Mendel University in Brno
Faculty of Regional Development and International Studies

# Analysis of Risk Factors Associated with Dengue in Southeast Asia

Diploma thesis

Author: Bc. Klára Urbanová
Supervisor: Mgr. et Mgr. Miroslav Horák, Ph.D.
Brno 2016

## DECLARATION

I declare that I carried out this thesis independently, and only with the cited sources, literature and other professional sources.

I agree that my work will be published in accordance with Section 47b of Act No. 111/1998 Coll. on Higher Education as amended thereafter and in accordance with the Guidelines on Publishing University Student Theses.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that Mendel University in Brno has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Before closing a license agreement on the use of my thesis with another person (subject) I undertake to request for a written statement of the university that the license agreement in question is not in conflict with the legitimate interests of the university, and undertake to pay any contribution, if eligible, to the costs associated with the creation of the thesis, up to their actual amount.

In Brno, December 2015

……

Signature

**Acknowledgements**

I would like to express my greatest appreciation of the time, effort and relevant advice my supervisor Mgr. et Mgr. Miroslav Horák, Ph.D. has provided me with throughout the thesis creation process. My gratitude also goes to my consultant Ing. Jakub Hejč, Brno University of Technology, Department of Biomedical Engineering for his consultations on the mathematical and statistical data processing. Finally, my greatest thanks are to my partner and my closest family, who have given me the love and strength to continue my effort.

**Abstract**

The thesis focuses on analyses of relations between risk factors and dengue fever incidence in a selected area of Southeast Asia region. Specifically, it focuses on the Philippines, where the governmental organizations provide relatively sufficient amount of data. First, general problematic of Neglected Tropical Diseases with focus on dengue fever is introduced. Its transmission and influencing factors are specified. Then, current methods of dengue incidence and risk factors relationship modelling are reviewed. Furthermore, statistical processing of available data is carried out, especially, in terms of detailed analysis of the relationship between climatic factors and dengue incidence in the Philippines. Based on the obtained results, a mathematical model describing the relationship between dengue, cumulative precipitation and mean temperature on regional and weekly basis, is created. Model estimation is performed with generalized linear regression by applying negative binomial distribution. With the model, the dengue incidence dependency on selected risk factors was verified. Concurrently, overall complexity of the disease development and transmission was verified by the model.

**Key words**

Neglected Tropical Diseases, dengue, Southeast Asia, the Philippines, risk factors, generalized linear model

**Abstrakt**

Práce se zaměřuje na analýzu vztahů mezi rizikovými faktory a výskytem horečky dengue ve vybrané části Jihovýchodní Asie. Práce je konkrétně zacílena na Filipíny, jejíž vládní organizace poskytují relativně dostatečné množství dat. Na úvod je nastíněna obecná problematika tzv. zanedbaných tropických nemocí s podrobným zaměřením na horečku dengue, způsoby jejího šíření a ovlivňující faktory, včetně současných metod modelování vztahu mezi rizikovými faktory a výskytem horečky dengue. Praktická část se věnuje statistickému zpracování dostupných dat, zejména podrobné analýze vztahu mezi klimatickými faktory a výskytem dengue na Filipínách. Na základě obdržených poznatků je vytvořen matematický model popisující vztah mezi výskytem dengue, kumulativními srážkami a průměrnou teplotou na regionální a týdenní bázi. Odhad modelu je proveden pomocí zobecněné lineární regrese s využitím negativně binomického rozdělení. Pomocí modelu byla ověřena závislost incidence dengue na vybraných rizikových faktorech, ale také celková komplexnost rozvoje a šíření nemoci, kterou je nutné sledovat na regionální úrovni.

**Klíčová slova**

Zanedbávané tropické nemoci, dengue, Jihovýchodní Asie, Filipíny, rizikové faktory, zobecněné lineární modely

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **ARMM** | Autonomous Region of Muslim Mindanao |
| **ASEAN** | Association of Southeast Asian Nations |
| **ASEAN-NDI** | ASEAN Network for Drugs, Diagnostics, Vaccines and Traditional Medicines Innovation |
| **AR** | Autoregressive |
| **ARIMA** | Autoregressive integrated moving average |
| **ARMA** | Autoregressive moving average |
| **AYP** | Average year population |
| **CAR** | Cordillera Administrative Region |
| **CI** | Cumulative incidence |
| **DF** | Dengue fever |
| **DHF** | Dengue haemorrhagic fever |
| **DALYs** | Disability-adjusted life years |
| **DSS** | Dengue shock syndrome |
| **GIS** | Geographically Integrated Systems |
| **GWR** | Geographically weighted regression |
| **ID** | Incidence density |
| **IP** | Incidence proportion |
| **I** | Integrated |
| **Lao PDR** | Lao People's Democratic Republic |
| **LHM** | Lag at half of maximum |
| **MA** | Moving average |
| **MDGs** | Millennium Development Goals |
| **NCR** | National Capital Region |
| **NTDs** | Neglected Tropical Diseases |
| **NZDs** | Neglected Zoonotic Diseases |
| **R&D** | Research and development |
| **RNA** | Ribonucleic acid |
| **SARIMA** | Seasonal autoregressive integrated moving average |
| **SEA** | Southeast Asia |
| **SEAR** | South-East Asia Region |
| **SDGs** | Sustainable Development Goals |
| **UN** | United Nations |
| **US** | United States |
| **UHC** | Universal health coverage |
| **VAR** | Vector auto-regression |
| **WHO** | World Health Organization |
| **WHO-TDR** | World Health Organization Special Programme for Research and Training in Tropical Diseases |
| **YLDs** | The years lived with a disability |
| **YLLs** | The years of life lost to due to premature mortality |

# 1 INTRODUCTION

The World Health Organization created a list of 17 Neglected Tropical Diseases (NTDs). Such diseases became a priority of focus in terms of Neglected Tropical Diseases. They affect over 1 billion population worldwide and are often referred to as communicable and infectious.[1] Generally, it is believed that these diseases have been neglected by national and worldwide governments and organizations for decades. Therefore, there have been insufficient amount of involvement e.g.; prevention and control programs, invention and supply of drugs and investment in general. According to some, the diseases have been also neglected because they influence the world's poorest population. The diseases are characterized by occurrence in hot, tropical geographical areas, affecting rural or suburban populations and their presence in low income countries. They are disabling, disfiguring and stigmatizing.[2]

Within the Southeast Asia region a large hidden burden of poverty and NTDs is present. Large portion of the population living in the region is affected by at least one of NTDs. They represent a crucial public health issue within the region. The diseases promote poverty in the region, as they are chronic and debilitating. Thus, seriously affect the productive and social lives of the population. The two countries at high risk are Indonesia and the Philippines, as they share 30% of the region's population living in extreme poverty. It is essential, for the Southeast Asia region as a whole and each specific country, to focus on active surveillance and defining the extent or burden of the diseases. Simultaneously, a need for new medicine, diagnostics and vaccines is in place.[3]

Since the 19$^{th}$ century, when dengue was a scarcely recognized disease, it arose to be the most important mosquito borne viral disease in the world. Currently, it is present in 112 countries all over the world. It occurs mainly in tropical and sub-tropical areas with incidence primarily located in urban and suburban locations. Around 2.5 to 3 billion people worldwide are estimated to be at risk.[4] It occurs in most Asian countries and is locally a leading cause of hospitalization and death.[5] The vector responsible for majority of dengue infections and its transmission is the *Aedes aegypti* mosquito.[4] Dengue represents a health burden in most of the Southeast Asian countries, Cambodia, Malaysia, Vietnam and the Philippines account for over 90% of the total cases registered within the region. In the Philippines, dengue is the most significant vector-borne disease.[6]

Epidemiology is an important field of study and research in understanding and implementing knowledge to control diseases and advance public health. There exist many different sources of data for epidemiology e.g.; census, vital registrations etc. The quality of data is essential for epidemiology, although, there exist distinct sources of errors. Epidemiology

defines and utilizes many terms such as; incidence, prevalence or risks and odds.[7] It also uses different statistical and mathematical methods and models when studying aspects and behaviors of distinct diseases, for example, Logistic regression[8] or Poisson regression model[7].

When studying diseases, specifically dengue fever, it is essential to focus on the risk factors which occur in the places of incidence and facilitate the transmission. Such factors can be; demographic, economic, behavioral, social and environmental. It is crucial to understand these factors and to review their impact.[9] According to some current studies; temperature, rainfall and relative humidity as climatic factors and urbanization, population growth, trade and transport as non-climatic risk factors, represent major contributing factors to dengue incidence and transmission.[10] Many of the recent studies put focus on studying the relationship between climatic risk factors and dengue incidence, as it is perceived as essential.[11]

In this work, Poisson regression model was first selected for determination of the relation between dengue incidence and climatic risk factors. However, further analysis showed that due to the fact, that dengue data do not meet one of Poisson's regression prerequisites, another mathematical approach was utilized, the Negative Binomial regression. Geographical area of focus, the Philippines, was selected based on the information about current situation of dengue incidence within the country and also in connection to the availability of data, surveillance systems within the country and its diverse climatic conditions. The raw data obtained for the climatic risk factors, population data and dengue cases were processed with basal statistical methods e.g.; conversion, formula derivation, averaging, interpolation, correlations etc., in order to be able to further obtain statistically significant data and then, results. Each of the dengue cases, population and climatic data sets were first processed and adjusted separately. Then, the data were put together within a mathematical interaction and compared and accounted for. At last, the dengue incidence predictive model was proposed, based on the final assessed climatic variables and dengue incidence rate. Finally, results of the model were reviewed, critically discussed and further, a conclusion was provided.

# 2 NEGLECTED TROPICAL DISEASE

Neglected Tropical Diseases (NTDs) are known worldwide as a group of 17 communicable; infectious diseases. Although there exist more than 40 of the diseases worldwide, the World Health Organization (WHO) prioritized 17 of them as these affect over 1 billion people worldwide, especially those living in poor conditions.[1] According to some resources the diseases have been neglected for several decades. This fact is supported by the connection of those diseases to the developing world and the general disregard towards it. The latter primary focus on HIVS/AIDS, malaria and tuberculosis impacted had an impact as well. Thus, until recently research, investment, pharmaceuticals invention, prevention and control programs appointed by policy makers have been scares.[2]

There are several common features to justify the reason for the diseases to be placed into a group as a whole. The diseases are disabling, disfiguring and stigmatizing and thus, believed to be contributing as one of the causes of poverty. Their predominant occurrence can be determined in hot, tropical areas geographically near the equator. There is a close coherence to remote rural areas, urban outskirts and displaced populations. According to some authors NTDs should be considered as the diseases of the "bottom billion"; the poorest of the world population. The diseases cause acute illnesses, long term disabilities and early deaths. All of the low income countries are believed to be affected by at least five NTDs concurrently. As important contributors to the diseases' incidence and transmission are considered, among others: the access to safe water, sanitation or housing conditions. Such listing suggests the preventability and eradicable potential of the diseases.[2] Among other collective characteristics of the NTDs belong: impact on developing countries in terms of disease burden, quality of life or loss of productivity. The population affected by these diseases has generally no visibility or voice in terms of political influence on administrative or governmental decisions. In contrast with HIV/AIDS, malaria and tuberculosis the NTDs do not spread across vast areas and therefore do not, in general, affect populations of high income countries.[1]

As mentioned above, the diseases cause disabilities and disfigurement leading to social discrimination and stigma, which can highly affect women in terms of marriage prospects, vulnerability to abuse or adverse pregnancy. When compared to the morbidity and mortality numbers of HIV/AIDS, malaria and tuberculosis the numbers for NTDs seem visibly lower, however, late evidence published in peer reviewed medical and scientific journals suggests the severity of such morbidity and mortality numbers for NTDs. Although recent evaluations have convinced the governments, pharmaceutical industry, donors and other agencies to invest in prevention and control of the diseases still, more research is needed in order to develop new diagnostic tools, medicines and complication management.[1]

Out of the 17 mentioned diseases, 9 are caused by microparasites and 8 by macroparasites. Most of the microparasites are believed to have simple life-cycles and tend to replicate within their host. They transmit either directly, through environmental contamination or through intimate contact, including transplacental route. Or they transmit indirectly, through a vector either being or not an intermediate host and through blood transfusions or organ transplants. The infections caused by microparasites range from acute (death or recovery), recurrent (repeated growth and decay of organisms in the host) or inapparent (dormant and difficult to detect) to subclinical (symptomless but detectable).[1]

On the contrary, macroparasites have complex life-cycles involving both intermediate and reservoir hosts, a tendency is seen not to replicate within the definitive human host. Exceptions exist in soil-transmitted helminths which do not require intermediate host. They transmit directly, through ingestion from contaminated environment or through skin penetration. Or indirectly, through ingestion of an infected intermediate host or tissues of a reservoir host, through a vector which serves as an intermediate host. The infections are mostly chronic with rather low mortality rates.[1]

It is even more complex to cure some of the micro and macroparasitic infections as some exploit a zoonotic component. Zoonotic infection is described as such in which a human becomes incorporated into the transmission cycle of a pathogen responsible for disease in wild or domestic animal.[1]

The following diseases are caused by the microparasitic pathogens: Buruli ulcer, Chagas disease, Dengue, Human African trypanosomiasis, Leishmaniases, Leprosy, Rabies, Trachoma and Treponematoses. The diseases caused by the macroparasitic pathogens are: Cysticercosis, Dracunculiasis, Echinococcosis, Foodborne trematodiases, Lymphatic filariasis, Onchocerciasis, Schistosomiasis and Soil-transmitted helminthiases.[12]

Another division of the diseases is based on the causative pathogen they result from: virus, including Dengue/Severe dengue and Rabies, protozoa, including Chagas disease, Human African trypanosomiasis and Leishmaniases, helminith, including Cysticercosis, Dracunculiasis, Echinococcosis, Foodborne trematodiases, Lymphatic filariasis, Onchocerciasis, Schistosomiasis and Soil-transmitted helminthiases and bacteria, including Buruli ulcer, Leprosy, Trachoma and Treponematoses.[13]

Tab. 1 comprises the NTDs and their stratification into groups according to the causative pathogen and further lists the causative agents of each disease. It continues by listing global distribution, transmission and generally used control of the diseases. The last two columns provide data for approximate prevalence and people at risk for each disease.

