model uses the default parameter settings found in the original "monomvn" R package [150]. For HEMMF, various R packages are employed, with details provided in the supplementary materials.

3.3 Physical hydrological models

In my endeavor to reconstruct the spatial runoff field, this study utilizes four contemporary Budyko models. These models provide estimates of long-term water balance, focusing on the interplay between long-term annual precipitation and E_0 [151, 152].

The fundamental tenets of these models can be articulated as:

$$Q = P - E, \quad \text{and} \quad E = f(P, E_0). \tag{3.1}$$

Here, P, Q, E, and E_0 signify annual precipitation, runoff, actual evapotranspiration, and potential evapotranspiration, respectively [153].

The Budyko models' potential evapotranspiration (E_0) is deduced from

Table 3.1: Overview of Budyko hydrological models adopted in this research.

Budyko Model Evapotranspiration	Model Name	Reference	Originator
$E = 1 - \exp\left(-\dfrac{E_0}{P}\right)$	QS	[39]	Schreiber
$E = \dfrac{E_0}{P} \tanh\left(\dfrac{E_0}{P}\right)^{-1}$	QO	[40]	Ol'Dekop
$E = \dfrac{\dfrac{E_0}{P}}{\left[\left(\dfrac{E_0}{P}\right)^2 + 1\right]^{\frac{1}{2}}}$	QTP	[41, 42]	Turc,Pike
$E = \sqrt{\dfrac{E_0}{P} \tanh\left(\dfrac{E_0}{P}\right)^{-1}\left[1 - \exp\left(-\dfrac{E_0}{P}\right)\right]}$	QB	[119]	Budyko

temperature, as proposed by the Oudin method [154]. Table 3.1 concisely represents the chosen non-parametric Budyko models. This work incorporates four such non-parametric models in my study to estimate runoff [119, 39, 40, 41, 42].

3.4   Spatial autocorrelation

All raster-based inputs within the HEMMF dataset exhibit varying SPA patterns. This research work employ the Local Moran's I, a local indicator of spatial association (LISA) devised by Anselin in 1995, to encapsulate SPA [155].

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, i \neq j}^{n} w_{ij} \left(x_j - \bar{x}\right), \tag{3.2}$$

27

where,

$$S_i^2 = \frac{\sum\limits_{j=1,i\neq j}^{n} w_{ij}(x_j - \bar{x})^2}{N-1}. \tag{3.3}$$

In this research, I_i symbolizes the Local Moran's I for the grid cell i. The value of the grid cell i is denoted by x_i, while $\bar{x}$ represents the mean value across all grid cells. The spatial relationship between the grid cells i and j is captured by $w_{i,j}$. The term n signifies the total grid cell count. S_i^2 defines the cumulative spatial weight of these grid cells. In my research, This finding employed a 3×3 neighbourhood spatial weights matrix to determine the "Queen's case" Local Moran's indices.

These Local Moran's indices map out the spatial layout of all values across grid cells, serving as estimated input variables. The spatial fields show the autocorrelation for each dataset or attribute span an identical grid count as their input spatial field. In the context of my research, these attributes are depicted through the spatial layouts of precipitation, temperature, PDSI, and annual runoff, as deduced by four distinct Budyko models.

3.5   Two-stage validation

The evaluation of the reconstructed runoff undergoes a two-stage validation process across two distinct time frames. In the first stage, from 1901 to 1999, the reconstructed runoff is assessed against the spatio-temporal patterns of grids relative to the GRUN datasets. Meanwhile, in the second stage covering 1800 to 1899, the runoff evaluation is based on grids spatially combined over the chosen river basin in relation to the GRDC datasets.

During each validation phase, this study determines evaluation metrics for every ML model. The foundational metrics for these learners include Root Mean Square Error (RMSE) [156], Mean Absolute Error (MAE) [156], and Coefficient of Determination (R^2) [156]. This research employs RMSE, MAE, R^2, and KGE metrics for training and validating ensemble learning outcomes using the first two input datasets. In contrast, ensemble learning results from the third input dataset are validated using two specific indices: spatio-temporal accuracy and the quality of fit for the reconstructed runoff. These are represented by the normalized root mean square error (NRMSE)

and Kling–Gupta Efficiency (KGE) [157]. The metrics utilized in the second stage evaluation, when applying the third input data, are RMSE, MAE, R^2, and KGE. These evaluation metrics have been described in Chapter 4 (Models evaluations metrics)

CHAPTER 4

MODELS EVALUATIONS METRICS

Evaluation metrics serve as quantitative indicators that gauge the effectiveness of models or algorithms, especially within machine learning, statistics, and data analytics. These indicators aid in understanding the proficiency of a model or algorithm relative to a specific standard or reference point. The relevant metric often depends on the task's nature, such as classification, regression, clustering, and so on.

4.1 Metrics

RMSE has been utilized to assess accuracy on a grid-by-grid basis when compared to benchmark data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.1}$$

Here, y_i represents the benchmark value, $\hat{y}_i$ denotes the reconstructed

value for the ith grid cell, and n stands for the overall count of non-missing grid cell values.

R^2: This coefficient of determination is employed to gauge the correlation magnitude between benchmarked and predicted values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.2}$$

Here, y_i symbolizes the benchmark value, $\hat{y}_i$ stands for the reconstructed value, $\bar{y}$ is the average of reconstructed values, i refers to the specific grid cell, and n indicates the total count of non-missing grid cell values.

MAE: This metric quantifies the average absolute disparity between benchmark and reconstructed data.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4.3}$$

In this context, y_i is the benchmark value, $\hat{y}_i$ signifies the reconstructed value, and i denotes the specific grid cell, with n as the total grid cells count.

NRMSE: This study employs NRMSE to assess temporal accuracy in relation to benchmark data.

$$NRMSE = \frac{RMSE_{EM}}{SD_{GRUN}} \tag{4.4}$$

Where, $RMSE_{EM}$ and SD_{GRUN} represent the temporal RMSE of the HEMMF output and the temporal standard deviation of GRUN data, respectively. A higher NRMSE value indicates greater bias, while a lower one implies reduced bias in model output.

KGE: Introduced by Gupta et al. in 2009, the Kling–Gupta Efficiency (KGE) index evaluates the goodness of fit [157].

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \tag{4.5}$$

Here, r is the linear correlation between simulated and observed runoff, σ_{sim} and σ_{obs} are the standard deviations of simulations and observations respectively, while μ_{obs} and μ_{sim} are the means of observations and simulations. KGE values can range from $-\infty$ to 1, with 1 indicating perfect model accuracy and larger negative values revealing a more biased model output. KGE is seen as a multi-objective criterion.

NSE (Nash-Sutcliffe Efficiency) is a commonly employed metric for gauging the efficacy of hydrological models. Introduced by J. E. Nash and J. V. Sutcliffe in their work [158], the NSE is formulated as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.6}$$

In this equation, y_i symbolizes the actual observations, $\hat{y}_i$ stands for the predicted values from the model, $\bar{y}$ is the average of the actual observations, and n represents the total count of observations. The NSE value can range from $-\infty$ to 1. A value of 1 implies an impeccable alignment between predictions and actual observations, whereas values drifting away from 1 signify deviations in the model's predictions.

IMPLEMENTATION

ML execution necessitates the careful configuration of hyperparameters to optimize predictive accuracy. Hyperparameters are pre-defined settings that govern how an algorithm learns from data. Unlike model parameters which are learned during training, hyperparameters must be set prior to this phase. Selecting the right combination can significantly influence model performance. Hence, systematic approaches such as grid search or random search are often used to test various hyperparameter combinations to ensure the most accurate and efficient predictions for the disaggregation of precipitation and historical runoff reconstruction.

5.1   Models execution

For the disaggregation of precipitation, we utilized the "xgboost" R package [79], exploring a spectrum of hyper-parameters: $nrounds = 50, 100, 200,$ $max\ depth = 2, 5, 10, eta = 0.1 - 0.3$ with increments of 0.1, gamma =

0, colsample bytree = ranging from 0.5 to 1.0 with an increment of 0.1, min child weight spanning from 1 to 10, and *subsample* set between 0.5 to 1.0, increasing by 0.05. The best-fit hyperparameters correspond to the lowest Root Mean Square Error (RMSE) as determined by ten-fold cross-validation in both case studies.

In the second case study, three distinct datasets are used (detailed in Chapter 6, section 6.2) to assess the robustness of HEMMF. The first dataset incorporates seven raster variables (including temperature, precipitation, PDSI, and runoff from four Budyko models). The second is solely the SPA of the initial dataset, while the third merges datasets one and two. All these datasets comprise 98 raster layers, showcasing spatial-temporal maps from 1902 to 1999. Fig. 3.1 illustrates HEMMF's process, detailing how various inputs, benchmarks (GRUN and GRDC), and ML models are integrated.

Spatially, our proposed approach is applied to every HEMMF ML model for runoff reconstruction using the three datasets. Training for these ML models uses 79 raster layers (spanning 1902–1980) for each input dataset, processed on a grid cell-by-cell basis. For validation (from 1981 to 1999), this case study set aside specific amounts of data from each dataset. Opting for a spatial approach ensures ML models receive adequate data. In contrast, a temporal model would provide limited data for training and validation per cell, potentially compromising model robustness.

Nine primary learners are trained with the three datasets, their cross-validated predictions serving as inputs for the four ensemble ML models within the HEMMF. Utilizing 10-fold cross-validation, each dataset is segmented into ten portions, with one acting as a test set for the other nine. Model accuracy hinges on the average error from the tested portions. Various metrics, including RMSE, R^2, and MAE, gauge each model's efficiency and accuracy.

All ensemble models predict HEMMF's final outputs. Leveraging extended input patterns and SPAs, the HEMMF effectively captures Europe's runoff pattern. Training incorporates GRUN data from 1902–1980, while validation (1981–1999) uses the expanded dataset. The HEMMF's assessment also includes a century's worth of river basin runoff, detailed further in supplementary materials.

The underlying assumption is that if HEMMF accurately captures data

spanning 200 years, it should equally effectively reconstruct data for the preceding 300 years. This is premised on the large dataset's capacity to offer a comprehensive runoff pattern view, making ML models more stable, as referenced in various studies.

Regarding hyperparameters, we've primarily used the default method from the R package for efficient ML algorithm performance at reduced computational costs. We've employed a grid search for default hyperparameter determination, adjusting ranges as necessary until attaining the desired performance. Both "caret" and "caretensemble" R packages serve this purpose, with specific hyperparameters detailed in the supplementary material, Table 5.1. Notably, the BayesGLM and BRRWMA models, as implemented through the "caret" package, don't require hyperparameters.

5.2 Models Execution R Packages

This study utilizes the default hyperparameters of the "caret" [159] and "caretensemble" [160] packages to execute each ML model and an ensemble of ML in the R programming language.

ML models, including QRNN, GAMBoost, RPART, CTree, XGBoost, MARS, RLM, KNN, and QRF, are implemented using respective R packages: "qrnn" [161], "mboost" [162], "rpart" [163], "ctree" [164], "xgboost" [165], "earth" [166], "MASS" [167], "kknn" [168], and "quantregForest" [144].

Ensemble models like RF, BayesGlm, and Bayesian Ridge Regression with Model Averaged are implemented using the "randomForest" [169], "arm" [170], and "monomvn" [150] packages, respectively.

5.3 HPC utilisation

This study harnesses the immense computational power of the Czech National Grid Infrastructure (NGI), made possible through the cutting-edge facilities provided by Massive Computations (MetaCentrum) resource. To conduct our comprehensive case studies, this thesis leverages the capabilities of a parallel computing cluster, employing anywhere from 20 to 80 cores within the available RAM range of 100 to 7000 GB. It is worth not-

Table 5.1: Table shows each model's range value of hyperparameters with an optimal model with the minimum value of RMSE.

Models	Hyperparameter	Explored range of value	Optimal model where minimum value of RMSE			R package
			First input data	Second input data	Third input data	
QRNN	n.hidden	1, 3, 5	5	5	5	qrnn
	penalty	1e-04 to 0e+00	0	0	0	
	bag	constant at a value of FALSE	FALSE	FALSE	FALSE	
GBCSS	mstop	50, 100, 150	150	150	150	mboost
	prune	constant at a value of no	no	no	n0	
RPART	cp	0.08144097 to 0.52457837	0.08144097	0.01783384	0.08529554	rpart
CTREE	mincriterion	0.01, 0.50, 0.99	0.5	0.5	0.5	party
XGBoost	nrounds	50, 100, 150	150	150	150	xgboost
	max_depth	1,2,3	3	3	3	
	eta	0.3 to 0 .4	0.4	0.4	0.4	
	gamma	constant at a value of 0	0	0	0	
	subsample	0.50 to 1	1	1	1	
	colsample_bytree	0.6 to 0.8	0.8	0.8	0.8	
	rate_drop	0.01 to 0.50	0.01	0.01	0.01	
	skip_drop	0.05 to 0.95	0.95	0.95	0.95	
	min_child_weight	constant at a value of 1	1	1	1	
MARS	nprune	2 to 17	11	11	17	earth
	degree	constant at a value of 1	1	1	1	
RLM	intercept	FALSE, TRUE	FALSE	TRUE	FALSE	MASS
	psi	huber, hampel, bisquare	huber	hampel	huber	
KNN	kmax	5 to 9	9	9	9	kknn
	distancel	constant at a value of 2	2	2	2	
	kerneL	constant at a value of optimal	optimal	optimal	optimal	
QRF	mtry	2 to 14	7	4	14	quantregForest
RF	mtry	2,5,7	2	2	2	randomForest
LM	intercept	TRUE	TRUE	TRUE	TRUE	
BayesGLM						arm
BRRWMA						monomvn

ing that, in the context of R programming, the number of cores used directly influences the RAM requirements. This work is executed through a meticulously designed "qsub" shell script, ensuring efficient and high-performance computation.

Part III

Data

<div style="text-align: right">

CHAPTER 6

</div>

REQUIRED DATA

Runoff reconstruction and precipitation disaggregation are advanced hydrological techniques to refine water flow and rainfall data for improved water resource management. Runoff reconstruction involves using available hydrological data to estimate river discharge or surface runoff patterns, often in areas or periods with missing data. Precipitation disaggregation, on the other hand, entails breaking down coarse temporal or spatial rainfall data into finer scales, enhancing its utility for localized or detailed analyses. The accuracy of these methods heavily relies on the quality and reliability of input data, such as precipitation records, satellite observations, ground-measured station data, and digital elevation models, ensuring that results reflect actual hydrological conditions.

6.1   Precipitation and its linked variable

The disaggregation accuracy depends on the quality and reliability of input data [171]. Satellite observations of precipitation from IMERG, which are less biased, are being used in this study. IMERG offers a global precipitation

product with a spatial resolution of 0.1 degrees and is available in half-hourly, daily, and monthly temporal resolutions. This study employs the IMERGV06 final run (gauge corrected) on a monthly scale. This product maintains high accuracy compared to ground-measured stations [172] and is available from June 2000 onwards. Precipitation patterns are observed to have a direct link with elevation, as highlighted in the study by [173]. Given this relationship, elevation data is being incorporated into my analysis.

The study uses the Shuttle Radar Topography Mission (SRTM) Digital Elevation Data Version 4, accessed online at [174]. This dataset, available from the 11th to the 22nd of February 2002, provides high-quality, global-scale data with a spatial resolution of 90 meters.

In-situ data from twenty-seven rain-gauge locations spread across the Czech Republic (Chapter 7 Figure 7.2 *a*) serve to validate the disaggregated precipitation datasets. The study assumes that there isn't any heterogeneity in precipitation at a one-kilometer spatial resolution. Disaggregated precipitation undergoes validation at twenty-seven stations, matching each pixel within its spatial coordinates from 2000 to 2021.