Tab. 1: NTDs, causative pathogen stratification, causative agents, global distribution, transmission/control, prevalence and population at risk.

| **NTDs**[a] | **Causative agent**[1, 13] | **Global Distribution**[1, 13] | **Transmission/ Control**[1, 13] | **Prevalence (million)**[14, 15] | **Population at risk (million)** [14, 15] |
|---|---|---|---|---|---|
| Viral infections | | | | | |
| Dengue | *Aedes aegypti, Ae. albopictus* | the Region of the Americas, the South-East Asia Region and the Western Pacific Region, African and Eastern Mediterranean region | Bites of infected Aedes mosquitoes/ Environmental management and vector control | 50 | Unknown, but increasing numbers at risk |
| Rabies | | Global distribution (especially Asia and Africa) | Bites or scratches from rabid animals (dogs, bats etc.) / controlling rabies in both wild and domestic animals, pre-exposure immunization to humans, post-exposure prophylaxis | Unknown; 0.05 | Unknown |
| Protozoan infections | | | | | |
| Chagas disease | *Trypanosoma cruzi* | Latin America | Contact with the faeces of a triatomine bug/ parasite and vector control | 8-9 | 25 |
| Human African trypanosomiasis | *Trypanosoma brucei gambiense, Trypanosoma brucei rhodesiense* | Africa | Bite of the tsetse fly (*Glossina spp.*)/ control of the animal reservoir, surveillance | 0.3 | 60 |
| Leishmaniases | different species of *Leishmania e.g. Leishmania donovani, L. tropica etc.* | African, Americas, South-East Asia and the Eastern Mediterranean regions | Bites of infected sandflies/ active case-detection, early treatment | 12 | 350 |

| Helminith infections | | | | | |
|---|---|---|---|---|---|
| Cysticercosis | *Taenia solium, Taenia saginata* | Africa, Asia and Latin America | Ingesting the tapeworm's eggs/ surveillance mechanisms, more reliable epidemiological data | 50 | Unknown |
| Dracunculiasis | Dracunculus medinensis | Ethiopia, Ghana, Mali and Sudan | Ingestion of water containing infected Cyclops / community-based surveillance systems, access to safe sources of drinking-water, vector control | 0.01 | Unknown |
| Echinococcosis | *Echinococcus granulosus, E. multilocularis* | Global distribution (pastoral communities), particularly South America, Mediterranean, Eastern Europe, Near and Middle East, East Africa, Central Asia, China, Russia | Ingestion of eggs through direct contact with definitive hosts (dogs) or indirectly through food, water or soil contaminated with eggs/ deworming of dogs, public information campaigns | Unknown | Unknown |
| Foodborne trematodiases (Clonorchiasis, opisthorchiasis, fascioliasis, and paragonimiasis) | *Clonorchis sinensis, Opisthorchis viverrini or O. felineus, Fasciola hepatica or F. gigantica, Paragonimus spp.* | Eastern Asia, Southeast Asia, Africa, Americas | Ingestion of food contaminated with the minute larval stages of the worm (metacercariae)/ Preventive chemotherapy, case-management approach | 20–40 | Unknown |
| Lymphatic filariasis | *Wuchereria bancrofti, Brugia malayi, B. timori* | Africa, Asia, Central and Southern America | Mosquito bites/ mass drug administration programmes | 120 | 1.3 bil. |
| Onchocerciasis | *Onchocerca volvulus* | Africa, small parts in Southern and Central America | Bite of infected black flies/ vector control (aerial application of insecticides), mass treatment | 37 | 90 |
| Schistosomiasis | | | | 207 | 779 |
| Soil-transmitted | *Ascaris* | Global | Environment | Ascariasis: 807 | Ascariasis: 4.2 |

| | | | | | |
|---|---|---|---|---|---|
| helminthiases | *lumbricoides, Trichuris trichiura, Ancylostoma duodenale, Necator americanus* | distribution | (water, soil)contaminated by the worms' infective stages/ deworming, vaccination, access to safe water and proper sanitation | Trichuriasis: 604 Hookworm: 576 | bil. Trichuriasis: 3.2 bil. Hookworm: 3.2 bil. |
| Bacterial infections | | | | | |
| Buruli ulcer | *Mycobacterium ulcerans* | Australia, Guyana, Malaysia, Mexico, Papua New Guinea, Peru, Sri Lanka, West and Central Africa | Mode of transmission is unknown/ Early detection and treatment, vaccine development | 0.05 | Unknown |
| Leprosy | *Mycobacterium leprae* | tropical and sub-tropical regions | Mode of transmission is unknown/ early diagnosis, multidrug therapy | 0.4 | Unknown |
| Trachoma | *Chlamydia trachomatis* | Africa, Asia, Central and South America, Australia and the Middle-East | Living in close proximity to an infected person/ lid surgery, drug treatment, environmental improvement | 84 | 590 |
| Treponematoses (bejel, pinta, yaws) | *Treponema pallidum* | Global distribution (highest rates in tropical regions) | Poor personal hygiene and overcrowding / identification, case finding, treatment, surveillance | Unknown | Unknown |

## 2.1 The WHO and NTDs

As it is described in the first WHO report on neglected tropical diseases, WHO strongly believes that it has never overlooked or neglected the NTDs, as already in 1952 during the Fifth World Assembly in Geneva, it addressed the need of countries for technical assistance to deal with treponematoses, rabies, leprosy, trachoma, hookworm, schistosomiasis and both forms of filariasis. In 2003 WHO shifted the focus of elimination and control from specific diseases towards the health needs of poor communities.[16] It has also shown its belief that many of the 17 diseases can be either effectively controlled or eliminated or even eradicated.[17] Since 2004 the effort and recognition of the NTDs' importance has emerged. WHO has produced many documents which remain a baseline for elimination and control programs. Objectives for control, elimination and eradication of the diseases have been published in many of the World Health Assembly resolutions.[18]

The WHO developed two important strategic interventions; preventive chemotherapy and intensified disease management. However, some of the 17 diseases require specific intervention approaches e.g., dengue, dracunculiasis and human dog-mediated rabies. Still, other interventions are needed to support the above mentioned.[17] Thus, the WHO developed and recommended five public-health strategies for the prevention and control of NTDs: preventive chemotherapy; intensified case-management; vector control; the provision of safe water, sanitation and hygiene; and veterinary public health. According to the evidence more effective control results are obtained when all five strategies are combined and implemented.[1]

The main goal of preventive chemotherapy strategy is to control morbidity in populations at risk of infection and illness. It uses large-scale distribution of high quality, safe, single-dose medicines currently used for four helminthiases: lymphatic filariasis, onchocerciasis, schistosomiasis and soil-transmitted helminthiases. The results of this measure depend on the use and mass distribution of seven broad-spectrum anthelminthic medicines: albendazole, diethylcarbamazine, ivermectin, levamisole, mebendazole, praziquantel and pyrantel. The medicines show efficacy, safety profile, and minimal side-effects and are easy to administer.[1]

The aim of the intensified disease management strategy is to reduce morbidity, prevent mortality and interrupt transmission. The key aspects used to fulfill this aim are: early diagnosis, provision of specialized care and treatment, and management of complications. This strategy substitutes in diseases for which there are no preventive chemotherapy medicines. The focus is on the following diseases: Buruli ulcer, Chagas disease, human African trypanosomiasis, leishmaniasis, leprosy and yaws. Most of the NTDs involve vector transmission e.g. insects, snails, crustaceans transmit infectious agents. Therefore, it is essential to understand the vector biology in order to be able to explain and predict the epidemiology of vector-borne diseases.[1]

A crucial vector-borne disease control measure is the use of pesticides, thus, it is essential to ensure the efficiency, ecological soundness and sustainability of such measures.[17] There are 9 NTDs related to water and sanitation in terms of their development and transmission, thus, it is important to focus on the safe water, sanitation and hygiene strategy. As there still are 900 million people with no access to safe-drinking water and 2,500 million people who lack access to improved sanitation. It is essential to improve the situation otherwise both MDG 7 "Ensure environmental sustainability" and the elimination and eradication of NTDs will not be met.[1]

Several of the NTDs are caused by agents originating from or involving vertebrate animals in their life-cycle. Those are called neglected zoonotic diseases (zoonoses) e.g. brucellosis cysticercosis, echinococcosis, foodborne trematodiasis, human African trypanosomiasis, leishmaniasis and rabies. It is essential to understand the veterinary sciences and through an integrated human and animal health approach improve the prevention and control of zoonoses.[17]

In 2007 the first Global Partners' Meeting took place and as a result there was a rise in shared commitment to support WHO strategies and targets. It triggered scaling up of control and elimination programs, enhanced access to medicines for the poorest. The first WHO report on neglected tropical diseases was published in 2010 describing both progress and challenges which occurred since 2007. The WHO Strategic and Technical Advisory Group for Neglected Tropical Diseases and partners adopted a roadmap for control, elimination and eradication in 2011. The main goal of the roadmap was to set targets for NTDs for 2015 and 2020. The targets for neglected zoonotic diseases were published in a separate report of the Interagency Meeting on Planning the Prevention and Control of Neglected Zoonotic Diseases (NZDs) also in 2011.[17]

The WHO effort has brought over the recent years several improvements e.g.; increased advocacy for new approaches to the control, elimination and eradication of NTDs, commitment from pharmaceutical companies to supply drugs, renewed government commitment and especially, the acknowledgement that the diseases ought to be addressed as part of the Millennium Development Goals (MDGs) agenda. Where they are considered within MDG 6 "Combat HIV/AIDS, malaria and other diseases", as "other diseases".[18] It is perceived that relieve in burden from suffering the NTDs will also bring an immense contribution to the achievement of the MDGs. By solving or improving the issues related to NTDs, does not only the NTD community help in achieving the MDG 6. It will impact other MDG outcomes e.g. by treating school-aged children for schistosomiasis and soil-transmitted helminthiases, helps to improve their nutritional and educational status (MDGs 3, 4, 5 and 6).[1]

As discussed earlier, there has been a shift towards universal health coverage (UHC) instead of aiming the focus solely on disease-specific goals. The term universal health coverage has been defined by WHO as: "ensuring that all people can use the promotive, preventive, curative, rehabilitative and palliative health services they need, of sufficient quality to be

effective, while also ensuring that the use of these services does not expose the user to financial hardship".[19]

In connection to the UHC the Sustainable Development Goals (SDGs) also recognize the need to tackle inequity and provide health for all. The NTDs are included in SDG 3 "Ensure healthy lives and promote well-being for all at all ages" within which the target is to "end the epidemics of AIDS, tuberculosis, malaria, and neglected tropical diseases and combat hepatitis, water-borne diseases, and other communicable diseases" by the year 2030. The inclusion of NTDs was crucial not only because of their influence at country level, but especially due to their impact on the distribution within populations e.g. across socioeconomic groups. There are other SDG targets e.g. target 3.8. or 13.3. which although indirectly, still provide framework for investment in universal coverage against NTDs and simultaneously, for long-term investment in vector control.[19]

After attaining a growth in the global community recognition of the severity of morbidity and mortality resulting from NTDs. And a change of thinking and approach towards prevention and control. The focus has also turned to strengthening health systems and programs in countries where the NTDs hit the hardest in terms of health and productivity.[1] In capacity building WHO is responsible for formulating appropriate training and strengthening existing capacity. All this in order to respond more effectively to the integrated delivery of control strategies. Three tools are used: monitoring, evaluation and surveillance to verify and improve the quality of interventions and to determine whether a specific program delivered expected outcomes.[17]

## 2.2 NTDs in Southeast Asia

First, it shall be explained the diversity of perspective according to which countries belong to the list of Southeast Asia (SEA) countries. Because further in this chapter it might be referred to some countries as being part of distinct regions according to the source used. The WHO distincts the South-East Asia Region (SEAR) of the World Health Organization into a list of 11 countries: Bangladesh, Bhutan, Democratic People's Republic of Korea, India, Indonesia, Maldives, Myanmar, Nepal, Sri Lanka, Thailand, Timor-Leste.[20]

Another division is closely linked to the Association of Southeast Asian Nations (ASEAN) which constitutes of 10 member states: Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic (Lao PDR), Malaysia, Myanmar, the Philippines, Singapore, Thailand, Viet Nam.[3]

Other sources divide SEA into 11 distinct countries: Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic (Lao PDR), Malaysia, Myanmar, the Philippines, Singapore, Thailand, Timor-Leste and Viet Nam.[21]

It ought to be mentioned that out of the 37 countries listed as countries of WHO Western Pacific Region some would be, according to above defined distribution, perceived as countries of SEA: Brunei Darussalam, Cambodia, Lao PDR, Malaysia, the Philippines, Democratic People's Republic of Korea, Singapore and Viet Nam.[22]

Despite of several ASEAN member countries being economic powers, there still is a major hidden burden of poverty and NTDs within the region.[3] Concurrently, NTDs represent an immense public health issue within the SEAR of WHO.[20] Many of the near 200 million people living within ASEAN region are affected by at least one NTD.[3] Within the WHO SEAR the diseases do not only affect vast number of people and cause high morbidity and mortality, they seriously affect the productive and social lives of the people.[20] In ASEAN region roughly 30% of the population live in extreme poverty, the highest share, three-quarters of the poor, is divided among Indonesia and the Philippines. As it was already explained in previous chapters, NTDs do not only affect those living in extreme poverty, but too promote poverty into larger depth, as they are known to be chronic and debilitating.[3]

There are 13 major NTDs affecting ASEAN countries, these are selected from the WHO's list of 17 NTDs. A stratification according to the causative pathogen and the list of the diseases is shown in Tab. 2. There are other simultaneous major NTDs affecting the ASEAN countries, however, these do not correspond with the WHO's list of 17 NTDs.

Tab. 2: List of major NTDs present in ASEAN countries stratified according to their causative pathogen.[3]

| **Helminths** | | **Bacteria** | |
|---|---|---|---|
| Cysticercosis | Lymphatic filariasis | Buruli ulcer | Trachoma |
| Echinococcosis | Schistosomiasis | Leprosy | Yaws |
| Foodborne trematodiases | Soil-transmitted helminthiases | | |
| **Protozoa** | | **Virus** | |
| Leishmaniasis | | Dengue/Severe dengue | Rabies |

The overview of the neglected helminth infections follows: it has been determined that there are 126.7 million people in Southeast Asia infected with *Ascarsis* roundworms, there are 115.3 million people infected with *Trichuris* whipworms, and 77.0 million infected with hookworms. High proportion is accounted for by the *Ancylostoma ceylanicum* a unique zoonotic hookworm infection found especially in Malaysia, Thailand, Cambodia and Lao PDR. Therefore it is possible to determine that around one-half of those people living in poverty in SEA, have one or more soil-transmitted helminth infection. Within the region, hookworm infections have been known for being the leading cause of anemia and iron deficiency anemia which leads to low-birth-weights, inadequate growth and mental development of children, high maternal mortality and low productivity in adults.[3]

According to the WHO database ASEAN countries account for more than 13% of global population eligible for deworming for soil-transmitted helminth infections. Simultaneously, ASEAN countries account for more than 13% of global population requiring mass treatment for Lymphatic filariasis.[3]

Out of the protozoan infections none are considered as highly endemic to Southeast Asia. Nevertheless, Visceral Leishmaniasis has recently emerged in Thailand.[3]

Leprosy and Trachoma are considered as the major bacterial infections present in SEA. Trachoma is endemic solely in Cambodia and Lao PDR, in Viet Nam the surveillance for elimination is currently applied. It is estimated that ASEAN countries account for 10% of the world's registered leprosy cases, with three-quarters in Indonesia. Yaws remains endemic in parts of Indonesia, in general, Buruli ulcer and Yaws can be found in SEA even though not to such an extent as in other world regions.[3]

Viral infections in SEA are represented by two major viral NTDs: dengue and rabies. Dengue infection within ASEAN countries accounts for more than 17% of global disease burden. An important public health threat is represented by canine rabies, especially, in the poorest areas of Indonesia and the Philippines.[3]

There is a summary showing number of cases for some of the major NTDs within ASEAN countries in Tab. 3. Not all listed diseases are provided with number of cases due to the unavailability of data both locally and regionally. This is also the reason for some of the other major NTDs present in ASEAN countries not being listed at all. Tab. 3 also provides the information on Disability-adjusted life years (DALYs) for SEAR. The numbers presented as DALYs are per 100 000 population.

The overall burden of any disease can be assessed by DALYs, it is a time-base measure combining years of life lost due to premature mortality (YLLs) and years of life lost due to time lived in states of less than full health, or years of healthy life lost due to disability (YLDs). One DALY is represented as one lost year of 'healthy' life. The burden which is measured by using DALYs is: the gap between a population's health status and that of a normative reference population. By using DALYs it is possible to compare the burden of diseases that cause premature death but little disability and those that do not cause death but do cause disability. In other expression, DALYs for a given disease or condition are the sum of (YLLs) and (YLDs) due to prevalent cases of the disease or health condition, in a population.[23]

Tab. 3: Number of cases and DALYs for SEAR in selected NTDs.

| **Disease** | **Number of cases ASEAN [millions][3]** | **DALYs SEAR 2012 (per 100 000 population)[24]** |
|---|---|---|
| Lymphatic filariasis | N/A | 94.8 |
| Schistosomiasis | 1.0 | 0.0 |
| Soil-transmitted helminthiasis | Ascariasis: 126.7<br>Trichuriasis: 115.3<br>Hookworm infection: 77.0 | Ascariasis: 31.0<br>Trichuriasis: 7.3<br>Hookworm infection: 49.8 |
| Leishmaniases | N/A | 92.2 |
| Leprosy | 0.02 | 10.8 |
| Trachoma | N/A | 0.4 |
| Dengue | 68.2 | 35.3 |
| Rabies | N/A | 34.1 |

N/A - not available.

There is a need for active surveillance for NTDs and for defining the extent of NTD-caused illnesses within countries and the region as a whole. Besides other reasons, because of the incomplete information on burdens only being available. A need for new drugs, diagnostics and vaccines remains a goal.[3]

There has been success within the region in establishing some of the needed support mechanisms for development. As for example, in 2009 the ASEAN Network for Drugs, Diagnostics, Vaccines, and Traditional Medicines Innovation (ASEAN-NDI) was founded. The network is supported by the World Health Organization Special Programme for Research and Training in Tropical Diseases (WHO-TDR). Its activities include the assessment of the product research and development (R&D) landscape for the triple burden of disease in the region; one of the burden being infectious tropical diseases.[3]

Another Asia Pacific NTD Initiative supported by the WHO Regional Strategic Plan for Integrated NTD Control in the South-East Asia Region (2012–2016) and the WHO Regional Action Plan for Neglected Tropical Diseases in the Western Pacific (2012–2016) was established. The initiative serves as a framework for countries and areas, donors, research institutes and other partners. It supports the following activities; programme planning and capacity-building, health education, mass drug distribution, curative care and morbidity management, monitoring and evaluation and surveillance, and knowledge management and operational research. The supported countries are those included in the WHO's South-East Asia and Western Pacific regions.[25]

# 3 DENGUE

During the 19th century, dengue was perceived as a sporadic disease which caused epidemics only at long term intervals. However, changes took place and currently, dengue is considered as the most important mosquito borne viral disease in the world. Within the past 50 years its incidence has increased by 30-fold with outbreaks which occur in five of six WHO regions. Currently, dengue is present in 112 countries in Southeast Asia, the Pacific, the Americas and Africa. It is found mainly in tropical and sub-tropical regions worldwide, especially in urban and semi-urban areas. There are around 2.5 to 3 billion people estimated to be at risk of dengue.[4] There are 50–100 million new infections estimated to occur every year. Severe dengue was first recognized in the 1950s during dengue epidemics in the Philippines and Thailand. It occurs in most Asian countries and has become a leading cause of hospitalization and death.[26]

There are three forms of the dengue infection; dengue fever (DF) also known as break bone fever, in severe dengue there are: dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS).[5] DF is characterized as asymptomatic or self-limiting. In severe dengue, DHF is characterized by plasma leakage in severity grades 1 and 2 and it can lead to DSS a life-threatening syndrome with grades 3 and 4. However, the pathogenesis of DSS has not been yet fully understood. The severe dengue has also been defined by severe bleeding and organ impairment.[26]

Dengue is a systemic viral infection which can be found globally both in endemic and epidemic transmission cycles.[27] The dengue virus is a single stranded Ribonucleic acid (RNA) virus which belongs to the Flaviviridae family. There exist four serotypes of dengue (DEN 1-4), those are classified according to biological and immunological criteria. In general, a person infected by one serotype develops protective immunity against that specific serotype. However, it does not develop immunity against the remaining serotypes. Furthermore, when a person is infected again with another serotype a more severe infection may occur.[4] Based on a pathogenesis from a different source the term dengue viruses (DENV 1-4) can be also found. All four viruses evolved in non-human primates and each entered the urban cycle independently around 500–1,000 years ago.[27]

Despite differences among each DENV the serotypes cause nearly the same syndromes in humans and circulate in the same biological area. The syndromes produced are often conditioned by age and immunological status. Thus, for an example, in initial dengue infection children will most likely experience subclinical infection or mild undifferentiated febrile syndromes. On the other hand children with secondary dengue infection will experience a dramatic change of the pathophysiology of the infection e.g. particularly sequential infections,

meaning differential serotype order occurrences. This can lead to DSS, its severity is age-dependent.[27]

On the contrary, in adults, primary infections with each of the four DENV serotypes, often result in DF. Some outbreaks of primary infections have been predominantly subclinical. Still, infections in adults bring the tendency for bleeding which leads to DHF. Secondary dengue infections in adults can produce both DSS and DHF. As mentioned above, the immunological status can affect the infection process e.g.; in individuals with asthma, diabetes and other chronic diseases the infection can be life-threatening. Some host factors can also affect the increase of risk of severe dengue e.g. female sex or AB blood group etc., or decrease the risk e.g. race or degree of malnutrition etc.[27]

There are still no vaccines or drugs available for treating dengue, however it can be managed by careful monitoring of the warning signs and early initiation of aggressive intravenous rehydration therapy.[28] The efforts to diminish dengue transmission are focused on vector control, by using combination of chemical and biological targeting of the vector mosquitoes and management of their breeding sites.[27]

The vector of the dengue infection are mosquitoes from the aedes genus. The following species: *Aedes aegypti, Aedes albopictus,* and *Aedes polynesiensis* play an important part in transmission of dengue. The primary and most important vector of the three is the *Aedes aegypti*, the other two may act as vectors depending on the geographical location. For example, *Aedes albopictus* can be found as a vector in Thailand, Samui Island, India, Singapore, and Mexico. *Aedes aegypti* is found in tropical and subtropical areas, it bites during the day and breeds within containers. It rests indoors, within a house, it can be found especially in living rooms and bedrooms. Thus, it is rather complicated to control this vector, as it is out of the range of outside insecticides and it maximizes the man-vector contact. It breeds within polluted water or small containers such as flower vase or buckets and the eggs survive for long time periods.[4]

High mosquito densities in endemic areas can be caused among other causes by: improper disposal of garbage or inadequate wastewater drainage. Elevated mosquito larval populations can be found during rainy season. This is also a reason why epidemics of dengue coincide with the rainy season. The viral propagation in mosquitoes is affected by ambient temperature and humidity. The mosquitoes are not able survive cold winter, thus the temperature should not decline below 10 °C. After biting an infected human, the dengue virus enters an adult female mosquito and it then, further transmits the virus to another human. As the infected mosquitoes take longer to finish their blood meal, it is perceived as a contributing factor to the efficiency of this particular mosquito as a vector.[4]

## 3.1 Vector life cycle and dengue virus transmission cycle

The life cycle of *Aedes aegypti* involves 4 life stages: adult, egg, larva, pupa. The entire life cycle lasts around 10–12 days. An adult, female mosquito lays the eggs around inner, wet walls of containers with water, above the water line. The eggs stick to the wall and can survive up to 8 months of drying out. Mosquitos generally lay around 100 eggs at a time. When the water level rises (often after a rainfall) and covers the eggs, the larvae emerge. Larvae feeds on microorganisms in the water. There are 4 larval stages during which the larvae molts and sheds the skin three times, then larva becomes pupa. The pupae will develop, usually for 2 days, until the body of a new mosquito emerges from the pupal skin and leaves the water. The life cycle of the primary dengue vector *Aedes aegypti* mosquito is depicted in Fig. 1 a) below.[29]

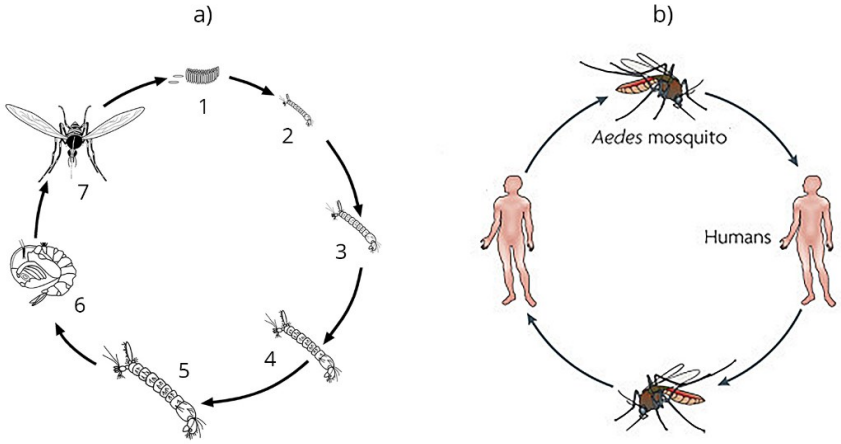


Fig. 1: a) Aedes aegypti mosquito life cycle; eggs (1), 1st larval stage (2), 2nd larval stage (3), 3rd larval stage (4), 4th larval stage (5), pupa (6), new mosquito (7). Source:[30]. b) The transmission human-to-mosquito-to-human life cycle of dengue virus. Source:[31].

Fig. 1 b) shows the human-to-vector (mosquito)-to-human dengue transmission cycle. An infected human is bitten by a mosquito, which then is infected, it further infects another human with the following blood feeding, whom then may infect another mosquito. The symptoms of the infection usually present 4–7 days after the mosquito bite, they typically last for about 3-10 days. Concurrently, a person will develop within 4 days from the bite a so called viremia. It lasts approximately 5 days, during this time there are high levels of the dengue virus in the blood. In order for the transmission to occur a mosquito must feed on a person during this 5 day period. After the virus enters the vector mosquito, it takes about 8–12 days of incubation for the virus to develop and then, it can be further transmitted.[31]

## 3.2 Dengue in South East Asia

Dengue is the most rapidly spreading arboviral disease in the tropics and subtropics.[32] It presents as a health burden in many countries of the world, but especially in South East Asia countries: Thailand, the Philippines, Indonesia, Singapore, Malaysia, Myanmar, Lao People's

Democratic Republic, Brunei Darussalam, Cambodia and Vietnam. Concurrently, it is recognized as a major health problem in the WHO Western Pacific Region countries, especially in Cambodia, Malaysia, the Philippines and Vietnam. These countries make up over 90% of the total cases reported in the region. In Vietnam it is a leading cause of hospitalization. In the Philippines, dengue is currently the most significant vector-borne disease. Although, the amount of dengue cases differs in various areas of the country. For an example, there were 36 cases per 100,000 population in 2010 in southern Autonomous Region in Muslim Mindanao and simultaneously, 372 cases were reported in northern Cordilera Admistrative Region. In addition, overall number of reported cases increased with the onset of rains.[6]

According to [27] there were 96 million estimated apparent dengue infections globally in 2010. Apparent dengue infection would be characterized as an infection with visible clinical symptoms e.g. rash, nausea etc. Thus, affecting the daily routines of an individual. Asia alone accounted for 70% of this burden. Asia is also typically perceived as disproportionate in distribution of the disease within its countries. Additionally, there were 294 million inapparent infections estimated worldwide in 2010. An inapparent infection would be one with no clinical symptoms visible and thus, has no impact on the daily life of an individual. This type of infection is not detected by the public health surveillance systems. However, it represents an immense potential reservoir of the infection.

Fig. 4 represents the annual number of infections for all ages as a proportion of national geographical area. It also depicts the above mentioned disproportionate distribution of the disease in Asia. The colored scale ranges from 0 infections in dark green color to 7.5–2.5 million infections in dark red color. As it is visible the highest amount of infections within South East Asia countries is present in Indonesia, the Philippines, Vietnam and Thailand.[27]

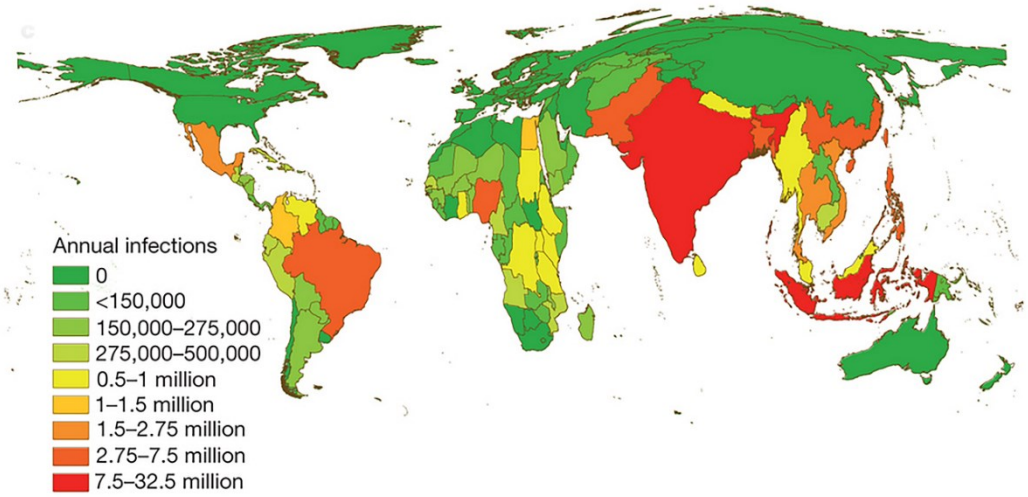


Fig. 2: Annual number of dengue infections for all ages as a proportion of national geographical area. Source:[27].

# 4 EPIDEMIOLOGY FUNDAMENTALS

According to Schoenbach & Rosamond[7] epidemiology is: "both a field of research to advance scientific understanding and also of application of knowledge to control disease and advance public health, a (primarily observational) science and a public health profession".