In predictor variables, this work considers climate-reconstructed gridded data represented by spatial fields of precipitation [45], temperature [46], and Palmer Drought Severity Index (PDSI) [48]. This research reconstructs precipitation, temperature, and PDSI through principal component analysis and point-by-point regression. I take all the aforementioned raster input data for 1500–1999 with a spatial resolution of 0.5 degrees.

This study uses the Global Gridded Runoff dataset (GRUN) [175] as the first benchmark data for HEMMF. I employ the GRUN benchmark data runoff fields as the target variable for the European continent available for 1902–1999 with a 0.5-degree spatial resolution and a temporal resolution of one month. This research derives this benchmark product using the RF algorithm with precipitation and temperature as predictor variables. The target variable is the in-situ measured streamflow observations available at the Global Streamflow Indices and Metadata Archive (GSIM) [175, 176]. This Archive provides station-based data with detailed information on spatio-temporal runoff at a global scale from 1902 to 2014.

The proposed HEMMF utilizes a different benchmark data set, which I refer to as the 'second benchmark data'. This data set provides a unique

historical discharge time series, representing spatially averaged river basin runoff. To validate the proposed HEMMF, I employ observed runoff time series data from 16 distinct river basins. These observed data were sourced from the GRDC database and are presented on a monthly scale [177]. However, as I focus on reconstructing annual runoff data, I have aggregated these monthly datasets into yearly time steps. Detailed information about the second benchmark data can be found in Table 7.7.

Table 6.1: Metrics for evaluating BayesGLM model output using third input data from 1800 to 1900, compared to GRDC river basin data.

GRDC NO	Station	River Basin	Catchment Area (km^2)	R^2	RMSE(mm/year)	MAE(mm/year)	KGE
6123300	Blois	Loire	40630.00	0.60	0.61	47.00	0.74
6335020	Rees	Rhine	159424.00	0.64	50.00	36.00	0.66
6335060	Kolen	Rhine	144232.00	0.71	42.00	32.00	0.74
6335301	Schweinfurt	Main	12715.00	0.47	71.00	56.00	0.39
6337100	Vlotho	Weser	17618.00	0.55	78.00	65.00	0.55
6335500	Wuerzburg	Main	14031.00	0.45	49.00	37.00	0.55
6337200	Intschede	Weser	34424.00	0.52	68.00	54.00	0.49
6337400	Hann.-Muenden	Weser	12442.00	0.61	44.00	36.00	0.67
6337514	Bodenwerder	Weser	15924.00	0.60	55.00	45.00	0.64
6340120	Dresden	Labe (Elbe)	53096.00	0.41	65.00	55.00	0.56
6343100	Wasserburg	Inn	11983.00	0.14	120.00	95.00	0.34
6343500	Burghausen	Salzach	6649.00	0.17	346.00	301.00	0.22
6742200	Orsova	Danube	576232.00	0.45	101.00	92.00	0.55
6935051	Basel	Rhine	35897.00	0.51	102.00	77.00	0.60
6935052	Basel, Schifflaende	Rhine	35905.00	0.51	102.00	77.00	0.60
6974150	Smalininkai	Nemunas - Neman	81200.00	0.11	43.00	34.00	0.31

Part IV

Results - case studies

Chapter 7

Case studies

In this PhD thesis, two pivotal case studies are presented: the first delves into the result of precipitation disaggregation over the Czech Republic, while the second focuses on runoff reconstruction across the European continent using hybrid modelling. The initial section provides a detailed description of the outcomes of the precipitation disaggregation study. This

is followed by exploring the results of the runoff reconstruction in the subsequent stage.

7.1 Precipitation disaggregation over the Czech Republic

This section shows each cluster's time series patterns of disaggregated precipitation data. Following this, the thesis presents the training and validation accuracy of the XGBoost model using designated evaluation metrics. Lastly, this chapter illustrates an evaluation of both IMERG and the disaggregated data compared with respect to ground-based observational station datasets.

7.1.1 Clustering

The IMERG data consists of 1096×255 usable pixels after removing invalid or NaN values. Each pixel is about 11 kilometers apart. The 255 refers to monthly data layers from June 2000 to September 2021. By using a median ensemble resampling method, the resolution improves to one kilometer, expanding the pixel count to 119851×255. This data is then split into eight groups. The pixel counts for these groups, from Cluster 1 to 8, are $19848 \times 255, 6157 \times 255, 18077 \times 255, 17491 \times 255, 2848 \times 255, 22953 \times 255, 12757 \times 255$, and 19684×255.

There's a time series plot for each pixel, as shown in Figure 7.1. By comparing this with elevation details in Figures 7.1 and 7.2 parts a), b), and c), it's clear that areas in the Czech Republic with higher elevations get more rain. Table 7.1 lists the range values of IMERG, disaggregated rain, and elevation for each cluster. This table highlights that while IMERG data does not change much with elevation, disaggregated rain rates vary across clusters.

Table 7.1: Disaggregated precipitation, IMERG, and elevation variation for each cluster.

Cluster	Elevation (m)	IMERG (mm/month)	Disaggregated (mm/month)
1	112-272	0.25-384	10-107
2	711-890	0.25-382	18-176
3	506-593	0.25-383	9-302
4	260-345	0.29-382	12-255
5	890-1528	0.26-375	29-477
6	429-507	0.26-375	11-182
7	593-714	0.26-383	13-157
8	344-430	0.27-381	12-272
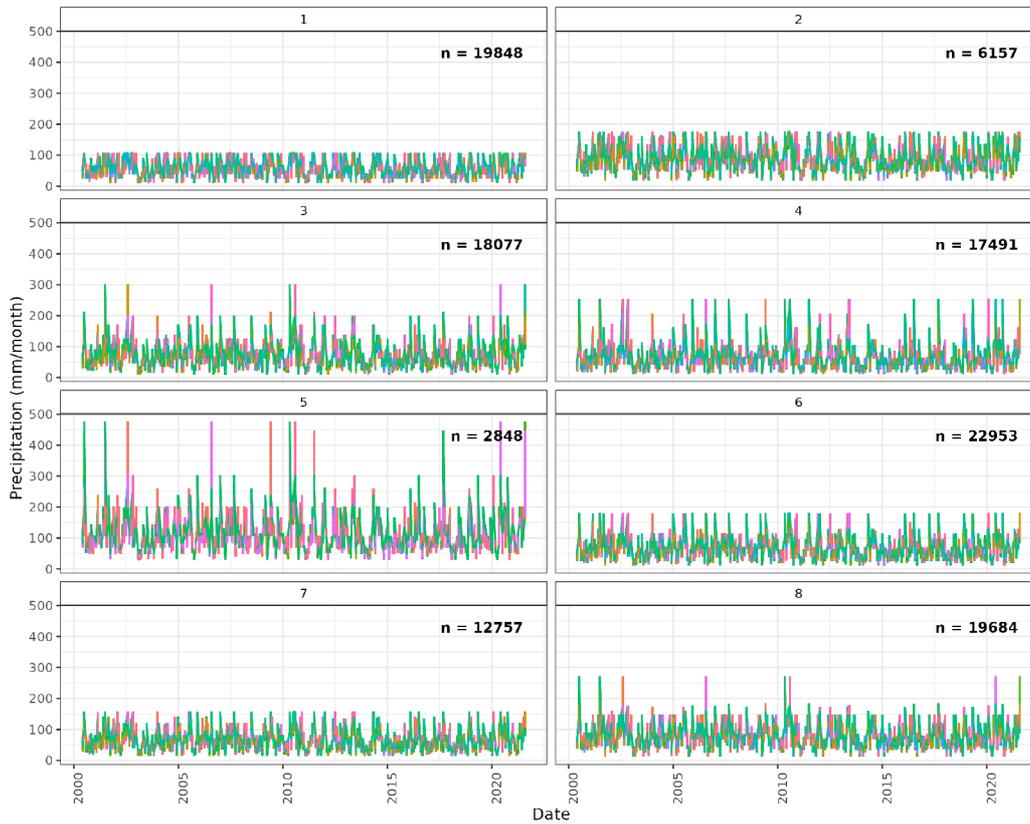
Figure 7.1: Time series graph of fuzzy spatial-temporal clusters using disaggregated rainfall data. Every pixel in a particular cluster is shown in this series, with 'n' indicating the count of raster pixels in each cluster.

51
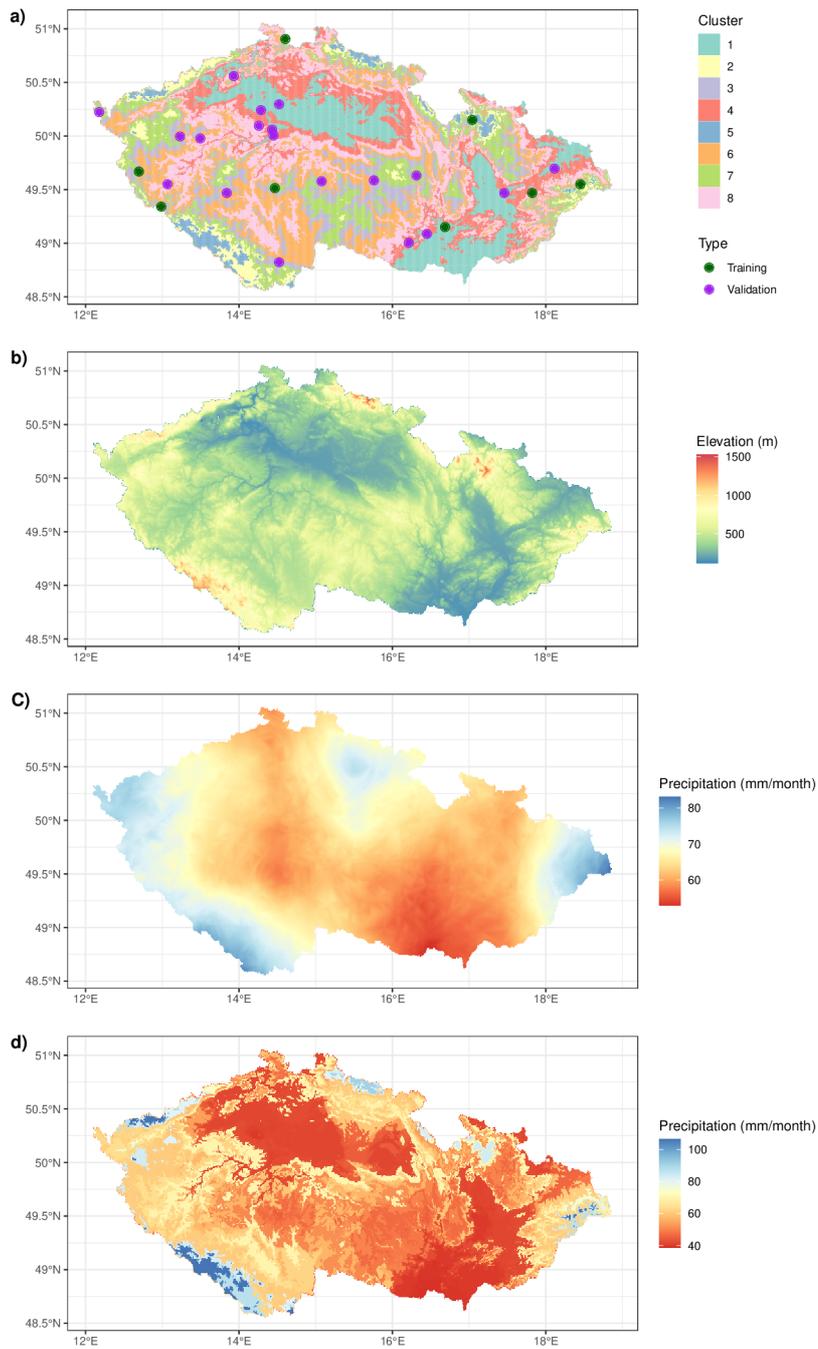
Figure 7.2: Figures marked as $a), b), c)$, and $d)$ display four distinct facets of our research at a 1km spatial granularity. Specifically, $a)$ highlights a multivariate cluster featuring training and validation stations, $b)$ presents topographic elevation, $c)$ depicts average rainfall data derived from the IMERG ensemble sample, and $d)$ demonstrates the disaggregated precipitation data.

7.1.2 XGBoost model training and validation results

In each group, the model trains on the top 80% of data from the eight most accurate stations relative to IMERG data, utilizing data from all the groups. High-quality data from these stations is key for training, helping the model predict rain patterns within the same group. The success of this approach is shown in Table 7.2 and Figure 7.3. The performance of each station is influenced by the quality of the training data, as evidenced by metrics such as R^2, $RMSE$, NSE, and KGE. IMERG rain data generally shows less bias at low places and more at high places. Table 7.2 shows that the data from the station at the highest elevation, O1LYSA01, often exhibits higher biases and more pronounced extreme values.
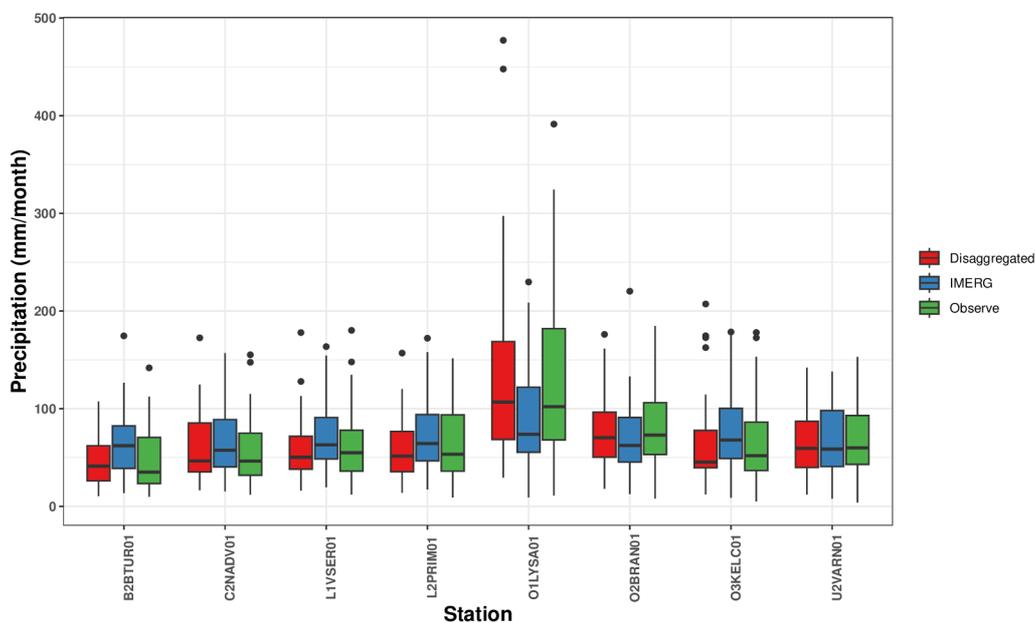
Figure 7.3: Box plot of validation data from training stations with comparison IMERG and disaggregated precipitation

7.1.3 Assessment of IMERG and disaggregated data with respect to ground observed station datasets

Our table (refer to Table 7.3) shows that most of the rainfall data from the station aligns closely with the IMERG data. This work has used specific

53

metrics to validate this, and they indicate a strong similarity in the precipitation patterns between the stations and the IMERG datasets. Key indicators like R^2, RMSE (mm/month) NSE, and KGE for the IMERG data vary in specific ranges across all stations. Data from IMERG at higher altitudes tends to have more bias than at lower altitudes. By looking at Figures 7.4 and 7.5, This thesis work can see how IMERG captures monthly and seasonal patterns. However, IMERG does not accurately represent the varied rainfall patterns over the Czech Republic.

When this work enhances the accuracy of IMERG data through disaggregation, it matches closely with the data from ground stations, as shown in Table 7.3. Disaggregated data also seems more accurate than IMERG at higher altitudes. This enhancement in disaggregated rainfall data means there's less bias in IMERG data. If the IMERG data is reliable, the disaggregated rainfall data is more precise. By increasing its spatial resolution to 1 km, disaggregation better represents rainfall variability.