The key aspects of epidemiology are the following: it deals with populations and thus involves rates and proportions, averages, dynamics etc. It involves measurements thus, the need for definition of the phenomena, spectrum of disease, source of data and compromise. It mostly involves comparison and thus, considers standards of reference for baseline risk, equivalent measurement accuracy and adjustment for differences. Finally, it is multidisciplinary as it must consider: statistics, biology, chemistry, physics, demography, geography, environmental science, policy analysis etc. It involves interpretation e.g. consistency, mechanisms e.g. environmental or economic and policy e.g. implications or recommendations.[7]

There exist two major sources of data for epidemiology: first, aggregate data, these are data obtained from some of the following sources: vital statistics (birth rates, death rates, etc.), data from the Census and other government data-gathering activities (demographic, housing, etc.), summaries of disease and injury reporting systems and registries, workplace monitoring systems, environmental monitoring systems (e.g. air pollution measurements) and production and sales data. Second type of data source is the individual-level data, vital events registration (births, deaths, etc.), disease and injury reporting systems and registries, national surveys, computer data files (e.g. health insurers), medical records, questionnaires, biological specimens.[7]

The data quality in epidemiology is essential, however, it can still become a challenge to ensure it. This is due to many possible sources of error, especially, in observational studies of human populations. It became a major topic in epidemiologic methods to identify, avoid and control the potential sources of errors.[7]

The best quantifiable and understood error yet is believed to be *sampling error*, it is a distortion that can occur from the "luck of the draw" in small samples from a population.

Perceived as more problematic is the *selection bias*, it is an error that occurs when the study participants are not representative of the population of interest. This error can result from some of the following: self-selection (volunteering), nonresponse (refusal), loss to follow-up (attrition, migration), selective survival, health care utilization patterns, systematic errors in detection and diagnosis of health conditions, choice of an inappropriate comparison group (investigator selection).[7]

Another error determined as highly problematic is *information bias*, it is a systematic error due to incorrect definition, measurement, or classification of variables of interest. Sources of such error can be; recall or reporting bias, false positives or negatives on diagnostic tests, errors in assignment of cause of death, errors and omissions in medical records. Another error relevant for epidemiologists is *confounding*, an error in the interpretation of comparisons between groups that are not truly comparable.[7]

## 4.1 Essential terms in epidemiology

Data collected routinely are usually not adequate for most epidemiologic studies and so data must be collected specifically for the purpose of a given study. Although it might show as time, effort and expense consuming, it provides estimates of measures that are more suitable for the acquired purposes. The following measures represent such examples: prevalence, incidence and case fatality.[7]

**Prevalence**

Prevalence is the most of basic of epidemiologic measures. It measures the proportion of cases within a population. It is defined as the number of cases divided by the population at risk[7]:

$$Prevalence = \frac{Cases}{Population\ at\ risk}. \quad (1)$$

It ought to be noted that prevalence is a proportion, so it must lie between 0 and 1, inclusive. Population at risk (PAR) means "eligible to have the condition". Prevalence can be used to estimate the probability that a person selected at random from the PAR has the disease Pr(D). Prevalence is defined to have three components: existing cases, populations "at risk" to have the condition and point (or a period) in time to which the prevalence applies.[7]

**Incidence**

Incidence in connection is the measure of the occurrence of new cases. It is defined as the number of new cases divided by the population at risk over time[7]:

$$Incidence = \frac{New\ cases}{Population\ at\ risk\ over\ time}. \quad (2)$$

Incidence includes three components: new cases, population at risk and interval of time. Note that incidence involves the passage of time. Incidence may be expressed as a proportion or as a rate. Incidence can be used to estimate the risk of an event during a stated period of time.[7]

There exist two major types of incidence measures. They differ primarily in the method of constructing the denominator: *cumulative incidence* (CI) also known as *incidence proportion* (IP) and *incidence density* (ID) also known as *incidence rate* (IR).

**Cumulative incidence**

CI represents the proportion of a population that experience an event or develop a condition during a stated period of time:

$$CI = \frac{New\ cases\ during\ stated\ period}{Number\ of\ persons\ at\ risk}. \quad (3)$$

Furthermore, the definition of CI is based on an "ideal" scenario; a population known to be free of the outcome is identified at a point in time (a cohort). All members of the cohort are at risk of experiencing the event or outcome (at least once) for the entire period of time, all first events or outcomes for each person are detected.[7]

It should be also noted that the period of time must be stated (e.g., "5-year CI") or be clear from the context (e.g., acute illness following exposure to contaminated food source). Because CI is a proportion, each person can be counted as a case only once, even if she or he experiences more than one event. As a proportion, CI can range only between 0 and 1 and thus, can be used to estimate risk or the probability of an event.[7]

**Incidence density**

ID represents the rate at which new cases develop in a population, relative to the size of that population:

$$ID = \frac{New\ cases\ during\ stated\ period}{Population\ time}, \quad (4)$$

where the term "population time" stands for number of person years of observation e.g. person months.[7]

Note that ID is a relative rate, not a proportion. The units of time must be stated, otherwise the numeric value is ambiguous (e.g., 15 cases/100,000 person-years = 15 cases/1,200,000 person-months). Ideally, incidence density is the instantaneous rate of disease occurrence at each moment in time. In practice, epidemiologists generally compute average ID during one or more periods.[7]

**Risks and odds**

It is essential to define the terms *risk* and *odds* in terms of epidemiology. In general, the term risk means the probability ($p$) that an event will occur in a given stated or implicit time interval. Furthermore, within its epidemiologic use, risk is a conditional probability, because it is the probability of experiencing an event or becoming a case conditional on remaining "at risk" (eligible to become a case) and "in view" (available for the event to be detected).[7]

Concurrently, any probability can be transformed into a related measure, the odds. Odds are defined as the ratio of the probability of an outcome to the probability of another outcome. When the only outcomes are (case, non-case), then the odds are the ratio of the probability of becoming a case, to the probability of not becoming a case. If the risk or probability of becoming a case Pr(D) is $p$, then the odds of becoming a case are $p/(1-p)$. If the risk, or probability, of developing disease X is 0.05 (5%), then the odds of developing disease X are 0.05/0.95 = 0.0526 (the odds always exceed the risk, especially for large risks).[7]

The advantage in using odds is in its mathematical properties. For an example, probabilities are restricted to the 0 to 1 interval. On the contrary, odds can be any nonnegative number

e.g. odds = 1.0 ("fifty-fifty") corresponds to probability = 0.5; the middle of the set of possible values. The logarithm of the odds can therefore be any real number, with log(odds) = 0 corresponding to the middle of the set of possible values.[7]

## 4.2 Epidemiologic models

In general, it is essential to understand statistical methods, especially statistical methods in connection to medicine, to understand what it is, given the type of data that has been collected. However, it is also crucial to understand, that different statistical methods have much in common and thus, it helps by understanding one method, to also understand another. The following methods apply to many types of exposure and outcome variables.[8]

### 4.2.1 Simple and multiple linear regression model

The method of linear regression is used to estimate the best-fitting straight line to describe the relationship between a dependent variable $y$ (for medical data also known as numerical outcome) and an independent variable $x$ (for medical data also known as numerical exposure).[8]

The simple linear regression considers only one independent variable.[33] It also provides an estimate of a correlation coefficient, which measures the closeness (strength) of the linear association.[8] In other words, linear correlation coefficient (e.g. Pearson) usually abbreviated "r", measures the degree to which the association between two variables is linear.[7] The equation of the regression line is:

$$y = \beta_0 + \beta_1 x + e\,, \tag{5}$$

where $\beta_0$ and $\beta_1$ are the parameters or regression coefficients of the linear regression: $\beta_0$ is the intercept (the value of $y$ when $x = 0$), and $\beta_1$ the slope of the line (the increase in $y$ for every unit increase in $x$). The letter $e$ represents the error and is normally distributed with mean zero and standard deviation $\sigma$.[8]

On the contrary, multiple linear regression model considers more than one independent variable. It can be carried out with any number of variables, however, it is recommended for the number to be kept reasonably small, as with larger numbers the interpretation becomes more complex.[8] The general form of the multiple regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i + e\,, \qquad (6)$$

where the quantity, $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i$, is known as *linear predictor* of the outcome *y*, given particular values of the independent variables $x_1$ to $x_i$. The *error*, *e*, is normally distributed and with mean zero and standard deviation $\sigma$, which is estimated by the square root of the residual mean square.[8]

It can be often found that the relationship between the dependent variable and independent variable is non-linear. There are three possible ways of incorporating such independent variable in the multiple regression equation. The first possibility is to redefine the variable into distinct subgroups and include it as a categorical variable using indicator variables, rather than as a numerical variable. The second approach would be to find a suitable transformation for the independent variable. And the third option is to find an algebraic description of the relationship.[8]

### 4.2.2 Logistic regression model

In statistics, logistic regression model is also known as logit regression or logit model.[33] It is most commonly used to analyze *binary* dependent variables. The dependent variable is categorical. The binary dependent variable can take only two values 0 and 1 e.g. healthy or diseased. It also provides a flexible mean of analyzing the association between a binary dependent and a number of independent variables.[8] It can do so by estimating probabilities using a logistic function.[33]
The general form of the logistic regression model is similar to that for multiple regression:

$$log\ odds\ of\ dependent\ variable = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i, \qquad (7)$$

the difference is that a transformation of the dependent variable, namely the *log of the odds of the dependent variable*, is modeled. The $\beta$'s are the regression coefficients associated with the *p* independent variables.[8]

The transformation of the probability, or risk, π of the dependent variable into the log odds is known as the logit function:

$$logit(\pi) = \log\left(\frac{\pi}{1-\pi}\right), \qquad (8)$$

as defined in the above description, probabilities must lie between 0 and 1, odds can take any value between 0 and infinity (∞). The log odds are not constrained at all; they can take any value between −∞ and ∞.[8]

The key epidemiologic assumptions of the logistic model are: the log odds of disease are linearly related to each of the risk factors, or equivalently, the disease odds are exponentially related to each of the risk factors. Or equivalently, the disease risk is related to each of the risk factors by the logistic (sigmoidal) curve and that the joint effects of the risk factors are multiplicative on disease odds.[8]

### 4.2.3 Poisson regression

It is a method used by epidemiologists to control for confounding and to obtain adjusted measures of effects. It is used for the analysis of rates.[8] This is used to estimate rate ratios comparing different independent variable groups.[33] Like logistic regression models, Poisson regression models are fitted on a *log scale*. The results are then antilogged to give rate ratios and confidence intervals. The general form of Poisson regression is similar to the one of logistic regression and multiple regression. It relates the log rate to one or more independent variables:

$$\log(rate) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i. \quad (9)$$

Again, the quantity on the right hand side of the equation is the *linear predictor* of the log rate, given the particular value of the $i$ independent variables $x_1$ to $x_i$. The $\beta$'s are the regression coefficients associated with the $i$ independent variables.[8]

Since log(rate) = log($d/T$) = log($d$) - log($T$), the general form of the Poisson regression model can also be expressed as[8]:

$$\log(d) = \log(T) + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_i x_i, \quad (10)$$

The term log($T$) is known as an *offset* in the regression model. In order to use statistical packages to fit Poisson regression models, it is essential to specify the dependent variable as the number of events and give the independent variable time $T$, which is then included in the offset term, log($T$).[8]

### 4.2.4 Bayesian statistics

Bayesian statistics used to formulate statistical models requires the specification of prior distributions for any unknown parameter. A prior distribution probability of an uncertain quantity is the probability distribution expressing one's beliefs about this quantity before some evidence is taken into account.[34]

Bayesian approach to statistical inference is based on the Bayes' formula for relating conditional probabilities.[8] It describes the probability of an event based on a condition that

might be related to the event. It produces the posterior probability distribution, which is the conditional distribution of the uncertain quantity given the data. In other terminology, the posterior probability is the conditional distribution that is assigned after the relevant evidence has been taken into account.[34]

The following equation represents Bayes' theorem:

$$p(A\ given\ B) = \frac{p(B\ given\ A) \times p(A)}{p(B)}, \tag{11}$$

where, $p$ represents probability and $A$ and $B$ represent events. The $p(A)$ is the prior probability of $A$ before the event $B$ is observed. The $p(B)$ is the probability of $B$ with no regard towards $A$. The $p(A\ given\ B)$ is the posterior probability and it is the probability of event $A$ given that $B$ is true. The $p(B\ given\ A)$ is the probability of event $B$ given that $A$ is true.[34]

In the Bayesian interpretation, probability measures a degree of belief. Bayes' theorem links the degree of belief in a proposition before and after accounting for evidence.[34]

### 4.2.5 Other models

Among other models used in epidemiology and statistics in general are: time-series analysis and spatial models. In terms of spatial models, these are very similar to those used for time-series analysis, only extended with the spatial aspect, thus sampling of an area of interest is incorporated. It uses topological, geometric and geographic prospects. Spatial models are influenced by some of the following: spatial autocorrelation statistics, Geographically Weighted Regression (GWR) or Geographically Integrated Systems (GIS).[35]

The time series analysis comprises methods for analyzing time series data in order to extract statistics and other characteristics of the data. Models for time series data can have different forms and processes, e.g. autoregressive (AR) models, the integrated (I) models and the moving average (MA) models. Combinations of these result in autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. These classes can be also extended to deal with vector-valued data and can be found under the term multivariate time-series models and sometimes extend the previously mentioned acronyms with "V" for vector. Therefore, will be found as vector auto-regression (VAR) models. It is common to include seasonal terms, represented with "S" and the product will be seasonal autoregressive integrated moving average (SARIMA) models.[36]

Compartment models belong to another group of modeling approaches. The compartment is a unit or block, where its characteristics are considered constant. Elementary epidemiological compartment model consist of three blocks – S, I and R. There, S stands for susceptible group of people, I stands for number of infectious and R number of recovered people. An interactions between each compartment or group of people are defined by differential equations.[37]

# 5 ANALYSIS OF RISK FACTORS

Dengue belongs among the major disease burdens in SEA and it has been hyper endemic for decades within the region. The region has the highest dengue incidence and the epidemics occurs in cycles every three to five years.[38] Factors such are: demographic, economic, behavioral and social often provide a key pace in control of communicable diseases, although, these factors have not been completely understood and their impact has not been sufficiently reviewed and confirmed in case of dengue.[9]

Both climate variables e.g. temperature, humidity, precipitation, wind speed or El Niño events and non-climatic trends such as socio-environmental factors e.g. population growth, migration, urbanization, housing, family income, international trade and travel and vector control[10], are expected to increase the prevalence of mosquito breeding sites, mosquito survival, the speed of mosquito reproduction, the speed of viral incubation, the distribution of dengue virus and its vectors, human migration patterns towards urban areas and displacement after natural disasters[39].

The geographic distribution of dengue viruses and their vector emerged in SEA after the World War II. It was, especially, due to the increased transport of equipment and movement of people which brought the spread of the vector mosquito to new geographic areas. Further, an ideal breeding habitat for the mosquito was ensured by inadequate water storage and systems and the presence of abandoned equipment and junk. The leaving troops, on the other hand, served as susceptible hosts for the virus.[40]

With all of the previously mentioned, it is still believed that it was the urbanization of the region that emerged after World War II, which provided the ideal conditions for virus propagation. Especially, the rural-urban migration to the cities which induced the unplanned growth of urban centers, followed by incapacity to provide adequate housing, water supply and sewerage systems. The urban population again served as susceptible host and resulted in an epidemic.[40]

Current studies define temperature, rainfall and relative humidity as important climatic factors in terms of the growth and dispersion of the mosquito vector and potential of dengue outbreaks. Other studies determine the population growth, uncontrolled urbanization, and spread of mosquito vector and movement of virus via international travel, as the major contributing non-climatic factors of recent dengue expansion in endemic areas of SEA region.[10]

Concurrently, there is a lack of more complex studies that would incorporate sophisticated multivariable predictive model with both; climatic and non-climatic data, which may help in controlling and preventing DF and the potential impact of these factors on dengue incidence and transmission.[10]

## 5.1 Climatic factors

Based on diverse sources of evidence the complexity of the relationship between climatic variables and dengue incidence is clear. There is a near unanimous scientific consensus that global temperatures are increasing, annual global rainfall will increase, although, differing in regions, flooding will become more severe and climate variability will increase.[39] The change in global temperature, precipitation and humidity will affect the biology and ecology of vectors and the risk of vector-borne diseases.[10] Thus, changes in climate will alter the spatial and temporal dynamics of DENV ecology.[41]

### 5.1.1 Temperature

It is predicted that the endemic range of DF will expand geographically with increasing temperatures. This will also allow for increased reproduction and activity and decreased incubation time of larvae, resulting in increased offspring productivity. Thus, increase in transmission potential and prevalence of DF will take place. Increased temperatures will most likely increase DF transmission by extending the season in which transmission occurs. Although, dengue incidence is related to temperature, the relationship varies by location. Higher temperature supports virus replication, vector proliferation and feeding frequency and thus, enhances transmission. However, it is important to note, that the impact of risen temperature is not immediate in terms of time. Various studies report differing lag times between increased temperature and an obvious increase in dengue transmission. The range was from 4 to 16 weeks, depending on the region and area observed.[10] On the contrary, increased temperatures in already warm areas may have negative effects on the range of the virus transmission by decreasing vector survival, reproduction and immature habitat. Furthermore, the increasing temperatures will change the latitudinal and elevational extent of the disease.[41]

Temperature will also influence the vector ecology. The ideal range of temperatures for *Ae.aegypti* and its survival, through all development stages, is between 20–30 °C. Evidence shows that mortality of an adult mosquito increases with prolonged extreme heat (over 40 °C) and cold (below 0 °C). The female mosquito's reproductive cycle is also governed by ambient temperature. The fertilization decreases below 20 °C. The temperature has a direct biophysical influence on viral replication and on vector development and survival.[41]

### 5.1.2 Precipitation

Precipitation provides essential habitat for the aquatic stages of the mosquito life cycle. The ideal habitat for the vector usually is; containers common in urban environments. However, in general any man-made container becomes the habitat for the mosquito pupae. Monsoon rains are associated with increased numbers of eggs and adults. Thus, precipitation has a strong influence on vector distributions. Mosquito range has been determined to expand during La Niña conditions (generally wetter). Simultaneously, intense rainfall may wash out breeding sites and therefore, have a negative effect on vector populations. Still, drier conditions e.g.; El Niño conditions, may also indirectly result in expansion of vector's range, as with droughts individuals will increase their water storage.[41] However, the rainfall and its influence on dengue transmission are inconsistent across geographical locations. In many countries of Asia-Pacific region, dengue outbreaks with positive association between wet season and dengue incidence and precipitation were reported.[10]

### 5.1.3 Humidity

Higher rates of precipitation in a combination with higher temperatures result in increased humidity. Higher humidity is associated with increased mosquito feeding activity, survival and egg development.[41] According to some sources it was found that humidity is the most important indicator of DF outbreak globally. It suggested higher incidence of DF by 30% in areas with higher humidity compared to areas with low humidity.[42]

It is important to stress the climate-dengue associations in connection to local climate variations. Variation in rainfall, humidity and temperature patterns among different areas, municipalities or provinces can be a large determinant of the strength and direction of associations between climate variables and DF incidence. Even though such fact is well known, availability of data often forces the researchers to scale up.[41]

## 5.2 Socio-environmental factors

It is believed that other factors contributing to dengue incidence are the increasing trends in population growth, uncontrolled urbanization, spread of mosquito vector and international trade and travel.[10]

### 5.2.1 Housing

The transmission of dengue may be influenced by the socio-economic status of people. People in developed countries have better living conditions, especially in terms of housing, which is an important factor for the dengue incidence. Facilities such as glazed windows, piped water, insect screening or air-conditioning effectively reduce contact with the vector mosquitoes and decrease their survival rate and reduce the risk of transmission. According to some studies, single houses e.g. on plantations, had 3-15 times higher risk of dengue compared to the town

houses and slum houses.[10] Traditional practices e.g. rainwater storage on roofs, expose the individuals to higher risk. The placement of the housing may be also determining, as areas close to markets and open sewers have 1.8 times higher risk of contracting the disease.[9]

### 5.2.2 Urbanization

Although, there has been an evidence of dengue transmission and outbreaks in rural areas in SEA, especially due to increased transport contact and spread of peri-urbanization. Urbanization is still strongly associated with DF cases. However, due to rural-urban migration, and thus, growth of urban population, dengue activity remains at significant level.[40]

Urban and suburban environments in tropical and sub-tropical regions are fragile due to rapid population movement, which causes unorganized urbanization with increasing housing densities and abundance of breeding sites for the mosquitoes.[43]

### 5.2.3 Trade and transport

International trade and transport are suggested to have an impact on geographical distribution of vectors and pathogen. Commercial shipping might be linked to spread of *A.aegypti* between regions. Air-travel has increased the dissemination of dengue viruses via rapid transit of infected individuals around the world. Thus the movement of the virus has been facilitated through, especially, air travel.[10]

### 5.2.4 Age, race and sex differences

DF is in general acknowledged to be a childhood disease and is often stated to be an important cause of pediatric hospitalization in SEA. Lately, there has been increasing evidence of the disease shifting towards older age groups. Thus, the epidemiological change in dengue infection within SEA region has been visible. Several studies showed that, especially in DHF and DSS cases, the disease occurred in patients older 15 years.[9]

Racial predisposition seems to be also one of factors which ought to be taken into account when dealing with human susceptibility to dengue infection. The major observation is that black individuals have a reduced risk for dengue fever compared to white or Hispanic individuals. [43] It is believed the reason for this predisposition to be a dengue-resistant genotype in black population. Further research for better understanding of dengue pathogenesis in connection to racial predisposition is suggested.[9]

It is essential to understand the male-female differences in infection rates and severity of disease for public health control programs. According to some studies the number of male compared to female patients was twice as big, sometimes even higher. Other sources show, despite higher incidence in males, severe illness to be higher among women. It is still unclear what the reason is for the male gender to be more susceptible. According to some, in many

Asian and other traditional societies, it is not ordinarily for the women to seek clinical care, instead they seek traditional practitioners who do not report to public surveillance systems.[9]

## 5.3 Current methods of dengue risk factors analysis

Diverse statistical analytical methods have been used for determination of the relationship between climate variables and dengue, e.g. cross correlations, Poisson, logistic or multivariable regression, SARIMA-time series etc. Many of these successfully established climate and dengue relationship and predictive models of dengue. Minimum, maximum and mean temperatures, relative humidity and rainfall were the most important climate variables for dengue prediction. Although, these variables are predictive at specific lags of time.[11]

Concurrently, problematic arose to be the data availability, thus, often use of aggregated data, over large spatial scales and long time periods was established. The issue with such data might be the inability to describe the influences happening over short time periods (daily, weekly) and at smaller spatial area (country level). Simultaneously, the data reported for dengue cases or incidence may be under or over reported, which might change over time and geographical area. Reported dengue cases may be also influenced by case definitions, availability of public health systems and subclinical cases documentation. Thus, it is advised to consider all the mentioned factors, before identifying the relationship between climate and dengue disease and obtaining dengue prediction.[11]

In endemic settings, dengue transmission is characterized by non-linear dynamics, strong seasonality, multi-annual oscillations and non-stationary temporal variations. Seasonal and multi-annual cycles in dengue incidence vary over time and space. According to the evidence, the inter-annual and seasonal climate variability have a direct influence on the transmission of dengue. Further suggesting the importance of climatic variables in dengue transmission.[11]

Another highlighted importance is the lag time, at monthly scales. The delayed effect (time lag) of climatic variables on dengue incidence may be explained by the climatic factors which affect the incidence indirectly. Thus, they affect the life-cycle dynamics of both vector and virus. The lag between climate data and incidence differs. The lag is expected to be shorter for minimum temperature and longer for higher temperatures and high relative humidity. Concurrently, the mean temperature is involved in all biological cycles of the mosquito that take more time to influence the incidence.[11]

In general, dengue fever disproportionately affects the poor. In order to be able to protect such vulnerable populations, it is essential to understand the effect of climatic and non-climatic trends on the relationship between vector, pathogen and host that drive the spread of DF.[39]

Many studies focusing on the relationship between climatic variables and dengue incidence selected for their dependent variables either dengue incidence or mosquito density and for their independent, climate variables; temperature (maximum, minimum, mean), amount of

precipitation (rainfall) and relative humidity, sometimes even El Niño Southern Oscillation (ENSO).[11] Those studies which also incorporated socio-environmental factors for their modeling selected some of the following variables: age, gender, occupation, window screen, water tank presence, housing pattern, mosquito control etc.[10]

Tab. 4 shows some of the published studies on either climatic or socio-environmental, or both, factors and their effect or relationship with dengue and its incidence. The table defines specific studies, the location of each study and the study time period, further, it states the used statistical model and variables used for modeling.

Tab. 4: Studies identifying relationships and impact of climatic, socio-environmental variables on dengue.

| **Study/ Source** | **Study area (Period)** | **Method** | **Risk factors** |
|---|---|---|---|
| **Arcari et al.**[44] | Indonesia (1992-2001) | Multiple regression, Pearson correlation | Temperature, rainfall, relative humidity |
| **Tipayamongkholgul et al.**[45] | Thailand (1996-2005) | Poisson regression | Temperature, relative humidity, ENSO |
| **Wu et al.**[46] | Taiwan (1998-2002) | Logistic regression, Spatial analysis | Temperature, rainfall, level of urbanization, % of elder population |
| **Hii et al.**[47] | Singapore (2000-2007) | Poisson regression, time-series analysis | Mean temperature, precipitation |
| **Cummings et al.**[48] | Thailand (1980-2005) | Linear regression, Wavelet analysis | Rainfall, population data (age, birth; death rate, household size etc.) |
| **Wu et al.**[49] | Taiwan (1998-2003) | Time-series analysis, ARIMA | Temperature, rainfall, relative humidity, vector density |
| **Yu et al.**[50] | Taiwan | Spatiotemporal analysis, Bayesian analysis | Climate variables (temperature, rainfall etc.) |
| **Xuan et al.**[51] | Vietnam (2008-2012) | Poisson regression | Temperature, rainfall, relative humidity |
| **Pham et al.**[52] | Malaysia (2001-2010) | Linear regression model | Mean temperature, rainfall |

# 6 METHODOLOGY

In order to obtain a sufficient background and knowledge about the problematic of neglected tropical diseases and their presence in Southeast Asia, with particular focus on dengue fever, a literature research was conducted. Concurrently, the aim was to carry out detailed analysis and to review the most current statistical methods used for modeling of dengue incidence and transmission worldwide, with a specific respect to SEA. The focus was directed towards interconnection with the major climatic and non-climatic factors affecting predictive modeling.

The research incorporated the use of available electronic databases e.g. PubMed, Scopus, Science Direct, Web of Science, Google scholar and WHO library to obtain the necessary information. The keywords used within the search were those in conjunction to all covered topics e.g. NTDs, WHO, SEA, dengue fever, transmission, incidence, modeling, statistical, climate, socio-environmental, factors, variables, projection, vector etc. Different combinations and forms of the key words, in order to identify potential articles and references, were applied. Where necessary and possible, the search was limited by the acquired subject as for example; Life science, Biology, Environmental science, Statistics etc. The search was also limited to journal articles, books and reports written in English or Czech language. The availability of full articles was also one of the determining factors. References and citations of the articles identified were checked, in order to ensure that all relevant articles were included.

The essential raw data were first processed by appropriate statistical methods. Further, these were used within a selected mathematical method, to propose a model for estimation of dengue fever incidence in the Philippines. The data were obtained from different available data sources. The population dataset was retrieved from the official 2010 Population Census published by the Philippine Statistics Authority. Any additional population data were obtained from previous 2000 and 2007 Population Census. The raw population data were first distributed into 17 groups, where each group represented one of the17 regions present in the Philippines. Further, the data were processed by using some of the basal statistical methods e.g.; estimation of population growth rate, average year population, interpolation of the data and population numbers for the years, where the data were missing, were estimated. The data on dengue cases in the Philippines were obtained from the Disease surveillance reports provided by the Republic of the Philippines Department of Health, National Epidemiology center, Public Health Surveillance and Informatics Division. The data were further processed into dengue incidence rate. The remaining data necessary for the model prediction were the climatic risk factors data. These data were obtained from the Global Surface Summary of the Day, National Oceanic and Atmospheric Administration, the U.S. Department of Commerce. The selected climatic

parameters of which sufficient amounts of data were available, were: temperature, precipitation and dew point. The data were also further statistically processed by for an example; linear interpolation, estimation of mean daily relative humidity etc. All of the basal statistical processing is further described, in more detail, within the chapter dedicated to data processing. For the estimation of the dengue incidence model a generalized linear model was utilized with negative binomial distribution. In order to process the data further, for the model estimation, Spearman correlation and non-hierarchical K-means clustering method was used. All statistical data processing, calculations, estimations or modeling were carried out in either Microsoft Excel 2013 or StatSoft Statistica 12 programs. MathWorks MATLAB R2013a program was utilized when generating most of the figures presented, as it provided adequate figure format which presented, as visually best readable and clear, and was best able to process large datasets which were used.

# 7 MODELING OF DENGUE INCIDENCE IN THE PHILIPPINES

This chapter describes in relevant detail the selected geographical area of focus; the Philippines. It further describes the basal statistical processing of the raw data, referring to the chosen risk factors. Those data used further for modeling of dengue incidence within regions present in the Philippines. It deals with creation of the prediction model. The creation of the prediction model composes of three phases; selection of parameters, model estimation and model evaluation. Finally, the results of the model are presented. If not stated otherwise, all tables and figures presented in this chapter were created as a result of author's own calculation, delineation and modeling and therefore, there are no sources stated.