Using station-specific disaggregation, IMERG data's accuracy improves and can even pick up extreme rainfall patterns, shown in Figure 7.6. Disaggregation maintains the varied rainfall patterns over the Czech Republic, something that IMERG data sometimes misses. This means disaggregation can represent monthly and seasonal rainfall patterns better than IMERG, as shown in Figures 7.4 and 7.5.

Table 7.2: Evaluating the accuracy of the chosen eight stations using the XGBoost model compared to the station's recorded data during the training and validation period.

Cluster	Station	Elevation	Training				Validation			
			R^2	RMSE (mm/month)	NSE	KGE	R^2	RMSE (mm/month)	NSE	KGE
1	B2BTUR01	231.98	0.78	14.21	0.69	0.79	0.69	16.60	0.56	0.75
2	O2BRAN01	774.46	0.81	18.49	0.76	0.84	0.80	18.05	0.79	0.88
3	C2NADV01	589.18	0.93	10.50	0.92	0.91	0.87	12.16	0.86	0.93
4	O3KELC01	308.43	0.88	14.61	0.84	0.86	0.74	22.19	0.72	0.85
5	O1LYSA01	1205.99	0.85	31.36	0.82	0.85	0.85	36.30	0.85	0.91
6	L1VSER01	437.32	0.85	14.45	0.81	0.86	0.76	18.16	0.71	0.84
7	L2PRIM01	613.36	0.88	11.92	0.86	0.88	0.87	13.71	0.82	0.85
8	U2VARN01	374.64	0.90	12.65	0.88	0.88	0.87	16.17	0.76	0.75
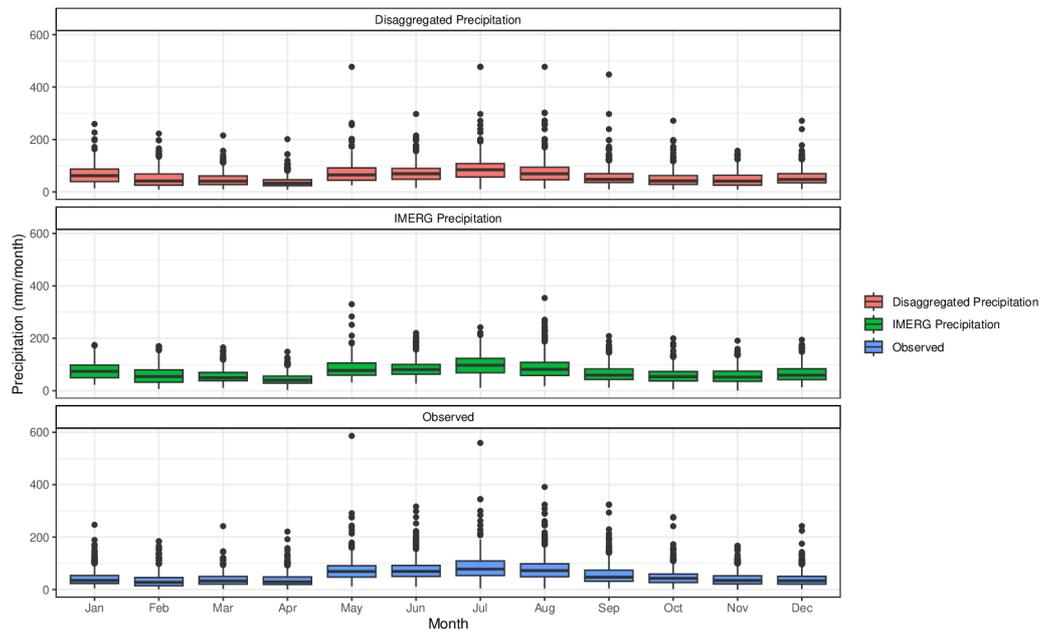
Figure 7.4: Monthly box plots showcasing observed, disaggregated, and IMERG precipitation across all stations.

Table 7.3: Evaluation metrics of the disaggregated and IMERG precipitation compared to data from 27 stations over 255 months.

Cluster	Station	Elevation	Disaggregataed Precipitation				IMERG Precipitation			
			R^2	RMSE (mm/month)	NSE	KGE	R^2	RMSE (mm/month)	NSE	KGE
1	P1KRAL01	239.92	0.59	20.09	0.40	0.70	0.67	30.74	0.16	0.59
1	P1PKAR01	238.70	0.63	19.12	0.44	0.71	0.68	32.80	0.04	0.53
1	P2TUHA01	163.39	0.63	19.07	0.47	0.72	0.73	28.66	0.29	0.62
1	B2BTUR01	231.98	0.76	14.72	0.67	0.78	0.72	23.97	0.42	0.68
1	O3PRER01	218.76	0.68	18.95	0.47	0.66	0.74	25.95	0.41	0.68
1	O1MOSN01	247.39	0.58	28.15	0.09	0.48	0.70	32.61	0.37	0.67
2	O2BRAN01	774.46	0.80	18.40	0.77	0.86	0.79	20.65	0.66	0.77
3	C1KOCE01	510.28	0.79	20.26	0.70	0.73	0.81	23.86	0.54	0.69
3	C2NADV01	589.18	0.92	10.85	0.91	0.92	0.90	14.56	0.83	0.83
3	P3NETV01	572.53	0.59	33.12	0.35	0.57	0.59	38.01	0.14	0.52
3	P3KOSE01	509.07	0.81	16.74	0.80	0.87	0.83	17.86	0.74	0.82
3	P3PRIB01	521.23	0.84	15.11	0.81	0.89	0.83	18.80	0.71	0.80
4	P1PLIB01	290.98	0.72	19.27	0.66	0.82	0.75	25.74	0.41	0.68
4	B2DZBA01	343.69	0.64	20.13	0.51	0.74	0.66	27.16	0.30	0.65
4	B2MBRA01	306.53	0.67	18.82	0.53	0.74	0.70	23.67	0.44	0.70
4	O3KELC01	308.43	0.84	16.41	0.81	0.88	0.79	23.35	0.62	0.76
5	O1LYSA01	1205.99	0.85	32.41	0.82	0.89	0.82	57.89	-0.50	0.12
6	L1VSER01	437.32	0.83	15.27	0.79	0.86	0.80	20.69	0.69	0.81
6	L2MANE01	485.23	0.62	25.30	0.35	0.64	0.61	35.68	-0.01	0.51
6	L2KRAL01	444.00	0.59	26.34	0.27	0.61	0.61	36.02	-0.06	0.49
7	L3AS0001	666.19	0.74	18.56	0.71	0.84	0.76	27.46	0.51	0.69
7	L2PRIM01	613.36	0.88	12.30	0.85	0.88	0.85	19.23	0.73	0.81
7	U1MILE01	663.96	0.65	20.58	0.49	0.74	0.65	30.41	0.24	0.61
7	B2NEDV01	710.53	0.76	17.19	0.61	0.72	0.75	22.62	0.54	0.74
8	L1STAN01	359.85	0.65	34.37	0.15	0.56	0.65	33.51	0.12	0.57
8	P1PRUZ01	371.17	0.66	30.84	0.12	0.60	0.65	31.30	0.13	0.59
8	U2VARN01	374.64	0.89	13.43	0.86	0.86	0.85	15.36	0.81	0.85
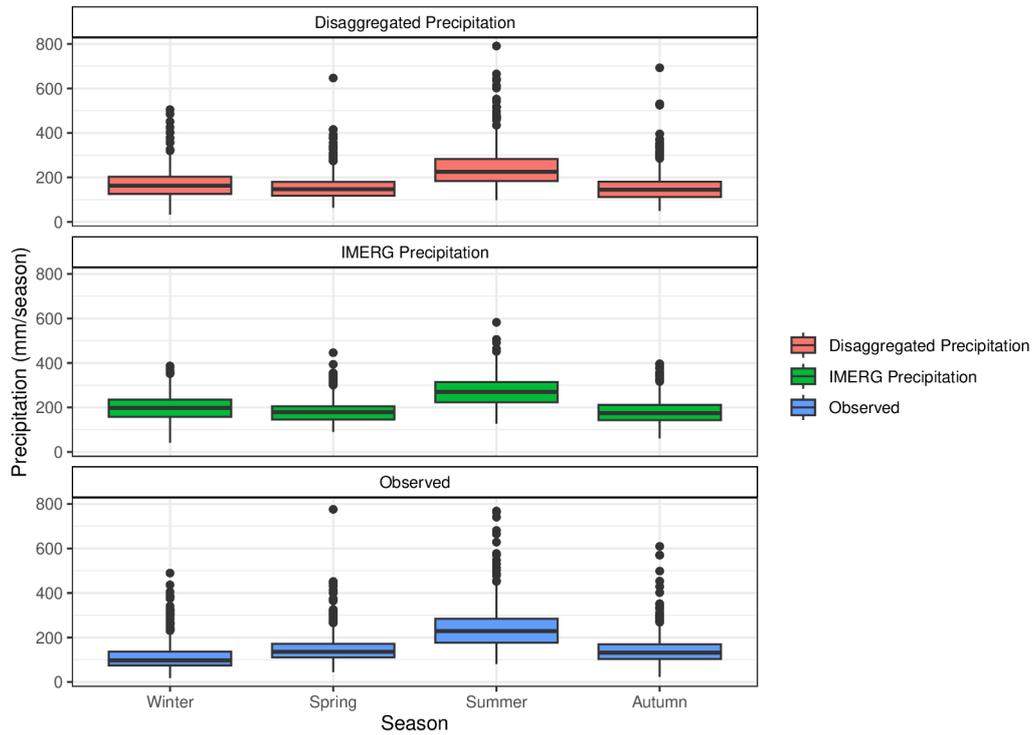
Figure 7.5: Box plots representing observed, disaggregated, and IMERG precipitation on a seasonal scale across all stations.
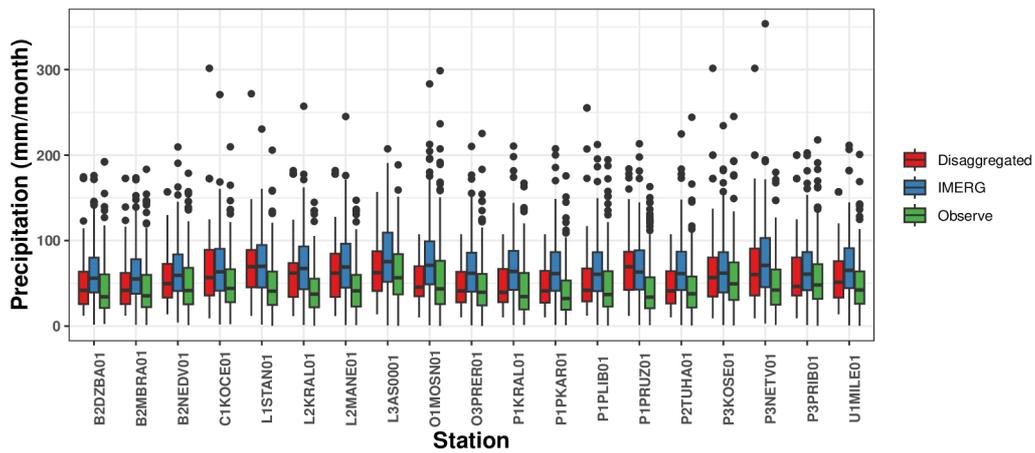
Figure 7.6: Box plot comparing the precipitation from validation stations with IMERG and disaggregated data.

7.2   Runoff reconstruction-Europe continents

In this section, this work assesses the performance of the Budyko models and the proposed HEMMF. This work evaluates HEMMF based on its efficiency and the capability of each input dataset to reconstruct the runoff pattern in Europe over a 500-year period (from 1500 to 1999).

7.2.1   Budyko models output

This section assesses various Budyko models, each with unique strengths and weaknesses, as highlighted by Edoardo Daly et al. (2019)[178]. Every Budyko model delivers distinct spatial-temporal time series patterns. These patterns aid the HEMMF in using the runoff time series as input data. This thesis work displays the accuracy of the Budyko model in two ways, related to different periods. For both methods, this work selects periods based on the ML model's training and validation years, facilitating a comparison of accuracy between hydrological and ML models for these periods. The first representation of accuracy is the spatially aggregated median, depicted in Fig. 7.7. The second displays accuracy by comparing grid cells, as detailed in Table 7.4.

This work has plotted the time series runoff, estimated by four distinct Budyko models in relation to GRUN, in Fig. 7.7 for a side-by-side comparison. This plot suggests that Budyko models can gauge benchmark data patterns annually, albeit with some uncertainties. Fig. 7.7 a) reveals that the Schreiber and Ol'dekop models struggle to replicate the annual GRUN runoff, underperforming in the process. On the flip side, the Budyko and Turk-Pike models fare better in mirroring the annual GRUN runoff. Yet, these two models have their limitations, such as biases in estimating runoff for specific years like 1925, 1940, 1960, and 1980, which the box plot in Fig. 7.7 b), c), and d) also supports. Notably, the validation data for all Budyko models tends to have more high-end values than the GRUN data runoff, as visualized in Fig. 7.7 d). Even with these biases, all four models maintain the GRUN data trend as mentioned in [91]. Both Fig. 7.7 and Table 7.4 indicate that solely relying on the hydrological Budyko models for runoff reconstruction might not yield satisfactory results. Therefore, this work's next move involves enhancing the accuracy of runoff reconstruction using

our proposed HEMMF hypothesis.
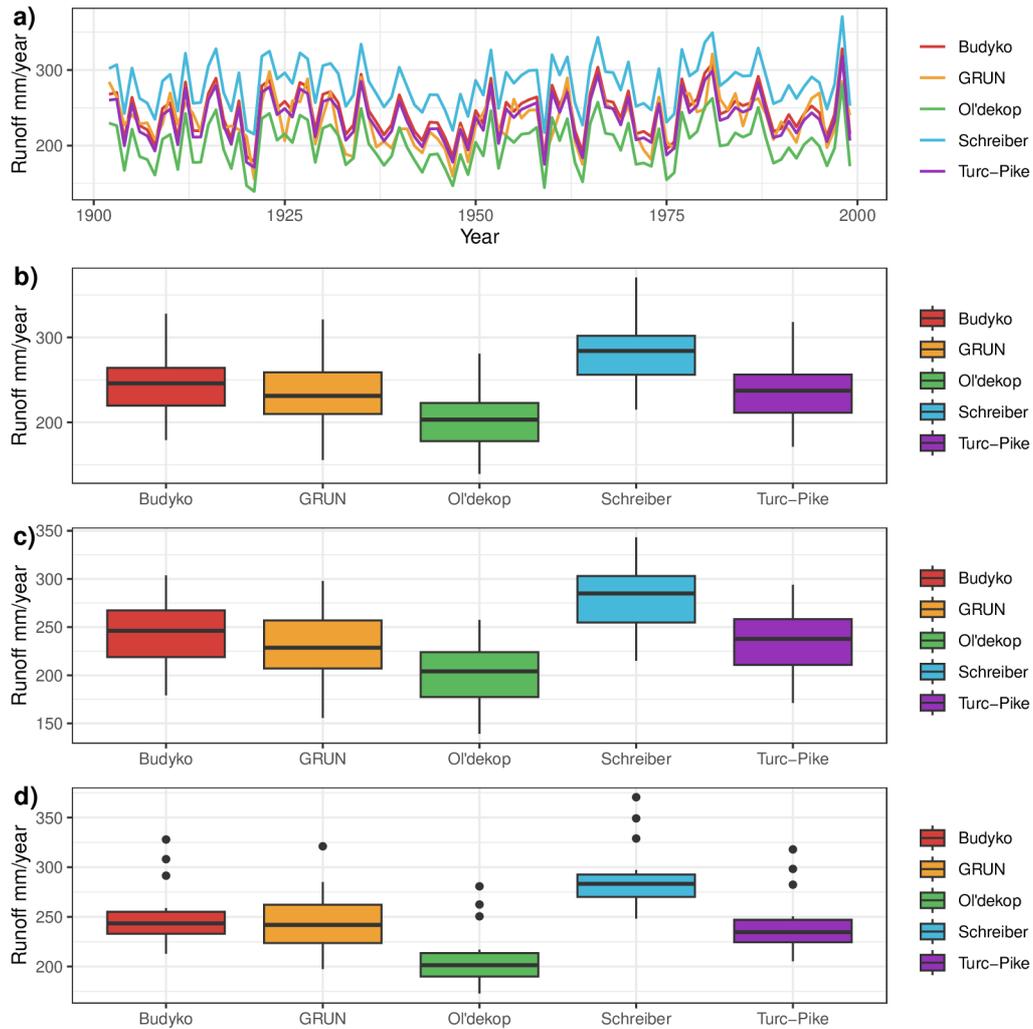
Figure 7.7: Median runoff estimates for Europe, derived from various Budyko hydrological models for distinct timeframes, are contrasted with the reference GRUN datasets. Figures a) and b) depict the runoff spanning 1902 to 1999 in the form of a time series and a box plot, respectively. Meanwhile, Figures c) and d) present box plots of the runoff during the training period (1902-1980) and the validation period (1981-1999), respectively.