## 7.1 The Philippines

The geographical area of the Philippines was chosen for the data processing and the prediction model, because dengue has been a growing concern in the Philippines. Another reason for deciding on the Philippines was its National Epidemic Sentinel Surveillance system which is managed by the National Epidemiology Center of the Department of Health and maintains surveillance of notifiable diseases, in this case dengue fever.[53]

Dengue is both the most significant vector-borne disease and a major health burden in the Philippines. In SEA the Philippines is one of the countries where the annual amount of dengue infection reaches the highest numbers, according to [27] it was within the range 2.75–7.5 million. Together with Indonesia, the Philippines account for three-quarters of those living in extreme poverty in terms of ASEAN region.[3] It belongs among four countries which account for more than 90 % of the total dengue cases reported within the region.[6]

The Philippines, officially known as the Republic of the Philippines, has the population about 92,337,852 according to 2010 census. The average annual population growth rate was 1.90 % between the period 2000 and 2010. The population percentage living in urban areas is around 58.5 %. A major proportion (37.3 %) of the population lives in only three regions; Calabarzon with 11.74 million people, the capital Metro Manila with 11.55 million people and Central Luzon with 9.72 million people. The Philippines comprises of 7107 islands, those are distributed into 3 island groups which are further subdivided into 17 regions. Fig. 3 (b) shows the geographical distribution of each of the 17 regions.[53]

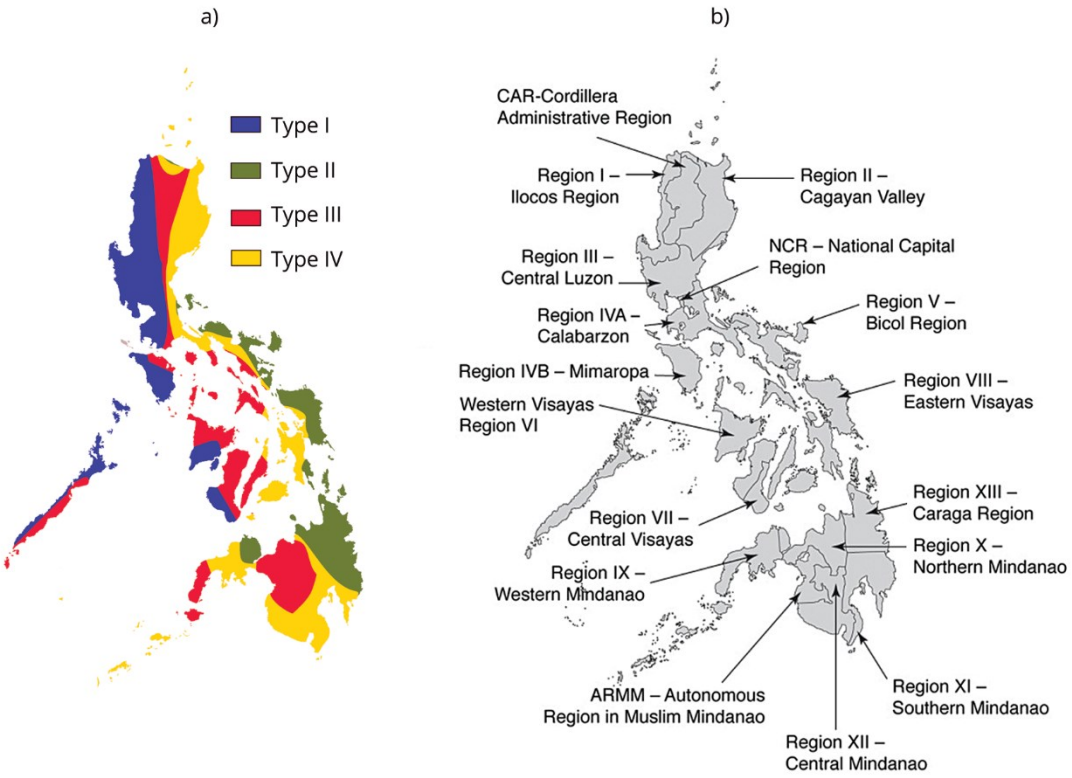


Fig. 3: Philippines (b) 17 administrative regions[53] and (a) four climate types[54]; Type I – two pronounced seasons, dry from November to April, wet during rest of the year. Type II – no dry season with a very pronounced rainfall from November to April and wet during rest of the year. Type III – seasons are not very pronounced; relatively dry from November to April, wet during rest of the year. Type IV – rainfall is more or less evenly distributed through the year.

The dengue cases within the country vary across the country's geographical area. It is often influenced by the geographical location of specific regions e.g.; northern part with mountain ranges which is cooler or southern part with higher temperatures and low lands.[6] The climate in the Philippines is tropical and maritime. It is characterized by high temperatures, humidity and rainfall. Based on temperature and rainfall two seasons are distinguished; 1) rainy season from June to November and 2) a dry season from December to May. However, the dry season is further subdivided into; a) dry cool season from December to February and b) dry hot season from March to May. Based on rainfall four climate types are distinguished, these are graphically and by description defined in Fig. 3.[54]

## 7.2 Data processing

### 7.2.1 Population dataset

The major population set was obtained from the official 2010 Population Census published by the Philippine Statistics Authority, other additional data on population numbers e.g. population predictions and estimated population growth rate were obtained from official 2000 and 2007 Population Census published also by the Philippine Statistics Authority. The population data were first sorted according to the 17 Philippine regions and years for which the data were available, those are; 2007 and 2010.

Then, the population growth rate for the period 2007–2010 was estimated. The formula used for this estimation was obtained from the official *population growth rate United Nations (UN) guideline* and is as follows[55]:

$$r = \frac{100 \ln(P_2 - P_1)}{t_2 - t_1}, \tag{12}$$

where *r* is the population growth rate and $P_1$ and $P_2$ are the number of persons at times $t_1$ and $t_2$. Tab. 5 shows the comparison of annual population growth rate (*gr*) for the time period of years 2000–2010 published within the 2010 Population Census[56] and author's estimated annual population growth rate for the time period of years 2007–2010.

Tab. 5: Annual population growth rate (gr) for the time period 2000-2010 and 2007-2010 expressed in [%] in each region.

| **Region** | **I** | **II** | **CAR** | **III** | **IV A** | **IV B** | **NCR** | **V** | **VI** |
|---|---|---|---|---|---|---|---|---|---|
| ***gr* (2000–2010) [%][56]** | 1.23 | 1.39 | 1.70 | 2.14 | 3.07 | 1.79 | 1.78 | 1.46 | 1.35 |
| ***gr* (2007–2010) [%]** | 1.46 | 1.90 | 2.08 | 1.44 | 2.34 | 2.34 | 0.87 | 1.99 | 1.26 |

| **Region** | **VII** | **VIII** | **IX** | **X** | **XI** | **XII** | **ARMM** | **Caraga** |
|---|---|---|---|---|---|---|---|---|
| ***gr* (2000–2010) [%][56]** | 1.77 | 1.28 | 1.87 | 2.06 | 1.97 | 2.46 | 1.51 | 1.49 |
| ***gr* (2007–2010) [%]** | 2.06 | 1.55 | 1.77 | 2.80 | 2.36 | 2.36 | -7.91 | 1.92 |

Based on the available population data and the estimated population growth rate the remaining population numbers for years 2009, 2011, 2012 and 2013 were estimated. From the obtained population estimations the average year population (AYP) was then calculated in order to get more corresponding population data for each year and region. Tab. 6 presents the overview of

all regions and the average year population estimations for years 2009, 2010, 2011, 2012 and 2013.

Tab. 6: Average year population for all regions for the time period 2009-2013.

| Region | AYP 2009 | AYP 2010 | AYP 2011 | AYP 2012 | AYP 2013 |
|---|---|---|---|---|---|
| **I** | 4,639,904 | 4,708,267 | 4,777,626 | 4,847,248 | 4,917,884 |
| **II** | 3,134,450 | 3,194,850 | 3,256,398 | 3,318,252 | 3,381,282 |
| **CAR** | 1,561,792 | 1,594,761 | 1,628,414 | 1,662,252 | 1,696,794 |
| **III** | 9,900,397 | 10,044,921 | 10,191,531 | 10,338,691 | 10,487,976 |
| **IV A** | 12,144,797 | 12,434,146 | 12,730,268 | 13,028,232 | 13,333,170 |
| **IV B** | 2,636,850 | 2,699,630 | 2,763,879 | 2,828,527 | 2,894,687 |
| **NCR** | 11,642,614 | 11,745,082 | 11,848,445 | 11,952,040 | 12,056,541 |
| **V** | 5,250,768 | 5,357,062 | 5,465,475 | 5,574,457 | 5,685,613 |
| **VI** | 6,956,771 | 7,044,958 | 7,134,252 | 7,223,835 | 7,314,543 |
| **VII** | 6,576,214 | 6,713,925 | 6,854,475 | 6,995,790 | 7,140,018 |
| **VIII** | 3,995,330 | 4,057,866 | 4,121,369 | 4,185,128 | 4,249,874 |
| **IX** | 3,308,069 | 3,367,480 | 3,427,944 | 3,488,688 | 3,550,509 |
| **X** | 4,106,489 | 4,223,874 | 4,344,544 | 4,466,122 | 4,591,103 |
| **XI** | 4,296,054 | 4,399,378 | 4,505,143 | 4,611,572 | 4,720,516 |
| **XII** | 3,958,762 | 4,054,062 | 4,151,615 | 4,249,781 | 4,350,269 |
| **ARMM** | 3,642,834 | 3,370,852 | 3,120,371 | 2,873,690 | 2,646,510 |
| **Caraga** | 2,355,303 | 2,401,261 | 2,448,104 | 2,495,183 | 2,543,168 |

When studied phenomena is age related, the standardization of population dataset is often applied for comparison of countries with different age structure. However, here the standardization of data was neglected. As it was verified by the Chi-square test that the distribution of population age structure is consistent, thus not showing any significant statistical deviations, for all 17 regions.

### 7.2.2 Dengue cases dataset

The data on dengue cases were obtained from the Disease surveillance reports provided by the Republic of the Philippines Department of Health, National Epidemiology center, Public Health Surveillance and Informatics Division. The data were provided for epidemiological weeks, there are 52 weeks in each year. The epidemiological week serves for a comparison of the same epidemiological data within different years. Although every few years one week is excluded from a given year as it would create 53$^{rd}$ week, which would not allow the comparison, because it would then modify the comparison time period, such exclusion was for example in 2008-2009 or 2014-2015. Available data were for years 2009, 2010, 2011, 2012 and 2013. The dengue

cases data except being distributed into weeks, were also distributed according to each region. Dengue cases surveillance reports were not available for all weeks within a given year, sometimes even several weeks were either not reported or the reports were not available.

Fig. 4 represents raw dengue cases dataset for all regions for the whole five year period; 2009-2013 in a relationship to epidemiology weeks. As it is visible where data were not available or reported the curves are disrupted. Colors of each curve represent distinct regions. Maximum occurrence of dengue cases is visible within each represented year. During 2010 the peak reached the highest point at around 2000 cases.

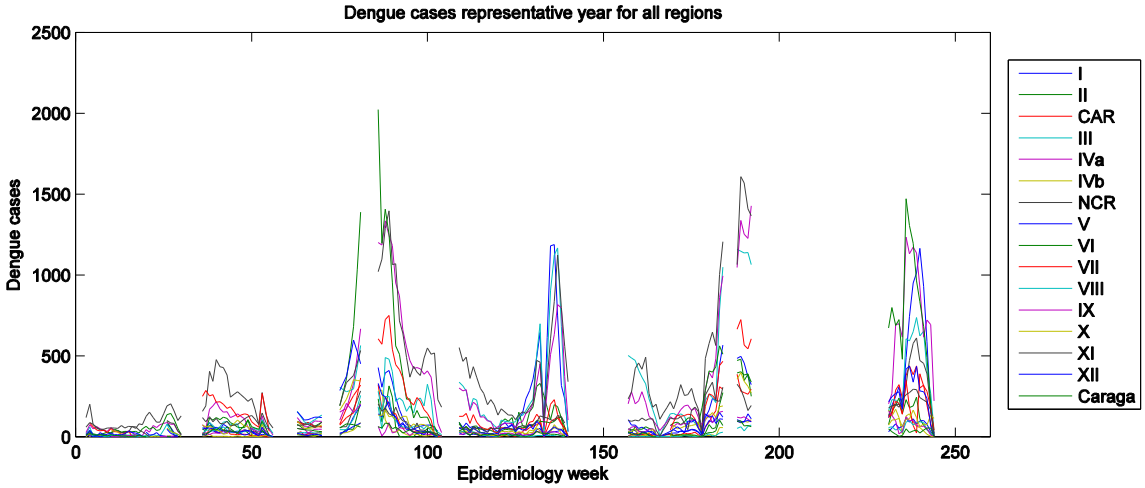


Fig. 4: Raw regional dengue data (2009-2013).

**Long term trend and seasonality**

In order to be able to observe the dengue incidence long term trend and seasonality the data were aggregated to representative year for all regions.

Furthermore, the dengue cases data together with the estimated average annual population data were used for the calculation of relative dengue *Incidence Density rate*, sometimes also defined only as *Incidence rate*. The formula used for calculation has been already introduced in the previous chapter concerning essential epidemiologic terms. The results were then interpreted as a number of dengue cases per 1 000 000 population.

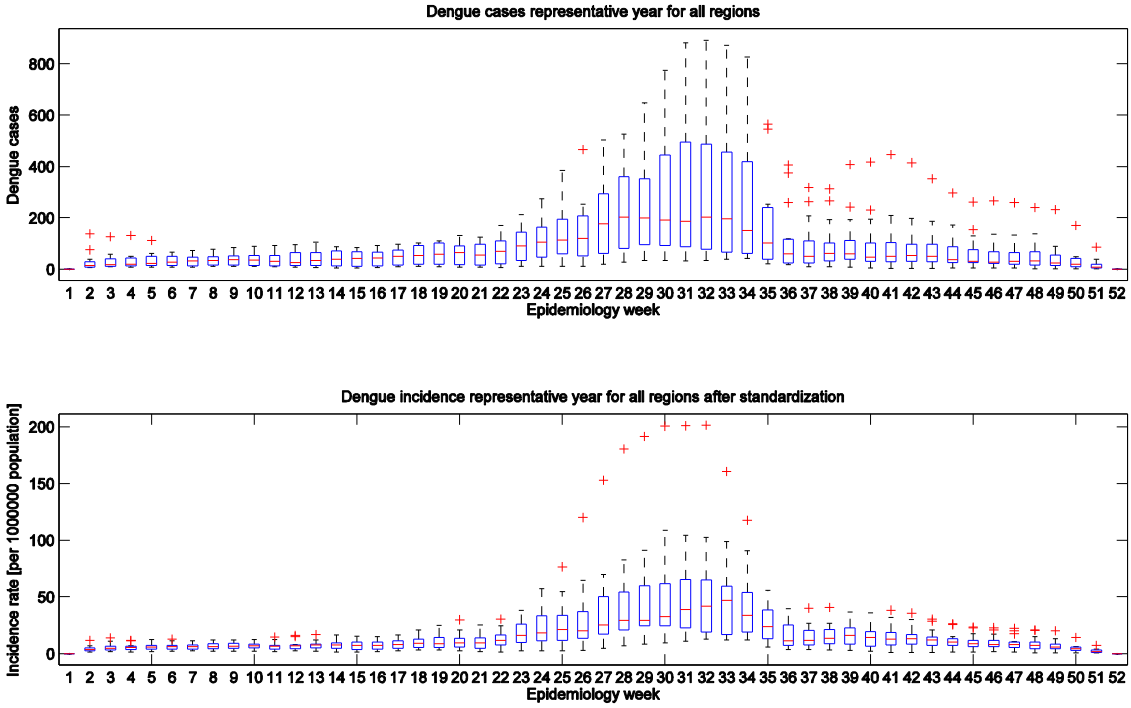


Fig. 5: Dengue cases and dengue incidence representative year for all regions, respectively.