Table 7.4: Evaluation metrics of Budyko models for both the training and validation phases, benchmarked against the GRUN data.

	Training Period				Validation Period			
	RMSE (mm/year)	MAE (mm/year)	R^2	KGE	RMSE (mm/year)	MAE (mm/year)	R^2	KGE
Budyko	155.09	96.43	0.77	0.84	155.56	93.17	0.80	0.85
Schreiber	159.05	105.34	0.77	0.82	159.67	102.43	0.80	0.83
Turc-Pike	155.29	95.49	0.77	0.83	155.80	92.23	0.80	0.84
Ol'dekop	160.06	97.24	0.77	0.78	160.43	94.26	0.80	0.79

7.2.2 HEMMF performance

In this section, the thesis displays how ML and ensemble models perform against the benchmark data (GRUN) for HEMMF's first validation step. The second validation step uses the pre-calibrated HEMMF and the aggregated GRDC data. Nine ML algorithms each utilize three input datasets for their training and validation. Every ML model and the combined ensemble produce three output sets corresponding to these input datasets. These datasets demonstrate the consistency of the ML models and the collective ensemble performance. Tables 7.5 and 7.6 present the performance of basic learners and ensemble learning, evaluated on a grid-by-grid basis using RMSE, MAE, and R^2.

7.2.2.1 First-stage validation from 1902-1999

This work is examining the performance of different machine learning models. Using three input datasets, these models' base learner outcomes and ensemble learning results are first validated. Table 7.5 (a) details their performance. QRNN tops the list based on metrics like RMSE, MAE, and R^2. XGBBoost is second, while RPART doesn't perform as well. Other models like CIT, QRF, KNN, GBCSS, MARS, and RLM have varying results, as displayed in the same table. Overall, QRF's training accuracy is the best across all input datasets.

In contrast, Table 7.6 (a) shows the results for ensemble machine learning models under the HEMMF using the first dataset. These ensemble models generally outperform the individual models. Three ensemble models, namely BayesGLM, MLR, and BRRWMA, demonstrate similar performance, while RF is a bit different, as seen in both the metrics and Fig. 7.8.

61

This figure also reveals that the RF's runoff reconstruction deviates slightly from the other three models, especially for the year 1995. All models seem slightly biased in their 1995 runoff estimates compared to the benchmark GRUN dataset.
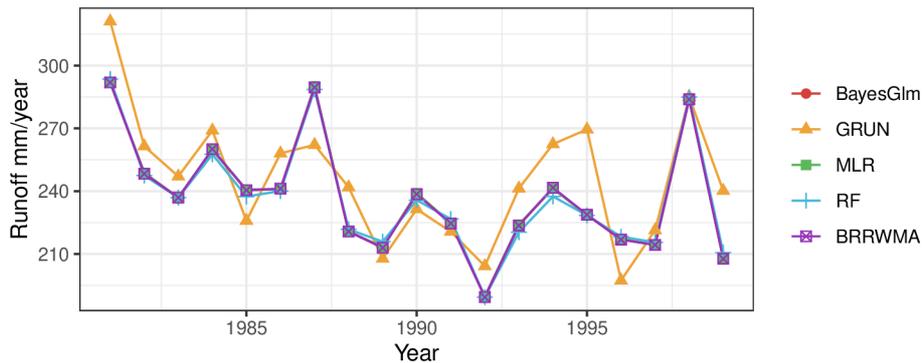
Figure 7.8: Annual European median runoff time series from HEMMF output based on the first input dataset spanning 1981-1999.

Each ML model's performance using the second input dataset is shown in Table 7.5 (b). QRF is the top model, with value of RMSE, MAE, and R^2 at 135.94 mm/year, 92.37 mm/year, and 0.79, respectively. In comparison, GBCSS is the weakest. Other models like KNN, CIT, QRNN, XGBoost, MARS, RLM, and RPART rank from 2nd to 8th in performance.

Table 7.6 (b) displays the ensemble learning performance of HEMMF using the second dataset. BayesGLM, MLR, and BRRWMA have similar performances, while RF lags behind. Fig. 7.9 shows yearly median runoff patterns for Europe and highlights that ensemble models have more bias in this second dataset. Finally, Table 7.6 reveals that neither ML nor ensemble models perform better with the second input than the first.
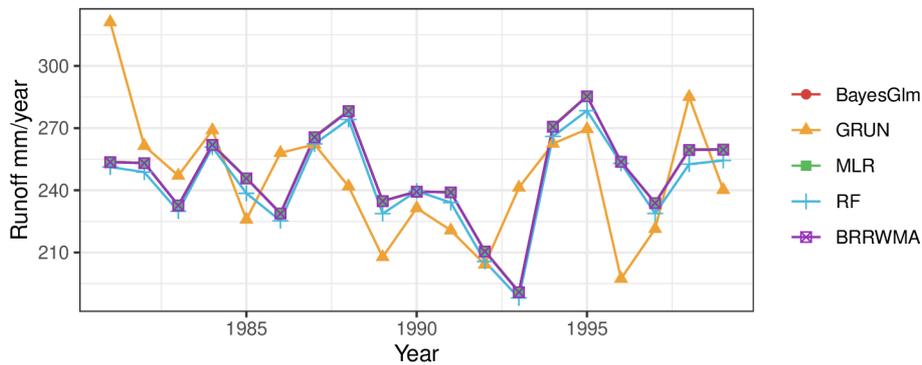
Figure 7.9: Annual European median runoff time series from the HEMMF output, using the second input dataset, spanning 1981-1999.

Table 7.5 (c) displays the cell-by-cell spatial performance of each base learner for the third input dataset. Each ML model shows improved performance using this third dataset compared to the first and second datasets. QRF is the top-performing ML model in validation, with metrics for RMSE, MAE, and R^2 at 127.83 mm/year, 84.47 mm/year, and 0.80 respectively. On the other hand, RPART performs the worst as indicated by Table 7.5 (c). The performance rank order for all ML models remains consistent across all three datasets.

Compared to the base learners of HEMMF, ensemble learning outputs perform consistently better. Table 7.6 (c) shows the performance of each of the four ensemble ML models using the third input dataset. Each ensemble model outperforms individual ML models across datasets. While the performances of BayesGLM, MLR, and BRRWMA are comparable, RF lags behind. Ensemble models using the third dataset always perform better than with the first and second datasets, as seen from Table 7.6 (a), (b) and (c). Fig. 7.10 further visualizes the robustness of these ensemble models.

Fig. 7.11 depicts the yearly median runoff across Europe. The figure, combined with metrics from Table 7.6 (c), reveals that BayesGLM, MLR, and BRRWMA have similar performances, with RF slightly trailing. Overall, ensemble models perform best using the third dataset. Fig. 7.13 displays a 500-year annual runoff reconstruction using the third dataset and BayesGLM. This thesis highlights BayesGLM because of its marginally superior accuracy over BRRWMA and RF, as supported by Table 7.6. The ac-

curacy of this ensemble model using the third dataset is further illustrated in Fig. 7.12.
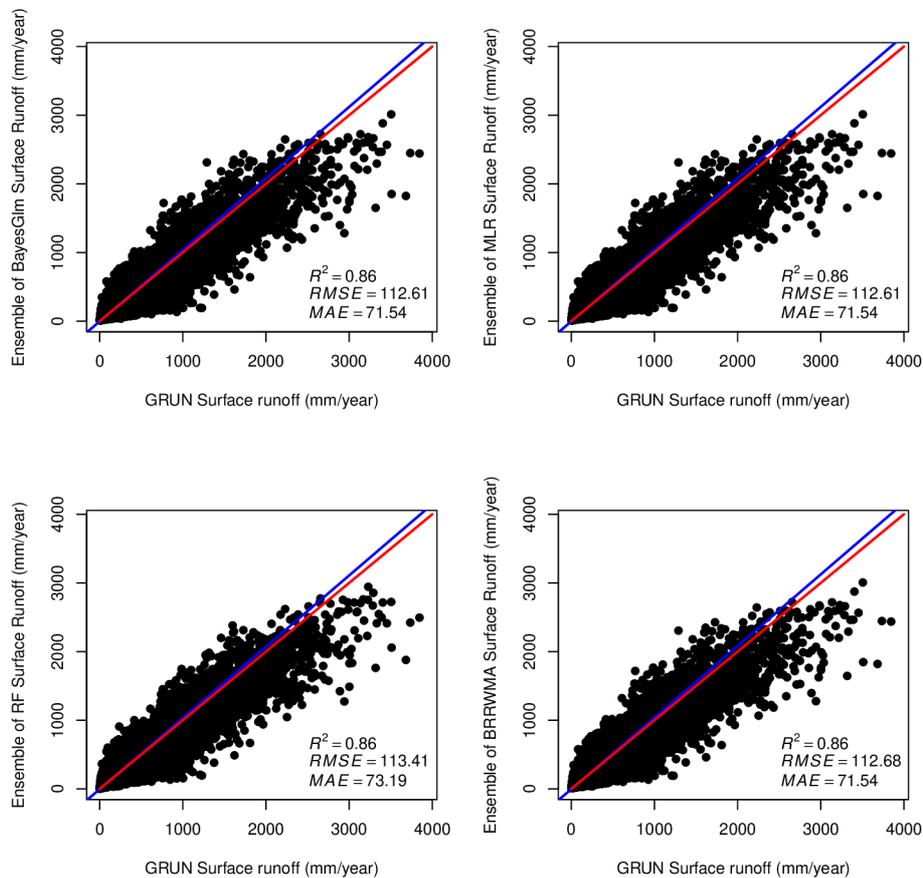
Figure 7.10: Scatter plot illustrating the outputs of four ensemble ML models using the third input dataset, spanning the validation period from 1981 to 1999. The blue color signifies the regression line, while the red colour denotes the 1:1 line.
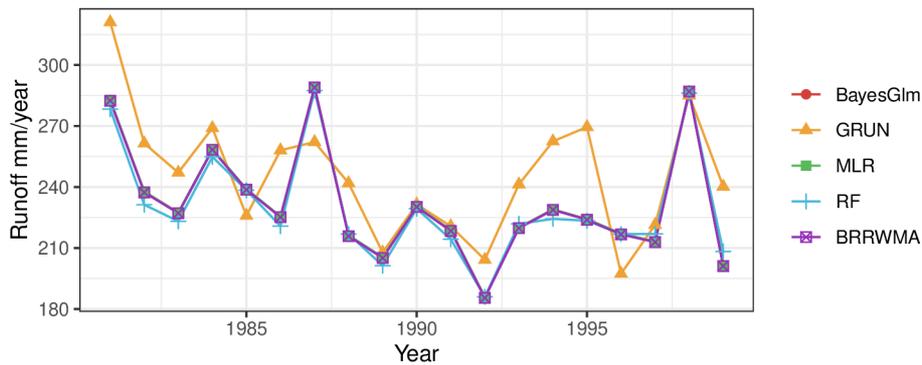
Figure 7.11: The proposed HEMMF model's predictions of the annual European spatial median runoff, using the third input dataset, span the years 1981-1999.

Table 7.5 (c) displays the cell-by-cell spatial performance of each base learner using the third input dataset. Every ML model performs better with this third dataset than with the first two. QRF stands out as the best-performing ML model in validation, presenting metrics for RMSE, MAE, and R^2 at 127.83 mm/year, 84.47 mm/year, and 0.80, respectively. Conversely, RPART is the least effective as seen in Table 7.5 (c). The ranking of all ML models is consistent across the three datasets.

Ensemble learning outshines the base learners of HEMMF in performance. Table 7.6 (c) highlights the achievements of the four ensemble ML models with the third input dataset, each surpassing individual ML models across all input datasets. While BayesGLM, MLR, and BRRWMA show comparable results, RF falls a bit short. Ensemble models using the third dataset consistently outdo their counterparts from the first and second datasets, as evident in Table 7.6 (a), (b), and (c). The robustness of these ensemble models is further portrayed in Fig. 7.10.

Yearly median runoff patterns across Europe are depicted in Fig. 7.11. This visualization, combined with metrics from Table 7.6 (c), indicates closely matched performances for BayesGLM, MLR, and BRRWMA, with RF slightly behind. Generally, the ensemble models excel using the third dataset. Fig. 7.13 presents a 500-year annual runoff reconstruction with the third dataset through BayesGLM, which is spotlighted due to its marginally superior accuracy over BRRWMA and RF, as corroborated by Table 7.6. The precision

of this ensemble model with the third dataset shines in Fig. 7.12.

Because it offers higher accuracy than the other two input datasets, this work uses predictions from the ensemble learning of the HEMMF model made with the third input dataset for spatial-temporal evaluation at the grid cell level. The validation metrics appear in Fig. 7.12. This figure presents three columns of Europe maps: the first displays estimated temporal mean runoff from all ensemble models against the benchmark GRUN dataset; the second presents the NRMSE validation metric; and the third shows the KGE value for the validation period 1981–1999. A model performs better when its NRMSE is lower and its KGE is higher. In northern Europe, the performance validation of each hybrid ensemble model stands out, evidenced by the high KGE shown in Fig. 7.12. However, hybrid ensemble models do not capture runoff exceeding 2000 mm/year in the north of Europe. Their estimated runoff aligns closely with the temporal mean runoff values relative to the benchmark dataset in the northern region. Contrarily, the Mediterranean region displays a higher NRMSE and a lower KGE, as seen in Fig. 7.12.
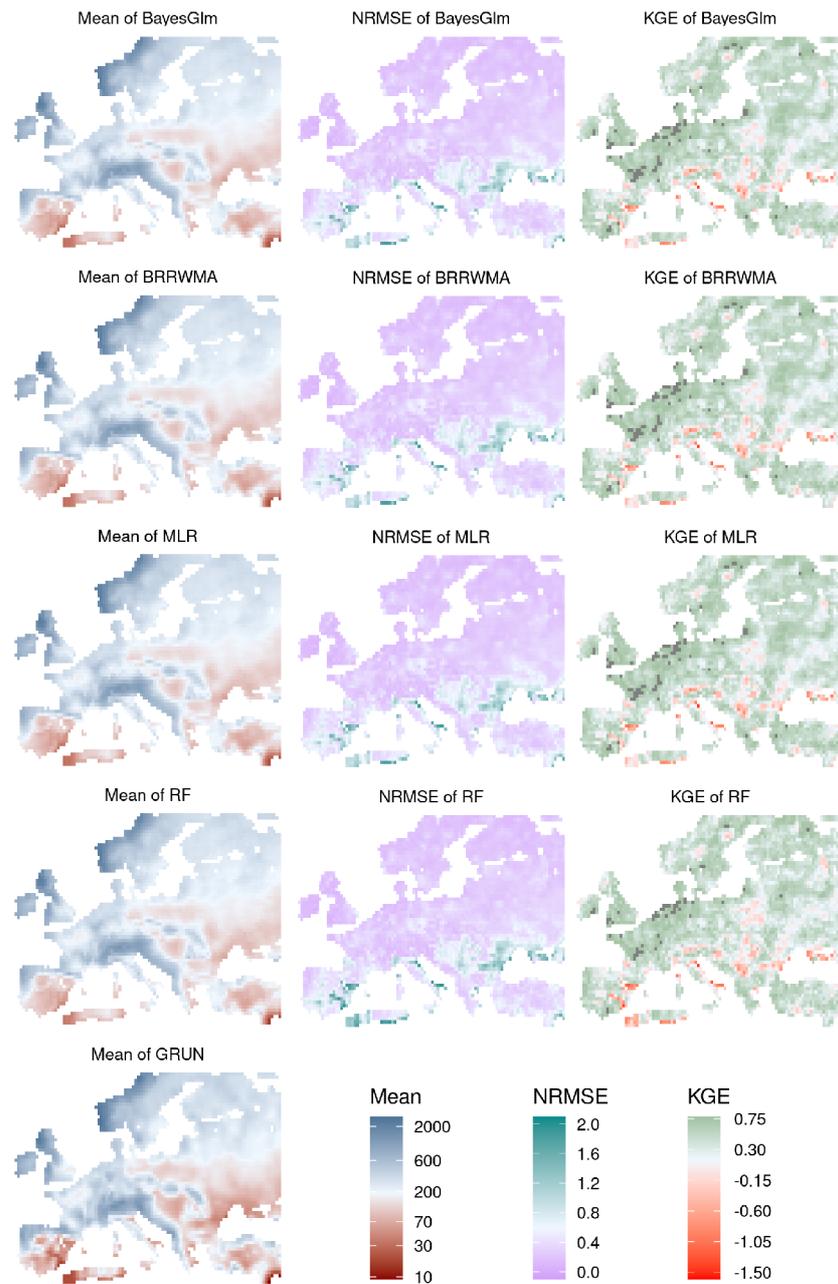
Figure 7.12: Normalized RMSE (NRMSE), mean, and Kling–Gupta effi-
ciency (KGE) of temporal runoff are reconstructed by ensemble ML models
using third input datasets for the validation period 1981–1999.
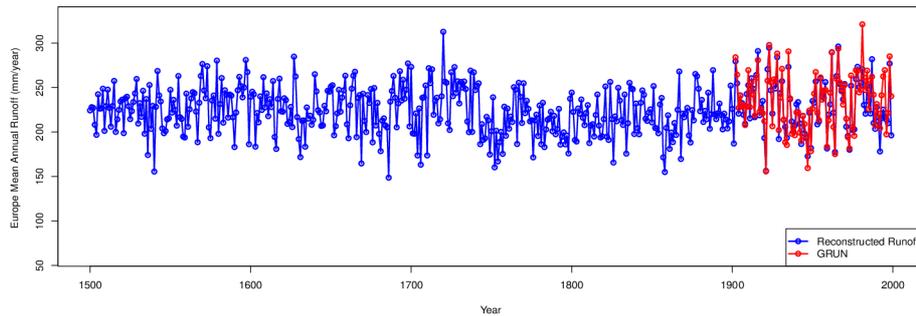
Figure 7.13: Five hundred years reconstructed annual spatial median runoff of Europe with respect to GRUN using BayesGLM.

7.2.2.2 Second-stage validation from 1800 to 1901

During the first validation stage, this work focuses on spatially distributed runoff patterns, while the second validation stage examines the overall performance of the current HEMMF model, representing the aggregated runoff pattern over the river basin. This research validates the reconstructed runoff of 16 chosen river basins using the GRDC dataset from 1800–1901. This thesis includes the results of this validation stage in the supplementary material. This thesis work also adds the reconstructed runoff using the BayesGLM model to the supplementary material, as its accuracy is a bit higher than the two ensemble models but demonstrates equivalent performance within the same ML category. Fig. **??** in the supplementary material presents a scatter plot of the reconstructed versus GRDC runoff for each GRDC station. Fig. **??** displays hydrographs of the averaged reconstructed and GRDC runoff time series over basin areas. Fig. **??** reveals that the GRDC hydrograph for station codes NO 6935051 and 6935052 are very similar, attributed to their similar catchment areas and their location within the same grid cell values. Fig. **??** illustrates the median European hydrograph for 500 years concerning reconstruction, training, and validation. Table 7.7 indicates that many GRDC stations align with the reconstructed runoff dataset, but some stations show a bias in the reconstructed runoff.

Table 7.5: Evaluation of various ML models' historical runoff predictions, serving as the foundational learners in the HEMMF, against the benchmark GRUN dataset. These forecasts rely on three distinct data sources integrated into the HEMMF framework.

(a) Evaluation metrics for various ML models applied to historical runoff prediction using the first input dataset.

	Training			Validation		
	RMSE(mm/year)	MAE(mm/year)	R^2	RMSE(mm/year)	MAE(mm/year)	R^2
GBCSS	134.03	85.33	0.77	135.38	84.04	0.80
RPART	176.36	111.29	0.61	188.17	113.41	0.60
QRNN	121.81	75.43	0.82	122.61	70.80	0.84
CIT	115.71	73.21	0.83	127.32	79.43	0.82
XGBoost	119.81	77.44	0.82	126.41	78.48	0.82
MARS	132.94	84.58	0.78	134.75	84.35	0.80
RLM	137.94	86.97	0.76	139.09	85.24	0.79
KNN	92.00	58.04	0.89	135.71	84.68	0.79
QRF	34.81	9.96	0.99	128.42	77.95	0.81
Mean	128.83	81.54	0.78	138.68	85.05	0.78
SD	38.57	27.77	0.09	19.74	11.88	0.07

(b) Evaluation metrics for various ML models when predicting historical runoff using the second input dataset.

	Training			Validation		
	RMSE(mm/year)	MAE(mm/year)	R^2	RMSE(mm/year)	MAE(mm/year)	R^2
GBCSS	209.61	148.74	0.45	218.64	156.54	0.47
RPART	202.39	139.98	0.49	209.74	146.37	0.51
QRNN	165.39	107.22	0.66	164.26	105.49	0.70
CIT	162.18	106.59	0.68	162.08	106.00	0.70
XGBoost	166.32	110.76	0.66	166.52	109.86	0.69
MARS	188.22	130.35	0.56	187.51	130.52	0.61
RLM	199.36	130.23	0.51	201.17	132.50	0.55
KNN	154.98	102.76	0.70	154.04	102.27	0.73
QRF	140.38	93.54	0.76	135.94	92.37	0.79
Mean	154.09	118.91	0.61	177.77	126.44	0.64
SD	61.70	18.88	0.11	27.85	25.20	0.11

(c) Evaluation metrics for various ML models when predicting historical runoff using the third dataset.

	Training			Validation		
	RMSE(mm/year)	MAE(mm/year)	R^2	RMSE(mm/year)	MAE(mm/year)	R^2
GBCSS	132.74	84.65	0.78	133.54	82.87	0.80
RPART	176.40	125.62	0.61	185.71	132.18	0.62
QRNN	116.73	73.05	0.83	119.73	73.38	0.85
CIT	109.89	69.73	0.85	122.79	78.08	0.83
XGBoost	113.43	74.40	0.84	74.24	74.40	0.85
MARS	125.41	80.91	0.80	127.80	79.24	0.82
RLM	134.80	84.47	0.74	135.69	82.85	0.80
KNN	84.70	52.75	0.91	134.45	84.17	0.80
QRF	27.06	6.45	0.99	116.56	73.04	0.85
Mean	113.46	67.23	0.82	127.83	84.47	0.80
SD	40.69	38.35	0.11	28.68	18.39	0.07

Table 7.6: Evaluation of the historical runoff's predictive accuracy across various machine learning models that constitute the ensemble in the HEMMF, benchmarked against the GRUN dataset. The predictions rely on three distinct input datasets fed into the HEMMF framework.

(a) Evaluation measures for various ML model ensembles predicting historical runoff using the first data set.

	Training				Validation			
	RMSE(mm/year)	MAE(mm/year)	R^2	KGE	RMSE(mm/year)	MAE(mm/year)	R^2	KGE
BayesGLM	89.46	55.75	0.90	0.90	120.59	75.24	0.84	0.86
RF	98.61	60.92	0.89	0.89	121.17	76.51	0.84	0.86
MLR	89.46	55.75	0.90	0.90	120.59	75.24	0.84	0.86
BRRWMA	89.44	55.70	0.90	0.90	120.62	75.22	0.84	0.86
Mean	91.74	57.03	0.90	0.90	120.74	75.55	0.84	0.86
SD	4.58	2.59	0.00	0.00	0.29	0.64	0.00	0.00

(b) Evaluation measures for various ML model ensembles predicting historical runoff based on the secondary input dataset.

	Training				Validation			
	RMSE(mm/year)	MAE(mm/year)	R^2	KGE	RMSE(mm/year)	MAE(mm/year)	R^2	KGE
BayesGLM	141.09	95.59	0.74	0.84	137.01	91.86	0.79	0.82
RF	146.01	97.26	0.74	0.83	143.87	96.24	0.77	0.81
MLR	141.09	94.59	0.75	0.84	137.01	91.88	0.79	0.82
BRRWMA	141.02	94.52	0.75	0.84	136.97	91.81	0.79	0.82
Mean	142.30	95.49	0.74	0.84	138.72	92.95	0.78	0.82
SD	2.47	1.28	0.01	0.00	3.44	2.20	0.01	0.00

(c) Evaluation metrics for various ML model ensembles in predicting historical runoff with the third input dataset.

	Training				Validation			
	RMSE(mm/year)	MAE(mm/year)	R^2	KGE	RMSE(mm/year)	MAE(mm/year)	R^2	KGE
BayesGLM	50.31	29.82	0.97	0.95	112.61	71.54	0.86	0.85
RF	64.44	32.20	0.95	0.93	113.41	73.19	0.86	0.86
MLR	50.31	29.82	0.97	0.95	112.61	71.54	0.86	0.85
BRRWMA	50.49	29.88	0.97	0.95	112.68	71.54	0.86	0.85
Mean	53.89	30.43	0.96	0.95	112.83	71.95	0.86	0.85
SD	7.04	1.18	0.01	0.01	0.39	0.82	0.00	0.00

Table 7.7: Evaluation Metrics for the BayesGLM Model Output Using Third Input Data from 1800 to 1900 in Comparison to GRDC River Basin Data.

GRDC NO	Station	River Basin	Catchment Area (km^2)	R^2	RMSE(mm/year)	MAE(mm/year)	KGE
6123300	Blois	Loire	40630.00	0.60	0.61	47.00	0.74
6335020	Rees	Rhine	159424.00	0.64	50.00	36.00	0.66
6335060	Kolen	Rhine	144232.00	0.71	42.00	32.00	0.74
6335301	Schweinfurt	Main	12715.00	0.47	71.00	56.00	0.39
6337100	Vlotho	Weser	17618.00	0.55	78.00	65.00	0.55
6335500	Wuerzburg	Main	14031.00	0.45	49.00	37.00	0.55
6337200	Intschede	Weser	34424.00	0.52	68.00	54.00	0.49
6337400	Hann.-Muenden	Weser	12442.00	0.61	44.00	36.00	0.67
6337514	Bodenwerder	Weser	15924.00	0.60	55.00	45.00	0.64
6340120	Dresden	Labe (Elbe)	53096.00	0.41	65.00	55.00	0.56
6343100	Wasserburg	Inn	11983.00	0.14	120.00	95.00	0.34
6343500	Burghausen	Salzach	6649.00	0.17	346.00	301.00	0.22
6742200	Orsova	Danube	576232.00	0.45	101.00	92.00	0.55
6935051	Basel	Rhine	35897.00	0.51	102.00	77.00	0.60
6935052	Basel, Schifflaende	Rhine	35905.00	0.51	102.00	77.00	0.60
6974150	Smalininkai	Nemunas - Neman	81200.00	0.11	43.00	34.00	0.31

Part V

Discussion

DISCUSSION

8.1   Context of precipitation disaggregation

High-resolution products for hydrological parameters are essential to understanding the patterns, causes, linkages, and impacts of weather and climate nationally. The present semi-modelled data products on hydrological parameters from various satellite constellations offer a fundamental data source, defining big data from every angle, albeit at a global level with coarse resolution and rapid temporal frequency. Integrated datasets, like the IMERG precipitation data, serve as critical ultra-big data sources for global weather and climate research. Yet, for national-level application, which aims to capture local patterns, there's a need to remodel these datasets for finer spatial and temporal resolutions. Most current disaggregation methods rely on expert-driven indices, multi-variate regression, or spatial interpolation, which confines the scope of datasets to specific time

steps or regions.

In contrast, this study introduces a non-parametric disaggregation method using the Czech Republic as an example. This method leverages ML and fuzzy clustering combined with elevation data to produce finer-resolution precipitation data from IMERG. Figures 7.4 and 7.5 show that IMERG data closely matches ground-observed data across stations, largely due to its calibration with rain gauge data [179]. However, discrepancies arise at higher altitudes, as seen at the O1LYSA01 station, due to significant precipitation variability within short distances [180]. These disparities stem from satellite data bias, given the considerable heterogeneity within each 11 x 11-kilometer pixel area [181].

Recognizing the IMERG data's bias concerning elevation, this study introduces a hybrid method, pairing multivariate clustering with the XGBoost model. This aims to disaggregate high-resolution precipitation and reduce elevation-induced biases. This work validates this method by comparing the high-resolution disaggregated precipitation and resampled IMERG precipitation to the ground-observed station and cluster data, ensuring they share similar precipitation patterns.

Although IMERG data often masks spatial variability due to its lower resolution [182], precipitation variability depends largely on elevation [183, 63]. Hence, this study employs multivariate clustering of elevation and precipitation, ensuring the disaggregated precipitation reflects higher spatial-temporal variability based on multivariate cluster outputs combined with the XGBoost model.

Using multivariate clustering, data groups form based on similar patterns and local climates. Pairing this with XGBoost increases accuracy by focusing on similar data patterns. This ensures the XGBoost-trained model, informed by a single station's observation, effectively reduces bias within each data group. As the XGBoost algorithm captures these patterns, it minimizes disaggregation bias. The station with the highest accuracy compared to IMERG data within each cluster is chosen for the best results, generating a robust prediction model.

The accuracy of this model heavily depends on the reliability of IMERG data. The data must be trustworthy even when the most accurate station within a cluster is selected. Therefore, this study relies on bias-corrected

IMERG monthly data calibrated with rain gauge information. However, this work's disaggregation further enhances the dataset's accuracy and resolution. This approach offers vast potential for global applications on IMERG data or other climate variables, creating high-resolution precipitation data in areas lacking reliable ground-observed stations.

8.2   Context of historical runoff reconstruction

Climate-reconstructed data contains uncertainty, and as a result, hydrological models, ML, and the ensemble of ML cannot reconstruct less biased runoff from biased input data. This work promotes a method that addresses this challenge by harnessing the capabilities of hydrological models, ML, and the ensemble of ML, collectively termed as HEMMF. This study aims to demonstrate the potential of HEMMF in reconstructing long-term runoff patterns using various extensive input datasets and two kinds of ML, accompanied by a two-stage validation. The comprehensive input data patterns also encompass the SPA of the spatial cell values of the extended inputs. To comprehend the potential contribution of SPA and Budyko models' output, the thesis work compares the effects of three distinct input datasets on two different types of ML models during the HEMMF implementation.