Fig. 5 shows two distinct box plots, where the top part of the figure defines the relationship of dengue cases and epidemiology week data. Together they show the representative year, composed of 52 weeks, generated from all of the data from the whole five year period. The missing dengue cases data were estimated by linear interpolation and smoothed out by the moving average method with the length of the time window 3 weeks. The representative weekly dengue cases for one region were set as a median of cases in corresponding weeks within each year. The bottom part of the figure defines the relationship of dengue Incidence rate, interpreted as a number of cases per 1 000 000 population, and epidemiology weeks. Together they show the dengue incidence representative year for all regions after standardization. The raw dengue cases data were standardized to 1 000 000 population, therefore the incidence rate was calculated.

### 7.2.3 Climatic dataset

The climatic dataset representing the chosen climatic risk factors were obtained from the Global Surface Summary of the Day, National Oceanic and Atmospheric Administration, the U.S. Department of Commerce. The database includes daily summary from 54 meteorological stations present in the Philippines and distributed among 16 out of the total 17 regions. There were no climatic data available for the ARMM therefore, this region is not further considered. Simultaneously, no sufficient extent of climatic measurement was available for 1 of 4 meteorological stations located in region II and 2 of 7 stations from region VIII thus, these stations were excluded from further processing.

Out of the available climatic risk factors the following were chosen; mean daily temperature, minimum daily temperature, and maximum daily temperature, also dew point and daily precipitation. The estimation of mean daily relative humidity was carried out based on the dew point and mean daily temperature data. Units of each parameter were converted into metric system of measurement. Further, the data were checked visually for any abnormal or extreme values. Then, missing or incorrect data were estimated by data obtained from linear interpolation. Some stations do not provide report in case of 0 precipitation, concurrently, some stations may not provide any report, and nevertheless, this fact would not mean there was no precipitation. Therefore, all days with no report were considered as days with 0 precipitation. One station from region VIII was excluded from the precipitation dataset due to insufficient valid data reporting. The differences in altitudes of the meteorological stations were neglected as most of the stations are located within such region's altitude, where majority of population is distributed. In case of regions with more than one meteorological station the daily value of a given parameter was set as an average of values from all available stations within that region.

**Long term temperature trends and seasonality**

Fig. 6 shows a) the relationship of minimum temperature and epidemiology week, which together present the minimum temperature representative year for all regions. The reduced minimum temperature variance of the regions is first clearly visible around week 20. The highest minimum temperature appears also around week 20 and is about 26 ºC. The lowest minimum temperature does not decline below 18 ºC. The extremes, visible as red crosses, are present primarily for CAR region, which is geographically located in cooler, northern, mountain part of the country. It is necessary to eliminate these extremes for further modeling, as those can present a problem since regression models are sensitive to such. Therefore, CAR has been excluded from further calculations. Part b) shows the relationship of maximum temperature and epidemiology week, which again together present the maximum temperature representative year for all regions. It is clearly visible, where the reduced maximum temperature variance of the regions occurs, in this case it is around week 20. The highest maximum temperature appears around week 20. The lowest maximum temperature does not fall below 25 ºC. Part c) presents the relationship of mean temperature and epidemiology week. Together they show the mean temperature representative year for all regions. The reduction in mean temperature variance among regions is again visible around week 20. It is visible from the Fig. that all three variables; minimum, maximum and mean temperature correlate. The statistical significance of correlation was confirmed by calculation of Spearman correlation coefficient, where the significance level was $p<0.05$.

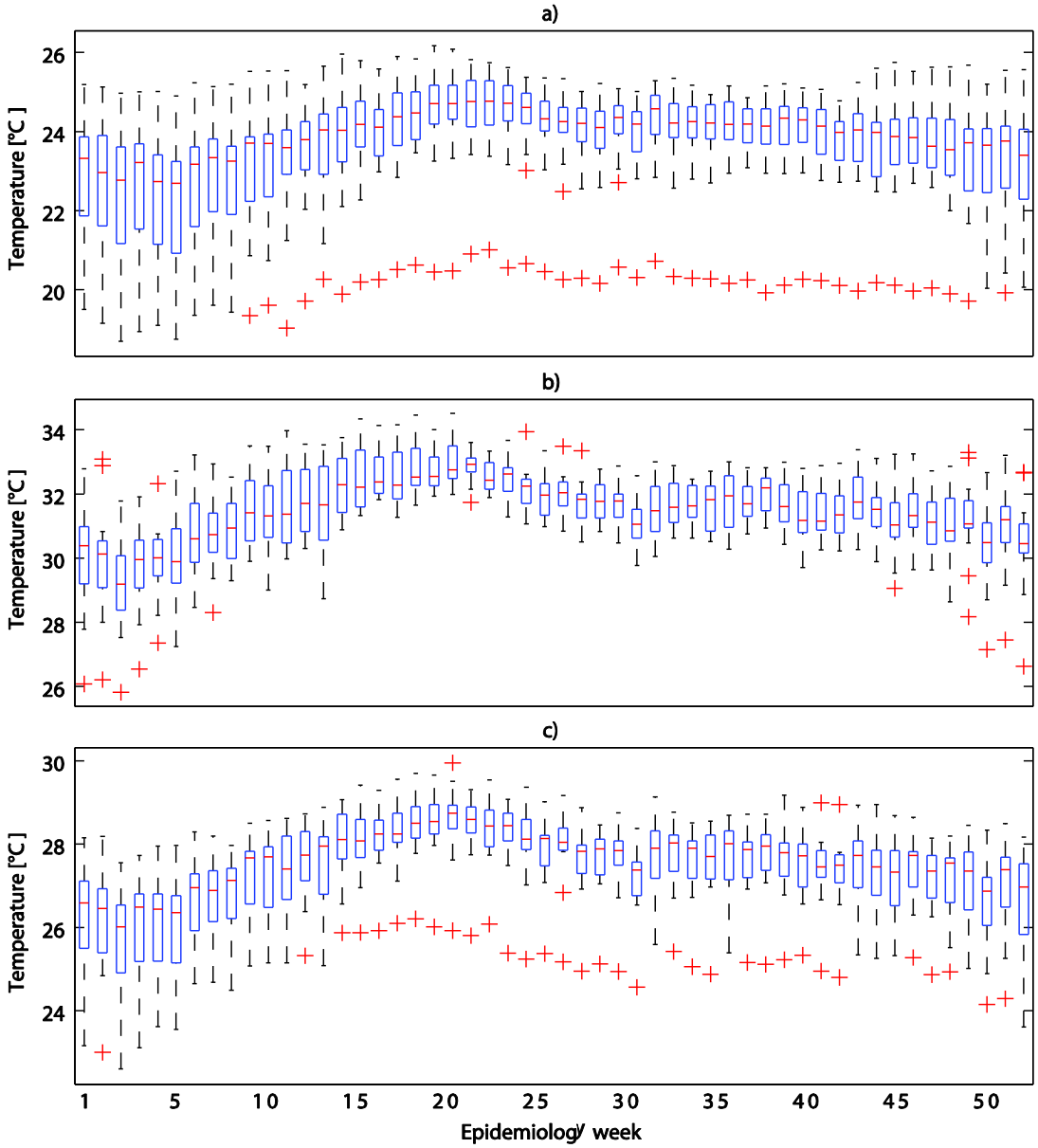


Fig. 6: Minimum (a), maximum (b) and mean (c) temperature [ºC] representative year for all regions.

**Long term precipitation trend and seasonality**

Fig. 7 shows another statistically processed climatic risk factor. It presents the relationship of precipitation and epidemiology week, thus a weekly cumulation of precipitation of all the regions. As a result the figure shows cumulative precipitation representative year for all regions. The highest value of median cumulative precipitation is visible around week 30. However, the reduction in variance within the regions is visible in between weeks 12 and 15. The highest amount of precipitation appears around week 30 and continues until about week 40. This trend corresponds with the rainy season in most of the regions.

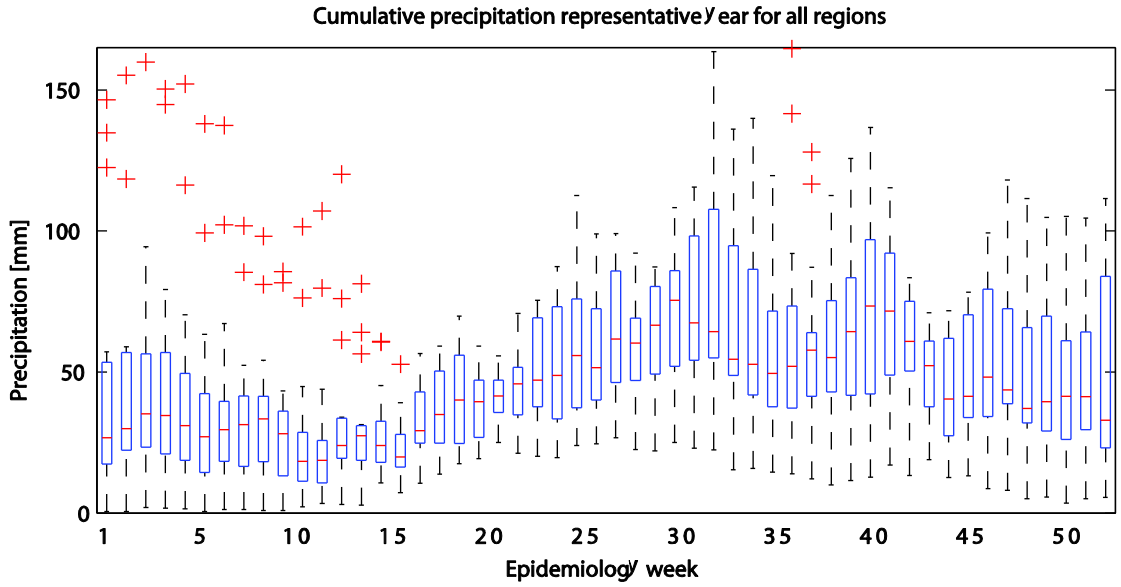


Fig. 7: Cumulative precipitation representative year for all regions.

**Long term dew point trend and seasonality**

Dew point is a measure of atmospheric moisture, it is a temperature to which the air must be cooled in order to reach saturation. Saturation is referred to as the maximum of water vapor possible within the air. The saturation is also defined as the dew point temperature and air temperature being equal. Therefore, the higher the dew point is, the more moisture in the air is present. It is sometimes referred to as Dew Point Temperature.[57]

Fig. 8 shows the dew point representative year for all the regions. The dew point is presented as temperature in ºC units. The decline in variance among the regions is visible starting around week 20 and continuing until about week 43. At the beginning and end of the representative year the variance is larger. The highest median of dew point is visible at around week 20.

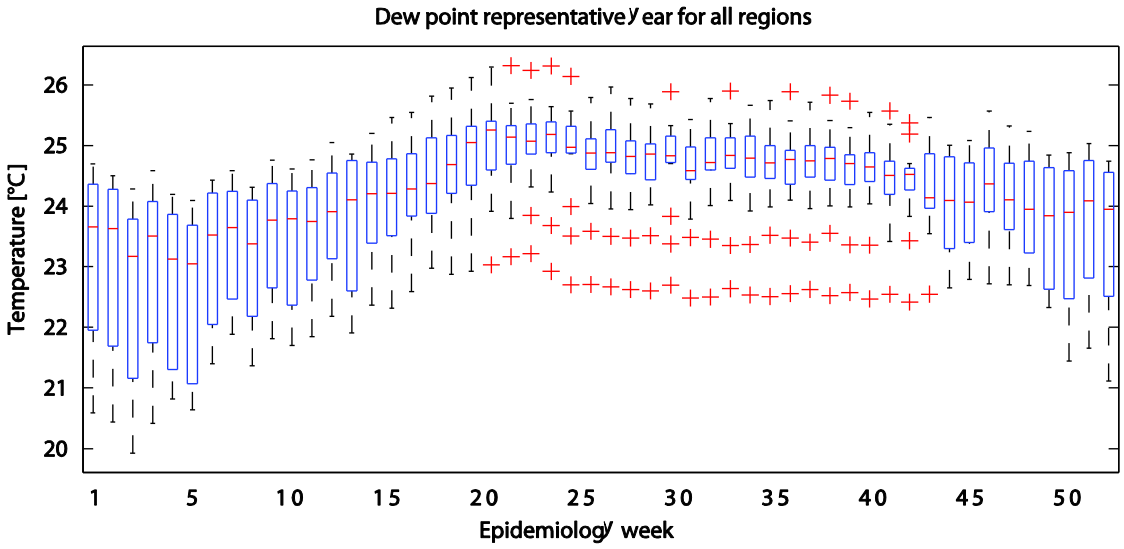


Fig. 8: Dew point representative year for all regions, interpreted as temperature.

**Long term relative humidity trend and seasonality**

The relative humidity was calculated based on the dew point and mean temperature data. Magnus formula was used for the calculation, it is as follows[58]:

$$RH = 100\left[e^{\left(\frac{17{,}625T_d}{243{,}04+T_d}\right)} / e^{\left(\frac{17{,}625T}{243{,}04+T}\right)}\right], \tag{13}$$

The following Fig. 9 shows the relationship of epidemiology week and relative humidity, which is defined in [%]. As a result the relative humidity representative year for all regions is presented. Around week 20 the variance in relative humidity for all regions reduces. This has been also visible for the previous variables; dew point, minimum, maximum and mean temperatures. Concurrently, the variance for the regions is higher at the beginning of the representative year and at the end. Similar trend has been again visible in minimum and mean temperature and dew point temperature. In general, the seasonality within the trend is low. Humidity is quite variable through thou the regions.

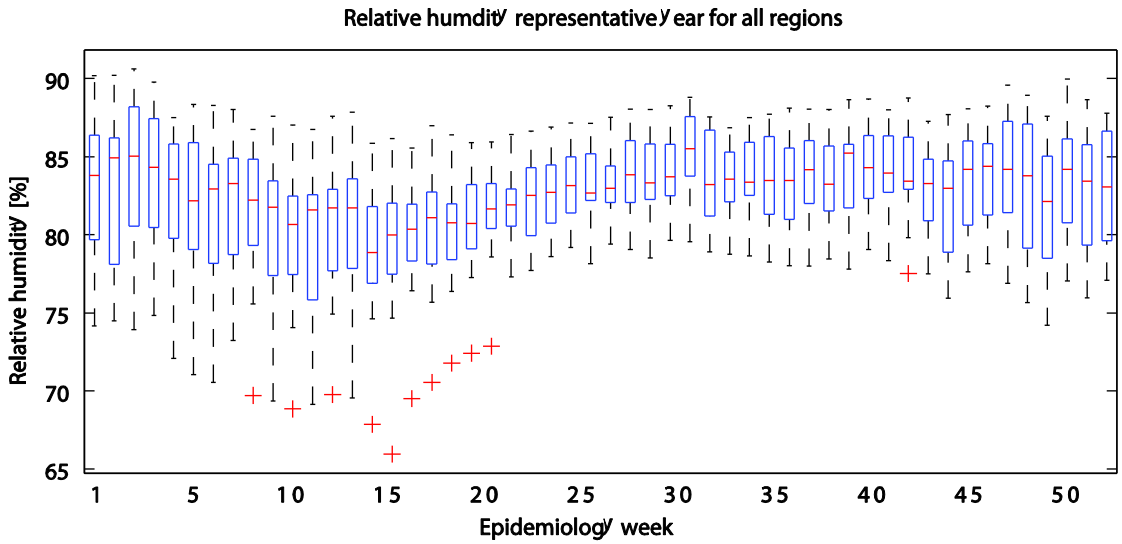


Fig. 9: Relative humidity [%] representative year for all regions.