Fig. 7.7 reveals that Budyko models exhibit bias when reconstructing accurate runoff from climate-reconstructed data due to uncertainties in the reconstruction of Pauling precipitation [184]. This primary input for the Budyko models is susceptible to these uncertainties. Such uncertainties in climate-reconstructed data also carry forward into the reconstruction of hydrological processes [185]. This work proposed that HEMMF notably reduces the uncertainty of hydrological balance models in achieving accurate runoff estimations using climate-reconstructed data. This thesis work validates this with this work results benchmarked against specific datasets (Figs. 7.10, 7.11). Hence, this thesis work supports and expands upon the insights from Konpala et al. [52], which advocate for integrating Budyko models with ML models to diminish bias and enhance adaptability in runoff predictions. This work approach infuses more flexibility by integrating SPA and an ensemble of ML models into HEMMF.

77

This work implements the proposed HEMMF using three separate datasets, evaluating the efficacy of each component [186], such as spatio-temporal correlation [187], the intricate hydrological patterns generated by the physical model [188], and ML combined with an ensemble of ML algorithms [128]. Subsequent investigations into HEMMF might look into utilizing diverse climate-reconstructed datasets or other products, such as hydrometeorological patterns produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), General Circulation Models (GCMs), and datasets crafted within the framework of Coupled Model Intercomparison Project Phase 5 (CMIP5) and Phase 6 (CMIP6) [53, 189]. While there are studies that analyze the aforementioned datasets [190, 191, 192] and report precise runoff estimations, they typically work on datasets with short temporal and spatial scopes and minimal intrinsic bias. Conversely, this work framework consistently excels on gridded, extensive historical continental scale datasets, especially considering the pronounced uncertainty inherent in precipitation data [184]. This framework holds promise in performing sturdily across various climate-reconstructed datasets that reconstruct diverse environmental spatial fields, such as runoff and evapotranspiration.

The outcomes of this study indicate that the input gridded data, which includes their information on SPA, yields higher accuracy in runoff pattern recognition compared to inputs solely composed of gridded data or the information from SPA gridded datasets. The performances of the ML and ensemble of ML models are the poorest when the inputs come only from the SPA of gridded data, based on the correlation of neighboring spatial fields. Yet, only varying the neighbourhood's correlation in gridded cells doesn't sufficiently capture the variation in the runoff pattern. Thus, ML and ensemble models don't represent the runoff patterns accurately using only the input from the neighbourhood correlation of precipitation, temperature, PDSI, and runoff data estimated by the four Budyko models.

The GRUN benchmark datasets undergo reconstruction using RF, so this study avoids using RF in the base learners of the thesis work proposed HEMMF, choosing QRF instead. However, RF is part of the ensemble learning of HEMMF. The QRF model, an extension of the RF family, displays the best overall performance (both in training and validation) among all nine ML models across all three input datasets, mainly because the GRUN benchmark dataset also comes from RF. Each base learner model's perfor-

mance hinges on the reliability or noise of the input datasets, the model architecture, and the strengths and weaknesses of the algorithm. The performance of all nine ML models improves with the third input dataset, which includes information on the SPA estimated for all gridded data. Furthermore, the accuracy of these nine ML models increases when used within an ensemble ML framework for runoff reconstruction.

The three ensemble models, MLR, BayesGLM, and BRRWMA, display similar performance in runoff reconstruction across all datasets, reflecting the diversity and complementarity of each ensemble's ML algorithms. The RF ensemble adopts a distinct algorithmic approach compared to the other three ensembles, resulting in slightly different performance. Generally, the outputs of all four ensemble models are expected to be similar, as ensemble models tend to offer comparable performance on large input data. This study demonstrates that the third input dataset holds the most promise for reconstructing runoff using the HEMMF. Some recent studies have shown that the BayesGLM model reliably forecasts both flash floods and historical tropical species in relation to climate.

Fig. 7.12 shows the comparative performance of the ensemble models using NRMSE and KGE metrics. Higher NRMSE and lower KGE values in certain Mediterranean regions indicate that the ensemble model does not predict the runoff accurately in these areas. This might stem from the distinct hydrological climate systems present in these regions and uncertainties in the predictor variables, such as precipitation [45] and temperature [46]. Since 1971, the Mediterranean region sees a relatively high percentage change in annual runoff [91]. The ensemble model struggles to account for this deviation in the runoff due to uncertainty in the predictor variables. Furthermore, the Mediterranean region displays significant runoff anomalies since 1950 [37, 193]. This anomaly suggests that the training data patterns differ from the validation data patterns. The predictor variables might not capture this difference, making the model's predictions less precise during the validation period.

There is an increasing tendency of runoff rate in northern Europe [175], and uncertainty exists in predictor variables. Because of this, the ensemble model does not capture runoff above 2500 mm/year; the values higher than 2500 mm/year appear in Fig. 7.10 and 7.12. Each ensemble ML model currently underestimates the runoff compared to benchmark data, especially

for extreme values (above 2500 and below 100 mm/year), as depicted in Fig. 7.10. This uncertainty indicates that the noise in predictor variables significantly affects the performance of hybrid frameworks, including the HEMMF framework this work proposes.

The main limitation of the hybrid framework this work presents is that its performance mainly depends on the individual ML models that form the ensemble, and it does not allow for optimization of the hyperparameters. This thesis assumes that non-optimal tuning of the ML hyperparameters on large input datasets either marginally affects performance or does not impact model performance at all [121]. Optimizing the hyperparameters of the HEMMF framework with extensive datasets does not justify the computational resources [121]. As this work mentions, the default selection of hyperparameters used in this study results in satisfactory performance of ML models with reduced computational costs [110, 194]. Besides this limitation, the framework's performance is also influenced by noise in the input data [121]

8.3 Future Perspectives

Hybrid modelling represents a compelling frontier for gridded climate data processing, particularly when addressing challenges related to gridded climate data disaggregation and reconstruction in both spatial and temporal gaps. Combining the strengths of multiple modelling approaches, hybrid models can leverage the best features of each, offering enhanced accuracy and flexibility in predictions. Especially in the context of climate change, where uncertainties in data are commonplace, hybrid frameworks like the HEMMF demonstrate the potential to mitigate errors arising from individual modelling techniques. By synergizing traditional hydrological models, clustering, machine learning, and ensemble methods, these combined approaches can achieve superior results in reconstructing past climate patterns, filling data voids, and enhancing the resolution of coarse datasets.

The evolution of hybrid modelling can pave the way for more robust climate data analyses, fostering better-informed decision-making in environmental planning, policy development, and disaster management. As this work witnesses a growth in computational capabilities and as method-

ologies become increasingly sophisticated, there is a growing expectation that hybrid models will be adept at managing more extensive datasets, refining hyperparameters with greater efficiency, and incorporating the latest machine-learning techniques to yield enhanced results. With the ongoing challenges posed by climate change, the role of advanced hybrid modelling in providing more accurate and comprehensive climate data representations will be indispensable, potentially revolutionizing how this work understands and addresses the changing environment and climate.

Part VI

Conclusion

SUMMARY AND CONCLUSION

9.1   Overall remarks on precipitation disaggregation and historical runoff reconstruction

This thesis has unveiled an advanced hybrid learning strategy that synergizes multivariate clustering and the XGBoost model to spatially disaggregate precipitation. This work bridged the gap between the IMERG data and ground-measured station inputs by capitalising on sensitive parameters like elevation and precipitation. This process culminates in a disaggregated precipitation dataset rigorously authenticated against ground observations. This work shows how leveraging SRTM elevation data aids in proficiently disaggregating monthly precipitation over the diverse Czech Republic landscape, producing a high-resolution precipitation dataset. The intricacies of multivariate clustering not only deliver the much-needed one-kilometer resolution and maintain consistency within each segmented cluster. This work's findings demonstrate a remarkable concordance be-

tween the disaggregated and actual observed precipitation, thus boosting the precision and practicality of monthly-scale IMERG data.

The preliminary coarse-scale IMERG precipitation data portrayed a growing bias at soaring elevations. Yet, this thesis's groundbreaking approach to precipitation disaggregation has risen to this challenge, enhancing its precision when juxtaposed with ground data. Although there are avenues to further perfect the accuracy, especially in high-altitude zones, this work methodology has augmented the spatial granularity and precision of precipitation data across all terrains, emphasizing elevated terrains. Given the transformative climate scenarios, the high-resolution precipitation dataset this work has formulated is a beacon of hope for myriad hydro-meteorological ventures.

Simultaneously, this thesis's paper delves deep into the merits of a groundbreaking hybrid ML framework, the Hybrid Ensemble Multi-Model Framework (HEMMF), purpose-built to revitalize historical runoff. This incorporates reconstructed climatic data, Budyko model outputs, and the nuance of spatial autocorrelation. The outcome of this work includes a historical runoff reconstruction characterized by minimal bias. These work endeavours underscore the merit of harmonizing hydrological and ML paradigms within an integrated framework like HEMMF, paving the way for the intricate mapping of runoff patterns. The HEMMF's prowess is steered by an ensemble of factors, including the nature of input data and ML model selection. The framework's calibre is palpable in its robust runoff reconstructions, especially when juxtaposed against substantial datasets. The reconstructed runoff coheres well with most GRDC runoff stations, although deviations exist due to the coarse resolution and spatial mismatches. The invaluable insights from the reconstructed runoff stand are poised to steer water resource management in Europe amidst the caprices of climate change.

Building upon this, the HEMMF boasts the versatility to interface with various climate datasets. Its efficacy hinges on the trustworthiness of input data and the ensemble of model performances. Despite some intrinsic limitations, this hybrid framework emerges as a beacon for spatio-temporal data reconstruction, primarily driven by the amalgamation of multi-ML, Budyko, and SPA models. Future trajectories will probe into the inclusion of eclectic components and advanced algorithms, refining the selection of single-based learners. These work methods signal a paradigm shift in data

disaggregation and historical runoff reconstruction, poised to guide policy-making and climate modelling endeavours.

Part VII

References

Bibliography

[1] Tim P Barnett, Jennifer C Adam, and Dennis P Lettenmaier. "Potential impacts of a warming climate on water availability in snow-dominated regions". In: *Nature* 438.7066 (2005), pp. 303–309.

[2] Samuel D Brody et al. "The rising costs of floods: Examining the impact of planning and development decisions on property damage in Florida". In: *J Am Plann Assoc.* 73.3 (2007), pp. 330–345.

[3] Zilefac Elvis Asong et al. "High-resolution meteorological forcing data for hydrological modelling and climate change impact analysis in the Mackenzie River Basin". In: *Syst. Sci. Data* 12.1 (2020), pp. 629–645.

[4] Jagadeesh Anmala, Bin Zhang, and Rao S Govindaraju. "Comparison of ANNs and empirical approaches for predicting watershed runoff". In: *Water Resour. Plan. Manag.* 126.3 (2000), pp. 156–166.

[5] HAP Hapuarachchi, QJ Wang, and TC Pagano. "A review of advances in flash flood forecasting". In: *Hydrol. Process.* 25.18 (2011), pp. 2771–2784.

[6] Timothy Dube et al. "Advancements in Earth Observation for Water Resources Monitoring and Management in Africa: A Comprehensive Review". In: *J. Hydrol.* (2023), p. 129738.

[7] Jonas Olsson et al. "Downscaling of short-term precipitation from regional climate models for sustainable urban planning". In: *Sustainability* 4.5 (2012), pp. 866–887.

[8] Janusz Sobieraj, Marek Bryx, and Dominik Metelski. "Stormwater Management in the City of Warsaw: A Review and Evaluation of Technical Solutions and Strategies to Improve the Capacity of the Combined Sewer System". In: *Water* 14.13 (2022), p. 2109.

[9] NL Poff, Mark M Brinson, and JW Day. "Aquatic ecosystems and global climate change". In: *Pew Center on Global Climate Change, Arlington, VA* 44 (2002), pp. 1–36.

[10] Phil L Kenkel and Patricia E Norris. "Agricultural producers' willingness to pay for real-time mesoscale weather information". In: *J. Agric. Resour. Econ.* (1995), pp. 356–372.

[11] Carolien Toté et al. "Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique". In: *Remote Sensing* 7.2 (2015), pp. 1758–1776.

[12] Christopher J Sergeant et al. "Risks of mining to salmonid-bearing watersheds". In: *Sci. Adv.* 8.26 (2022), eabn0929.

[13] Adam Luke et al. "Going beyond the flood insurance rate map: insights from flood hazard map co-production". In: *Nat. Hazards Earth Syst. Sci.* 18.4 (2018), pp. 1097–1120.

[14] Flannery Dolan et al. "Evaluating the economic impact of water scarcity in a changing world". In: *Nat. Commun.* 12.1 (2021), p. 1915.

[15] Shuai Xiao, Jun Xia, and Lei Zou. "Evaluation of multi-satellite precipitation products and their ability in capturing the characteristics of extreme climate events over the Yangtze River Basin, China". In: *Water* 12.4 (2020), p. 1179.

[16] Rajani K Pradhan et al. "Review of GPM IMERG performance: A global perspective". In: *Remote Sens Environ.* 268 (2022), p. 112754.

[17] Ali Karbalaye Ghorbanpour et al. "Comparison and assessment of spatial downscaling methods for enhancing the accuracy of satellite-based precipitation over Lake Urmia Basin". In: *J. Hydrol.* 596 (2021), p. 126055.

[18] Shaodan Chen et al. "Evaluation of Tropical Rainfall Measuring Mission (TRMM) satellite precipitation products for drought monitoring over the middle and lower reaches of the Yangtze River Basin, China". In: *J. Geogr. Sci.* 30.1 (2020), pp. 53–67.

[19] GM Foody. "Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship". In: *Remote Sens. Environ.* 88.3 (2003), pp. 283–293.

[20]   Shaodan Chen et al. "Spatial downscaling of Tropical Rainfall Mea-
       suring Mission (TRMM) annual and monthly precipitation data over
       the Middle and Lower Reaches of the Yangtze River Basin, China".
       In: *Water* 11.3 (2019), p. 568.

[21]   Shiguang Xu et al. "A new satellite-based monthly precipitation
       downscaling algorithm with non-stationary relationship between
       precipitation and land surface characteristics". In: *Remote Sens. En-
       viron.* 162 (2015), pp. 119–140.

[22]   Tao Zhang et al. "Spatial downscaling of TRMM precipitation data
       considering the impacts of macro-geographical factors and local
       elevation in the Three-River Headwaters Region". In: *Remote Sens
       Environ.* 215 (2018), pp. 109–127.

[23]   Celso Augusto Guimarães Santos et al. "Cluster analysis applied
       to spatiotemporal variability of monthly precipitation over Paraíba
       state using Tropical Rainfall Measuring Mission (TRMM) data". In:
       Remote Sens. 11.6 (2019), p. 637.

[24]   Baohong Ding et al. "The dependence of precipitation types on sur-
       face elevation and meteorological conditions and its parameteriza-
       tion". In: *J. Hydrol.* 513 (2014), pp. 154–163.

[25]   Hang Yu, Long Wang, and Maoling Yang. "Effects of Elevation and
       Longitude on Precipitation and Drought on the Yunnan–Guizhou
       Plateau, China". In: *Pure Appl. Geophys.* (2023), pp. 1–21.

[26]   Satya Prakash, RM Gairola, and AK Mitra. "Comparison of large-
       scale global land precipitation from multisatellite and reanalysis
       products with gauge-based GPCC data sets". In: *Theor. Appl. Cli-
       matol.* 121 (2015), pp. 303–317.

[27]   David M Straus. "Clustering techniques in climate analysis". In: *Ox-
       ford Research Encyclopedia of Climate Science.* 2018.

[28]   Ujjwal Singh et al. "Hybrid multi-model ensemble learning for re-
       constructing gridded runoff of Europe for 500 years". In: *Informa-
       tion Fusion* 97 (2023), p. 101807.

[29]   Carolyn Lober et al. "Bias correction of 20 years of IMERG satellite precipitation data over Canada and Alaska". In: *J. Hydrol. Reg. Stud.* 47 (2023), p. 101386.

[30]   Priyam Deka and Ujjwal Saha. "Introduction of k-means clustering into random cascade model for disaggregation of rainfall from daily to 1-hour resolution with improved preservation of extreme rainfall". In: *J. Hydrol.* 620 (2023), p. 129478.

[31]   Demetris Koutsoyiannis and Christian Onof. "Rainfall disaggregation using adjusting procedures on a Poisson cluster model". In: *J. Hydrol.* 246.1-4 (2001), pp. 109–122.

[32]   M-L Segond, C Onof, and HS Wheater. "Spatial–temporal disaggregation of daily rainfall from a generalized linear model". In: *J. Hydrol.* 331.3-4 (2006), pp. 674–689.

[33]   Marie-Laure Segond et al. "Simulation and spatio-temporal disaggregation of multi-site rainfall data for urban drainage applications". In: *Hydrol Sci J.* 52.5 (2007), pp. 917–935.

[34]   Taikan Oki and Shinjiro Kanae. "Global hydrological cycles and world water resources". In: *science* 313.5790 (2006), pp. 1068–1072.

[35]   MHJ Van Huijgevoort et al. "Global multimodel analysis of drought in runoff for the second half of the twentieth century". In: *J. Hydrometeorol.* 14.5 (2013), pp. 1535–1552.

[36]   Charles J Vörösmarty et al. "Global threats to human water security and river biodiversity". In: *Nature* 467.7315 (2010), pp. 555–561.

[37]   Lukas Gudmundsson, Sonia I Seneviratne, and Xuebin Zhang. "Anthropogenic climate change detected in European renewable freshwater resources". In: *Nat. Clim. Change.* 7.11 (2017), pp. 813–816.

[38]   Mikhail Ivanovich Budyko. *Climate and life.* Academic Press, Inc., 1974.

[39]   P Schreiber. "Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa". In: *Z. Meteorol* 21.10 (1904), pp. 441–452.

[40] EM Ol'Dekop. "On evaporation from the surface of river basins". In: *Meteorol. obs.* 4 (1911), p. 200.

[41] Lucien Turkish. "The water balance of soils: relations between precipitation, 'e vaporization and é flow". In: *Journ 'e es de l'hydraulique* 3.1 (1955), pp. 36–44.

[42] JG Pike. "The estimation of annual run-off from meteorological data in a tropical climate". In: *J. Hydrol.* 2.2 (1964), pp. 116–123.

[43] Yuting Yang et al. "Hydrologic implications of vegetation response to elevated CO 2 in climate projections". In: *Nat. Clim. Change.* 9.1 (2019), pp. 44–48.

[44] Xinyao Zhou et al. "Reconstructed natural runoff helps to quantify the relationship between upstream water use and downstream water scarcity in China's river basins". In: *Hydrol Earth Syst Sci* 23.5 (2019), pp. 2491–2505.

[45] Andreas Pauling et al. "Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation". In: *Clim. Dyn.* 26.4 (2006), pp. 387–405.

[46] Jürg Luterbacher et al. "European seasonal and annual temperature variability, trends, and extremes since 1500". In: *Science* 303.5663 (2004), pp. 1499–1503.

[47] Nathan Wells, Steve Goddard, and Michael J Hayes. "A self-calibrating Palmer drought severity index". In: *J. Clim.* 17.12 (2004), pp. 2335–2351.

[48] Edward R Cook et al. "Old World megadroughts and pluvials during the Common Era". In: *Sci. Adv.* 1.10 (2015), e1500561.

[49] Y Markonis et al. "Persistent multi-scale fluctuations shift European hydroclimate to its millennial boundaries". In: *Nat. Commun.* 9.1 (2018), pp. 1–12.

[50] Aijun Guo et al. "Uncertainty analysis of water availability assessment through the Budyko framework". In: *J. Hydrol.* 576 (2019), pp. 396–407.

[51] AP Sergeev et al. "Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals". In: *Catena* 174 (2019), pp. 425–435.

[52] Goutam Konapala et al. "Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US". In: *Environ. Res. Lett.* 15.10 (2020), p. 104022.

[53] Louise Slater et al. "Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models". In: *Earth Syst. Sci. Data Discuss.* (2022), pp. 1–35.

[54] Ujjwal Singh et al. "Disaggregating IMERG Satellite Precipitation Over Czech Republic: An Innovative Approach using Hybrid Extreme Gradient Boosting based on Fuzzy Spatial-Temporal Multivariate Clustering". In: (2023).

[55] Leonardo Brain García Fernández, Anna Diva Plasencia Lotufo, and Carlos Roberto Minussi. "Development of a Short-Term Electrical Load Forecasting in Disaggregated Levels Using a Hybrid Modified Fuzzy-ARTMAP Strategy". In: *Energies* 16.10 (2023), p. 4110.

[56] Honglin Zhu et al. "Towards an Accurate and Reliable Downscaling Scheme for High-Spatial-Resolution Precipitation Data". In: *Remote Sens.* 15.10 (2023), p. 2640.

[57] Badr-eddine Sebbar et al. "Machine-Learning-Based Downscaling of Hourly ERA5-Land Air Temperature over Mountainous Regions". In: *Atmosphere* 14.4 (2023), p. 610.

[58] Gabriel Jouan et al. "Gaussian mixture models for clustering and calibration of ensemble weather forecasts". In: *Discrete Contin. Dyn. Syst. - S.* 16.2 (2023), pp. 309–328.

[59] Huihui Zhang et al. "Spatial and temporal downscaling of TRMM precipitation with novel algorithms". In: *J. Hydrometeorol.* 21.6 (2020), pp. 1259–1278.

[60] Bramha Dutt Vishwakarma, Jinwei Zhang, and Nico Sneeuw. "Downscaling GRACE total water storage change using partial least squares regression". In: *Sci. Data* 8.1 (2021), p. 95.

[61]   Kou-lin Hsu et al. "Precipitation estimation from remotely sensed information using artificial neural networks". In: *J. Appl. Meteorol.* 36.9 (1997), pp. 1176–1190.

[62]   Li Chen et al. "Downscaling of GRACE-derived groundwater storage based on the random forest model". In: *Remote Sens.* 11.24 (2019), p. 2979.

[63]   Rabeea Noor et al. "Combining APHRODITE Rain Gauges-Based Precipitation with Downscaled-TRMM Data to Translate High-Resolution Precipitation Estimates in the Indus Basin". In: *Remote Sens.* 15.2 (2023), p. 318.

[64]   Arfan Arshad et al. "Reconstructing high-resolution gridded precipitation data using an improved downscaling approach over the high altitude mountain regions of Upper Indus Basin (UIB)". In: *Sci. Total Environ.* 784 (2021), p. 147140.

[65]   Gabriel G Katul et al. "Evapotranspiration: a process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system". In: *Rev. Geophys.* 50.3 (2012).

[66]   Paul Ekness and Timothy O Randhir. "Effect of climate and land cover changes on watershed runoff: A multivariate assessment for storm water management". In: *J. Geophys. Res.* 120.9 (2015), pp. 1785–1796.

[67]   Lei Dong et al. "Wind power day-ahead prediction with cluster analysis of NWP". In: *Renew. sustain. energy rev.* 60 (2016), pp. 1206–1212.

[68]   Huade Guan, John L Wilson, and Hongjie Xie. "A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain". In: *J. Hydrol.* 375.3-4 (2009), pp. 578–588.

[69]   Guanghua Xu et al. "Spatial downscaling of TRMM precipitation product using a combined multifractal and regression approach: Demonstration for South China". In: *Water* 7.6 (2015), pp. 3083–3102.

[70]   Robert J Hijmans et al. "Package 'terra'". In: *Maintainer: Vienna, Austria* (2022).

[71]   Ujjwal Singh et al. "ScatSat-1 leaf area index product: models comparison, development, and validation over cropland". In: *IEEE Geosci. Remote. Sens.* 17.4 (2019), pp. 563–567.

[72]   Dr S Santhosh Baboo and M Renuka Devi. "An analysis of different resampling methods in Coimbatore, District". In: *Glob. j. comput. sci. technol.* 10.15 (2010).

[73]   Shreyas Fadnavis. "Image interpolation techniques in digital image processing: an overview". In: *Int. j. eng. res.* 4.10 (2014), pp. 70–73.

[74]   Gregor Perich et al. "Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks". In: *Field Crops Res.* 292 (2023), p. 108824.

[75]   Wilhelm Burger and Mark J Burge. *Digital Image Processing: An Algorithmic Introduction.* Springer Nature, 2022.

[76]   James C Bezdek. *Pattern recognition with fuzzy objective function algorithms.* Springer Science & Business Media, 2013.

[77]   Alexis Sardá-Espinosa. "Comparing time-series clustering algorithms in r using the dtwclust package". In: *R package vignette* 12 (2017), p. 41.

[78]   Ángel López-Oriona et al. "Quantile-based fuzzy C-means clustering of multivariate time series: Robust techniques". In: *Int J Approx Reason.* 150 (2022), pp. 55–82.

[79]   Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* 2016, pp. 785–794.

[80]   Zhen Li et al. "An improved cyclic multi model-eXtreme gradient boosting (CMM-XGBoost) forecasting algorithm on the GNSS vertical time series". In: *Adv. Space Res.* 71.1 (2023), pp. 912–935.

[81]   Ali Mokhtar et al. "Estimation of SPEI meteorological drought using machine learning algorithms". In: *IEEE Access* 9 (2021), pp. 65503–65523.

[82]   Yongchan Kim et al. "The future water vulnerability assessment of the Seoul metropolitan area using a hybrid framework composed of physically-based and deep-learning-based hydrologic models". In: *Stoch. Environ. Res. Risk Assess.* (2023), pp. 1–22.

[83]   Shuai Chen and Xiaohong Ruan. "A hybrid Budyko-type regression framework for estimating baseflow from climate and catchment attributes". In: *J. Hydrol.* (2023), p. 129118.

[84]   John M Quilty, Anna E Sikorska-Senoner, and David Hah. "A stochastic conceptual-data-driven approach for improved hydrological simulations". In: *Environ Model Softw.* 149 (2022), p. 105326.

[85]   Yue Yu et al. "Intermittent solar power hybrid forecasting system based on pattern recognition and feature extraction". In: *Energy Conv. Manag.* 277 (2023), p. 116579.

[86]   Ammar Mohammed and Rania Kora. "A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges". In: *J. King Saud Univ. - Comput. Inf. Sci.* (2023).

[87]   Sani Isah Abba et al. "Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration". In: *IEEE Access.* 8 (2020), pp. 157218–157237.

[88]   Di Tian et al. "A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information". In: *Stoch Environ Res Risk Assess.* 36.8 (2022), pp. 2353–2375.

[89]   Jung-Woo Kim and Yakov A Pachepsky. "Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation". In: *J. Hydrol.* 394.3-4 (2010), pp. 305–314.

[90] Lukas Gudmundsson and Sonia I Seneviratne. "Observation-based gridded runoff estimates for Europe (E-RUN version 1.1)". In: *Earth Syst. Sci. Data* 8.2 (2016), pp. 279–295.

[91] Gionata Ghiggi et al. "GRUN: an observation-based global gridded runoff dataset from 1902 to 2014". In: *Earth Syst. Sci. Data* 11.4 (2019), pp. 1655–1674.

[92] Frederik Kratzert et al. "Toward improved predictions in ungauged basins: Exploiting the power of machine learning". In: *Water Resour. Res* 55.12 (2019), pp. 11344–11354.

[93] Hristos Tyralis et al. "Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS". In: *J. Hydrol.* 577 (2019), p. 123957.

[94] Yuhang Zhang et al. "Comparing machine learning and deep learning models for probabilistic post-processing of satellite precipitation-driven streamflow simulation". In: *Hydrol. Earth Syst. Sci.* (2022), pp. 1–41.

[95] Subhrendu Gangopadhyay et al. "A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles". In: *Water Resour. Res.* 45.6 (2009).

[96] CL Wu and Kwok-Wing Chau. "Data-driven models for monthly streamflow time series prediction". In: *Eng. Appl. Artif. Intell.* 23.8 (2010), pp. 1350–1367.

[97] Maha Shabbir, Sohail Chand, and Farhat Iqbal. "Prediction of river inflow of the major tributaries of Indus river basin using hybrids of EEMD and LMD methods". In: *Arab. J. Geosci.* 16.4 (2023), p. 257.

[98] Xiang Yu et al. "Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting". In: *J. Hydrol.* 582 (2020), p. 124293.

[99] Ruonan Hao and Zhixu Bai. "Comparative Study for Daily Streamflow Simulation with Different Machine Learning Methods". In: *Water* 15.6 (2023), p. 1179.

[100]   Zainab Abdulelah Al-Sudani, Sinan Q Salih, Zaher Mundher Yaseen, et al. "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation". In: *J. Hydrol.* 573 (2019), pp. 1–12.

[101]   Rana Muhammad Adnan et al. "Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs". In: *J. Hydrol.* 586 (2020), p. 124371.

[102]   Babak Vaheddoost, Mir Jafar Sadegh Safari, and Mustafa Utku Yilmaz. "Rainfall-Runoff Simulation in Ungauged Tributary Streams Using Drainage Area Ratio-Based Multivariate Adaptive Regression Spline and Random Forest Hybrid Models". In: *Pure Appl. Geophys.* (2023), pp. 1–18.

[103]   Sandra M Hauswirth et al. "The suitability of a hybrid framework including data driven approaches for hydrological forecasting". In: *Hydrol. Earth Syst. Sci. Discuss.* (2022), pp. 1–20.

[104]   Chenchen Zhang et al. "RUNOFF ESTIMATION BASED ON HYBRID-PHYSICS-DATA MODEL". In: *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 3 (2022), pp. 347–352.

[105]   Peiman Parisouj et al. "Physics-Informed Data-Driven Model for Predicting Streamflow: A Case Study of the Voshmgir Basin, Iran". In: *Appl. Sci.* 12.15 (2022), p. 7464.

[106]   Ayşe Doğru, Selim Buyrukoğlu, and Murat Arı. "A hybrid super ensemble learning model for the early-stage prediction of diabetes risk". In: *Med Biol Eng Comput.* (2023), pp. 1–13.

[107]   Selim Buyrukoğlu and Serkan Savaş. "Stacked-based ensemble machine learning model for positioning footballer". In: *Arab. J. Sci. Eng.* 48.2 (2023), pp. 1371–1383.

[108]   Mirza Mumtaz Zahoor et al. "A New Deep Hybrid Boosted and Ensemble Learning-Based Brain Tumor Analysis Using MRI". In: *Sensors* 22.7 (2022), p. 2726.

[109]   Joao Mendes-Moreira et al. "Ensemble approaches for regression: A survey". In: *ACM Comput. Surv.* 45.1 (2012), pp. 1–40.

[110]   Hristos Tyralis, Georgia Papacharalampous, and Andreas Langousis. "Super ensemble learning for daily streamflow forecasting: large-scale demonstration and comparison with multiple machine learning algorithms". In: *Neural. Comput. Appl.* (2020), pp. 1–16.

[111]   Lirong Cai et al. "Global models and predictions of plant diversity based on advanced machine learning techniques". In: *New Phytol.* 237.4 (2023), pp. 1432–1445.

[112]   J Sakari Salonen et al. "Calibrating aquatic microfossil proxies with regression-tree ensembles: Cross-validation with modern chironomid and diatom data". In: *The Holocene.* 26.7 (2016), pp. 1040–1048.

[113]   J Sakari Salonen et al. "A North European pollen–climate calibration set: analysing the climatic responses of a biological proxy using novel regression tree methods". In: *Quat. Sci. Rev.* 45 (2012), pp. 95–110.

[114]   Shicheng Li and James Yang. "Improved river water-stage forecasts by ensemble learning". In: *Eng Comput.* (2022), pp. 1–19.

[115]   Junliang Fan et al. "Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China". In: *Agric. For. Meteorol.* 263 (2018), pp. 225–241.

[116]   Zhong-kai Feng et al. "Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization". In: *J. Hydrol.* 583 (2020), p. 124627.

[117]   Mohammad Rezaie-Balf et al. "Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of Aswan High Dam". In: *Hydrol. Sci. J.* 64.13 (2019), pp. 1629–1646.

[118] Mohammad Reza Najafi and Hamid Moradkhani. "Ensemble combination of seasonal streamflow forecasts". In: *J. Hydrol. Eng.* 21.1 (2016), p. 04015043.

[119] MI Budyko. "The heat balance of the earth's surface, US Dept. of Commerce". In: *Weather Bureau, Washington, DC, USA* (1958).

[120] Omer Sagi and Lior Rokach. "Ensemble learning: A survey". In: *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 8.4 (2018), e1249.

[121] Sergio González et al. "A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities". In: *Inf. Fusion.* 64 (2020), pp. 205–237.

[122] Yannan Bin et al. "Prediction of neuropeptides from sequence information using ensemble classifier and hybrid features". In: *J. Proteome Res.* 19.9 (2020), pp. 3732–3740.

[123] Cheng-Wei Ju et al. "Stacked ensemble machine learning for range-separation parameters". In: *J. Phys. Chem.* 12.39 (2021), pp. 9516–9524.

[124] Artittayapron Rojarath, Wararat Songpan, and Chakrit Pong-inwong. "Improved ensemble learning for classification techniques based on majority voting". In: *2016 7th IEEE international conference on software engineering and service science (ICSESS)*. IEEE. 2016, pp. 107–110.

[125] Sareh Sayari et al. "Meta-learner methods in forecasting regulated and natural river flow". In: *Arab. J. Geosci.* 15.11 (2022), p. 1051.

[126] Omid Zandi et al. "Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area". In: *Atmos. Res.* 272 (2022), p. 106159.

[127] SR Divyasri, R Saranya, and P Kathiravan. "Comprehensive analysis of Classical Machine Learning models and Ensemble methods for predicting Crime in urban society". In: (2023). DOI: https://doi.org/10.21203/rs.3.rs-2550707/v1.

[128]   Mohammad Zounemat-Kermani et al. "Ensemble machine learning paradigms in hydrology: A review". In: *J. Hydrol.* 598 (2021), p. 126266.

[129]   Matt W Gardner and SR Dorling. "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences". In: *Atmos. Environ.* 32.14-15 (1998), pp. 2627–2636.

[130]   James W Taylor. "A quantile regression neural network approach to estimating the conditional density of multiperiod returns". In: *J Forecast*. 19.4 (2000), pp. 299–311.

[131]   Alex J Cannon. "Quantile regression neural networks: Implementation in R and application to precipitation downscaling". In: *Comput Geosci.* 37.9 (2011), pp. 1277–1284.

[132]   Colin Chen. "A finite smoothing algorithm for quantile regression". In: *J. Comput. Graph. Stat.* 16.1 (2007), pp. 136–164.

[133]   Peter Bühlmann and Bin Yu. "Boosting with the L 2 loss: regression and classification". In: *J. Am. Stat. Assoc.* 98.462 (2003), pp. 324–339.

[134]   Matthias Schmid and Torsten Hothorn. "Boosting additive models using component-wise P-splines". In: *Comput Stat Data Anal.* 53.2 (2008), pp. 298–311.

[135]   Kelly O Maloney, Matthias Schmid, and Donald E Weller. "Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages". In: *Methods Ecol. Evol.* 3.1 (2012), pp. 116–128.

[136]   Leo Breiman et al. "Classification and regression trees. Wadsworth Int". In: *Group* 37.15 (1984), pp. 237–251.

[137]   Torsten Hothorn, Kurt Hornik, and Achim Zeileis. "Unbiased recursive partitioning: A conditional inference framework". In: *J. Comput. Graph. Stat.stics* 15.3 (2006), pp. 651–674.

[138]   Achim Zeileis, Torsten Hothorn, and Kurt Hornik. "Model-based recursive partitioning". In: *J. Comput. Graph. Stat.* 17.2 (2008), pp. 492–514.

[139]   Carolin Strobl et al. "Bias in random forest variable importance measures: Illustrations, sources and a solution". In: *BMC Bioinf.* 8.1 (2007), p. 25.

[140]   Jerome H Friedman. "Multivariate adaptive regression splines". In: *Ann. Stat.* (1991), pp. 1–67.

[141]   Peter J Huber. "Robust estimation of a location parameter". In: *Ann. math. stat.* (1964).

[142]   J Doyne Farmer and John J Sidorowich. "Predicting chaotic time series". In: *Phys. Rev. Lett.* 59.8 (1987), p. 845.

[143]   Klaus Hechenbichler and Klaus Schliep. *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification.* 2004.

[144]   Nicolai Meinshausen. "Quantile regression forests". In: *J Mach Learn Res* . 7.Jun (2006), pp. 983–999.

[145]   Naomi Altman and Martin Krzywinski. *Points of Significance: Ensemble methods: bagging and random forests.* 2017.

[146]   Leo Breiman. "Random forests". In: *Mach. Learn.* 45.1 (2001), pp. 5–32.

[147]   Andrew Gelman et al. "A weakly informative default prior distribution for logistic and other regression models". In: *Ann Appl Stat.* 2.4 (2008), pp. 1360–1383.

[148]   Andrew Gelman et al. "Package 'arm'". In: *R topics documented* (2020).

[149]   Jim E Griffin, Philip J Brown, et al. "Inference with normal-gamma prior distributions in regression problems". In: *Bayesian Anal.* 5.1 (2010), pp. 171–188.

[150]   Robert B Gramacy, Maintainer Robert B Gramacy, and Monotone data augmentation extends this Bayesian. "Package 'monomvn'". In: *R topics documented* (2019).

[151]   Bhaskar J Choudhury. "Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model". In: *J. Hydrol.* 216.1-2 (1999), pp. 99–110.

[152] Yongqiang Zhang et al. "Decadal trends in evaporation from global energy and water balances". In: *J. Hydrometeorol.* 13.1 (2012), pp. 379–391.

[153] ML Roderick et al. "A general framework for understanding the response of the water cycle to global warming over land and ocean". In: *Hydrol. Earth Syst. Sci.* 18.5 (2014), pp. 1575–1589.

[154] Ludovic Oudin et al. "Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling". In: *J. Hydrol.* 303.1-4 (2005), pp. 290–306.

[155] Luc Anselin. "Local indicators of spatial association—LISA". In: *Geogr. Anal* 27.2 (1995), pp. 93–115.

[156] Christian W Dawson, Robert J Abrahart, and Linda M See. "HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts". In: *Environ. Model. Softw.* 22.7 (2007), pp. 1034–1052.

[157] Hoshin V Gupta et al. "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling". In: *J. Hydrol.* 377.1-2 (2009), pp. 80–91.

[158] J Eamonn Nash and Jonh V Sutcliffe. "River flow forecasting through conceptual models part I—A discussion of principles". In: *J. Hydrol.* 10.3 (1970), pp. 282–290.

[159] Max Kuhn et al. "Package 'caret'". In: *The R Journal* (2020).

[160] Zachary A Deane-Mayer and JE Knowles. "caretEnsemble: ensembles of caret models". In: *R package version* 2.0 (2016).

[161] Alex J Cannon and Maintainer Alex J Cannon. "Package 'qrnn'". In: (2015).

[162] Torsten Hothorn et al. *Package 'mboost'*. 2013.

[163] Terry Therneau et al. "Package 'rpart'". In: *Available online: cran. ma. ic. ac. uk/web/packages/rpart/rpart. pdf (accessed on 20 April 2016)* (2015).

[164] Torsten Hothorn, Achim Zeileis, and Maintainer Torsten Hothorn. "Package 'partykit'". In: (2021).

[165] Tianqi Chen et al. "Xgboost: extreme gradient boosting". In: *R package version 0.4-2* 1.4 (2015), pp. 1–4.

[166] Maintainer Stephen Milborrow. *Package 'earth'*. 2020.

[167] Brian Ripley et al. "Package 'mass'". In: *Cran R* 538 (2013).

[168] Klaus Schliep, Klaus Hechenbichler, and Maintainer Klaus Schliep. *Package 'kknn'*. 2016.

[169] Suggests RColorBrewer and Maintainer Andy Liaw. "Package 'randomForest'". In: *University of California, Berkeley: Berkeley, CA, USA* (2018).

[170] Andrew Gelman et al. "Package 'arm'". In: *data analysis using regression and multilevel/hierarchical models. R package, version* 9 (2016).

[171] Sebastian Eichhorn. "Disaggregating population data and evaluating the accuracy of modeled high-resolution population distribution—The case study of Germany". In: *Sustainability* 12.10 (2020), p. 3976.

[172] Wei Sun et al. "Evaluation and Correction of GPM IMERG Precipitation Products over the Capital Circle in Northeast China at Multiple Spatiotemporal Scales". In: *Adv. Meteorol.* 2018 (2018).

[173] Mohammad Reza Eini et al. "Characterization of Precipitation Concentration Indicators and Their Variations in a Central European Region". In: (2023).

[174] Earth Engine Data Catalog. *SRTM Digital Elevation Data Version 4*. Accessed: 2023-07-20. 2002. URL: https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4.

[175] G. Ghiggi et al. "GRUN: an observation-based global gridded runoff dataset from 1902 to 2014". In: *Earth Syst. Sci. Data* 11.4 (2019), pp. 1655–1674. DOI: 10.5194/essd-11-1655-2019. URL: https://www.earth-syst-sci-data.net/11/1655/2019/.

[176] Gionata Ghiggi et al. "G-RUN ENSEMBLE: A Multi-Forcing Observation-Based Global Runoff Reanalysis". In: *Water Resour. Res* 57.5 (2021), e2020WR028787.

[177] Balazs M Fekete, Charles J Vörösmarty, and Wolfgang Grabs. *Global, composite runoff fields based on observed river discharge and simulated water balances.* 1999.

[178] Edoardo Daly et al. "Linking parametric and water-balance models of the Budyko and Turc spaces". In: *Adv Water Resour.* 134 (2019), p. 103435.

[179] GJ Huffman et al. "GPM IMERG Final Precipitation L3 1 Month 0.1 Degree× 0.1 Degree V06". In: *Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA* (2019).

[180] Andrés Navarro et al. "Assessment of IMERG precipitation estimates over Europe". In: *Remote Sens.* 11.21 (2019), p. 2470.

[181] Jéssica G Nascimento et al. "Evaluating the Latest IMERG Products in a Subtropical Climate: The Case of Paraná State, Brazil". In: *Remote Sens.* 13.5 (2021), p. 906.

[182] Thomas Ramsauer, Thomas Weiß, and Philip Marzahn. "Comparison of the GPM IMERG final precipitation product to RADOLAN weather radar data over the topographically and climatically diverse Germany". In: *Remote Sens.* 10.12 (2018), p. 2029.

[183] Thomas Mölg et al. "Temporal precipitation variability versus altitude on a tropical high mountain: Observations and mesoscale atmospheric modelling". In: *Q J R Meteorol Soc.* 135.643 (2009), pp. 1439–1455.

[184]   Conor Murphy et al. "A 305-year continuous monthly rainfall series for the island of Ireland (1711–2016)." In: *Clim. Past* 14.3 (2018), pp. 413–440.

[185]   Matthew S Armstrong, Anthony S Kiem, and Tessa R Vance. "Comparing instrumental, palaeoclimate, and projected rainfall data: Implications for water resources management and hydrological modelling". In: *J. Hydrol.: Reg. Stud.* 31 (2020), p. 100728.

[186]   Michał Woźniak, Manuel Grana, and Emilio Corchado. "A survey of multiple classifier systems as hybrid systems". In: *Inf. Fusion.* 16 (2014), pp. 3–17.

[187]   Huajin Lei, Hongyu Zhao, and Tianqi Ao. "A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China". In: *Hydrol. Earth Syst. Sci* 26.11 (2022), pp. 2969–2995.

[188]   Basil Kraft et al. "Towards hybrid modeling of the global hydrological cycle". In: *Hydrol. Earth Syst. Sci.* (2021).

[189]   Boen Zhang et al. "A vine copula-based polynomial chaos framework for improving multi-model hydroclimatic projections at a multi-decadal convection-permitting scale". In: *Water Resour. Res.* (2022), e2022WR031954.

[190]   Hapu Arachchige Prasantha Hapuarachchi et al. "Development of a national 7-day ensemble streamflow forecasting service for Australia". In: *Hydrol Earth Syst Sci.* 26.18 (2022), pp. 4801–4821.

[191]   Hooman Gholami et al. "Multi-GCM ensemble model for reduction of uncertainty in runoff projections". In: *Stoch Environ Res Risk Assess.* (2022), pp. 1–12.

[192]   Isa Ebtehaj and Hossein Bonakdari. "A comprehensive comparison of the fifth and sixth phases of the coupled model intercomparison project based on the Canadian earth system models in spatiotemporal variability of long-term flood susceptibility using remote sensing and flood frequency analysis". In: *J. Hydrol.* 617 (2023), p. 128851.

[193] Daniele Masseroni et al. "65-year changes of annual streamflow volumes across Europe with a focus on the Mediterranean basin". In: *Hydrol. Earth Syst. Sci. Discuss.* (2020), pp. 1–16.

[194] Georgia Papacharalampous, Hristos Tyralis, and Demetris Koutsoyiannis. "Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A multiple-case study from Greece". In: *Water Resour. Manage.* 32.15 (2018), pp. 5207–5239.

Part VIII

Current research - publications

CHAPTER 10

PUBLICATION

10.1 List of publications

1. **Ujjwal Singh**, Petr Maca, Martin Hanel, Yannis Markonis, Rama Rao Nidamanuri, Sadaf Nasreen, Johanna Ruth Blöcher, Filip Strnad, Jiri Vorel, Lubomir Riha, Akhilesh Singh Raghubanshi, Hybrid multi-model ensemble learning for reconstructing gridded runoff of Europe for 500 years, Information Fusion, https://doi.org/10.1016/j.inffus.2023.101807

2. **Ujjwal Singh**, Sadaf Nasreen, Gaurav Tripathi,Pragya Mehrishi, Rajni Kumar Pradhan, Zuzana Bestakova, Vivek Vikram Singh, K C Gouda, Laxmi Kant Sharma, Kiran Jalem, Petr Maca, Rama Rao Nidamanuri,Akhilesh Singh Raghubanshi, Yannis Markonis,Rakovec Oldˇrich, Martin Hanel. Disaggregating IMERG Satellite Precipitation Over Czech Republic: An Innovative Approach using Hybrid Extreme Gradient Boosting based on Fuzzy Spatial-Temporal Multivariate Clus-

113

tering, Submitted in "Journal of Big Data"

3. Argha Banerjee, **Ujjwal Singh**, and Chintan Sheth. Disaggregating geodetic glacier mass balance to annual scale using remote-sensing proxies. Journal of Glaciology, DOI: https://doi.org/10.1017/jog.2022.89

4. Sadaf Nasreen, Markéta Součková, Mijael Rodrigo Vargas Godoy, **Ujjwal Singh**, Yannis Markonis, Rohini Kumar, Oldrich Rakovec, and Martin Hanel. A 500-year annual runoff reconstruction for 14 selected european catchments. Earth System Science Data, https://doi.org/10.5194/essd-14-4035-2022

5. Zuzana Beštáková, Filip Strnad, Mijael Rodrigo Vargas Godoy, **Ujjwal Singh**, Yannis Markonis, Martin Hanel, Petr Máca, and Jan Kyselʿy. Changes of the aridity index in europe from 1950 to 2019. Theoretical and Applied Climatology, https://doi.org/10.1007/s00704-022-04266-3