### 7.2.4 Time lag

In general a time lag is an interval of time between occurrence two related phenomena. In epidemiology modeling this may influence the explanatory power of a model.[44] In our case it was essential to look for a time lag in terms of the effect of climatic variables on the *aedes aegypti* life cycle and dengue transmission cycle, therefore the effect on the dengue cases or further, the dengue incidence rate. In other words, what time period it will take for increased or declined values of climate variables to affect the dengue incidence rate.

The relationship and comparison of; minimum, maximum and mean temperature, dew point, relative humidity, cumulative precipitation and dengue incidence rate and epidemiology week is shown in Fig. 10. Further, the time lag between each variable is delianeated. It presents

the temporal distribution of standardized climate variables at national level and incidence rate. All the variables were transformed to mean variables for the Philippines as a whole and then normalized by the min–max standardization method to dimensionless units [-]. This was in order to be able to compare them and show their relationship in only one figure.

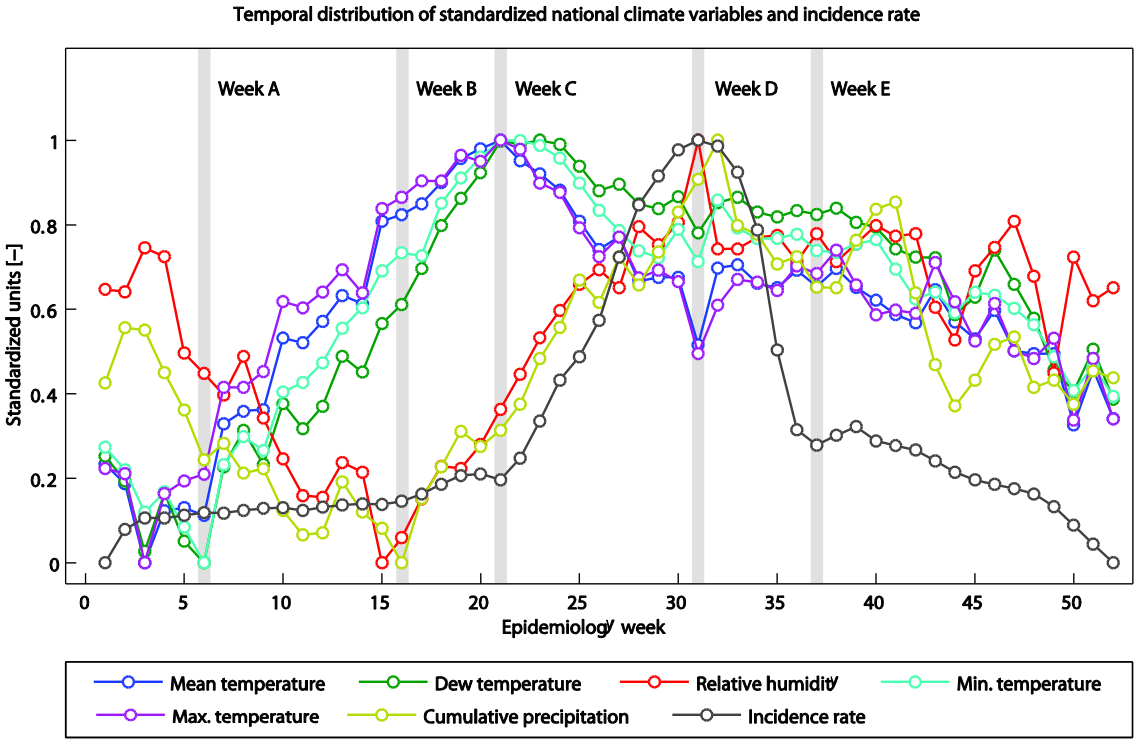


Fig. 10: Temporal distribution of standardized national climate variables and incidence rate. [-]

In Fig. 10, each variable is represented by a certain color to make the visibility of relationships simpler. Five week columns A to E were depicted to highlight different important facts of the relationships.

Week A, which is approximately week 6, shows the beginning of an increasing trend of the temperature variables e.g.; minimum and due point temperature, the maximum and mean temperature have already began the increasing trend 2 weeks earlier, but continue such trend over the 6$^{th}$ week together with the above mentioned temperature variables. At the point where other climate variables began their increasing trend the remaining climate variables; cumulative precipitation and relative humidity, began a declining trend. At this point of the year the incidence rate trend is approximately constant.

Week B, around week 16, shows the continuing, simultaneous increasing trend in all the temperature variables. Concurently, it shows the end of the declining trend in cumulative precipitation and relative humidity. At this point of the year the incidence rate begins a slightly increasing trend.

Week C, approximately week 21, clearly presents the peak of the increasing trend of all climate temperature variables and the beginning of their declining trend. Simultaneously, at this point relative humidity and cumulative precipitation have been increasing for about 5 weeks. The incidence rate begins its steep increasing trend.

When week D, about week 31, is reached the trends in previous week C are reversed. Thus, incidence rate, relative humidity and cumulative precipitation reach peak and begin steep declining trend, at the same time, the temperature variables experience increasing trend.

However, by week E, around week 37, all the variables experience gradual declining trend with a few week exception in relative humidity and cumulative precipitation. The highlited week, e.g. A and C concurently represent time lags, of length about 10 weeks, meaning the time period it takes before the effect of changed variable trend affects the incidence rate trend. To show a specific example; it took about 10 weeks, from week A to week B, of increasing climatic temperature trend, for the incidence rate trend to begin an increase as well. Thus, there is a visible effect of change in temperature variables on the incidence rate trend, however it takes about 10 weeks for the effect to be visible. The same but reverse situation is visible between weeks C and D, the time lag was again about 10 weeks, where after 10 weeks of declining climatic temperature variables' trend the incidence rate also began its declining trend.

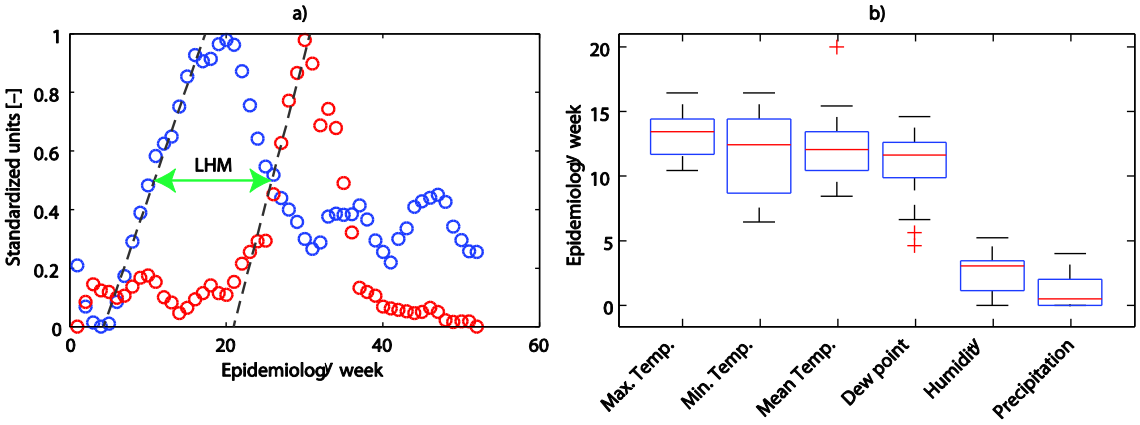


Fig. 11: a) hypothetical method of time lag measurement, b) summary of time lags for each variable and all regions.

Fig. 11 a) shows the method used for determining the lag time for individual regions. It shows in general, how it was visually measured. The blue part determines any variable and red part represents incidence rate. The units of any variable and the incidence rate were standardized by the min-max method in order to be comparable. It determines the distance in weeks between two hypothetic tangents which were depicted upon the upward slope. Lag at half of maximum (LHM) stands for the measurement of the time lag in half of the normalized height. In this case the LHM is approximately 10 weeks.

Fig. 11 b) shows the actual measurement of the time lags within regions. Based on the method in a) the time lags were summarized for each variable and all regions. It shows the time window

at which the variables are needed to be observed. Each box plot shows the variance of regions in weeks. For maximum temperature the median value is set at about 13.5 weeks. Minimum temperature median of lag time is about 12.5 weeks. Mean temperature is close to 11.5 weeks of lag time. The dew point median of lag time is about 10.5 weeks. Humidity time lag median is down to 3.5 weeks and precipitation is at about 1 week. The variance among regions within each variable differs. Reduction in variance of lag time for all regions is visible in precipitation and humidity, although, maximum, mean and dew point temperature are very close to the two as well. Only minimum temperature shows larger variance. As it was also visible in Fig. 10, here it is apparent that when both; precipitation and temperature have an increasing trend, as a result the incidence rate increases as well. On the other hand, when trends of these two variables differ, e.g. when temperature increases and simultaneously, precipitation declines, the incidence trend remains more or less constant and vice versa.

## 7.3 Selection of independent variables

Selection of proper input variable is crucial for feasible model estimation. The variable should be able to explain the dengue incidence. Concurrently, the independent variables need to be independent towards each other, because the dependency could decline influence of an important variable. This section deals with analysis and selection of proper climate variables for further model construction. Because the data do not correspond with the normal Gaussian distribution nonparametric Spearman correlation was performed for the climate variables and dengue cases as a measure of dependency.

Tab. 7 shows the Spearman correlation coefficients for all variables and provides basis for selection of the model input variables. In red, statistically significant numbers are represented. The extent of the numbers represents of how they relate, in other expression, how much they correlate with each other.

Tab. 7: Spearman correlation coefficient for all climate variables and number of dengue cases.

| | **Cases** | **Max. Temp.** | **Min. Temp.** | **Mean. Temp.** | **Dew point** | **Humidity** | **Cum. Prec.** |
|---|---|---|---|---|---|---|---|
| **Cases** | 1,00 | 0,52 | 0,30 | 0,49 | 0,28 | -0,06 | 0,30 |
| **Max. Temp.** | 0,52 | 1,00 | 0,30 | 0,81 | 0,32 | -0,29 | 0,24 |
| **Min. Temp.** | 0,30 | 0,30 | 1,00 | 0,72 | 0,79 | 0,19 | 0,34 |
| **Mean. Temp.** | 0,49 | 0,81 | 0,72 | 1,00 | 0,67 | -0,15 | 0,30 |
| **Dew point** | 0,28 | 0,32 | 0,79 | 0,67 | 1,00 | 0,47 | 0,51 |
| **Humidity** | -0,06 | -0,29 | 0,19 | -0,15 | 0,47 | 1,00 | 0,62 |
| **Cum. Prec.** | 0,30 | 0,24 | 0,34 | 0,30 | 0,51 | 0,62 | 1,00 |

In this case, mean temperature was selected as an adequate input variable. Even though maximum temperature correlated with number of cases slightly more, there is a possibility of maximum temperature variable to also contain extremes, which the regression model is sensitive to. As a second input variable the cumulative precipitation was selected. When compared to the dew point and humidity, it correlates with the dengue cases the most and concurrently, it correlates with both mentioned variables thus, is adequate to be used.

### 7.3.1 Adjustment of selected variables

Therefore, as the most adequate parameters for creation of the model two time series variables were selected; mean temperature and cumulative precipitation. According to crude delineation, there is a visible time lag between increasing mean temperature, cumulative precipitation and an increase in dengue incidence, the time lag is about 9-16 and 1-2 weeks respectively. Optimum time window and time lag were in case of both predictors further estimated by a crosscorrelation. Crosscorrelation can be considered as a measure of similarity of two series as a function of the lag of one relative to the other [59]. Based on the results obtained from the crosscorrelation two climatic indexes were created: $TA_{9w}$ a $PRC_{2w}$.

**Temperature index $TA_{9w}$**

The best outcome in case of mean temperature was obtained for both time window and time lag at 9 weeks. In case of precipitation, the most adequate outcome was time window of 2 and time lag of 1 week. First, moving average of mean temperatures was done within 9 week length time window. Value of $TA_{9w}$ corresponding to incidence week *k* is then given as value of moving average temperature data at week *k*-9. This can be also interpreted as average temperature or temperature trend within week *k*-13 to *k*-5.

**Precipitation index $PRC_{2w}$**

$PRC_{2w}$ for incidence week *k* was calculated as cumulative precipitation within week *k*-3 to *k*-1. For the climatic indicator $PRC_{2w}$ moving average was further also done with a time window 3 weeks, in order to obtain a smoothed trend time series and to reduce random errors, which could originate during delineation or reporting.

Even though the generalized linear models do not have a prerequisite of normal distribution for independent variables, the models are very sensitive to extreme values. Fig. 12 (a) shows, an asymmetry of the data with a right skewed tail and extreme values are visible. Because of this, a further transformation of the indicator was carried out, in order to obtain better symmetry and eliminate the extreme values.

The transformation was done based on the following equation:

$$PRC_{2w}^{t} = \ln(PRC_{2w}^{n} + 1)\,, \tag{14}$$

where, $\mathrm{PRC}_{2w}^{n}$ is the initial indicator and $\mathrm{PRC}_{2w}^{t}$ is the result indicator transformed by natural logarithm. A constant 1 was added in order to avoid negative values for weeks with zero reported precipitation.

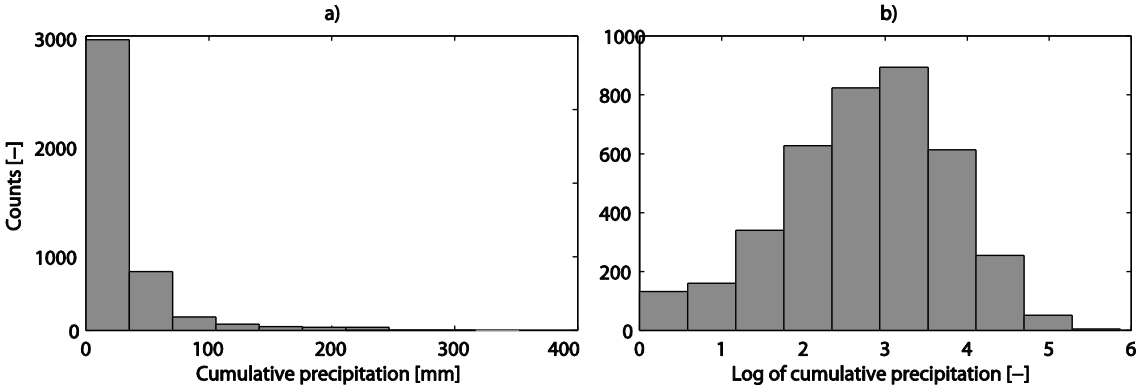


Fig. 12: Histogram of a) initial and b) transformed $\mathrm{PRC}_{2w}$.

Fig. 12 shows the transformation process between the (a) initial data of two week period cumulative precipitation and the (b) newly obtained transformed two week cumulative precipitation. It is clearly visible that the data are not normally distributed and there is huge asymmetry; the data are right skewed. The right histogram shows the result of transformation by equation (14). The data are centralized and influence of extremes was eliminated.

## 7.4 Model estimation

Dengue incidence is a variable, which represents counts in a time series and typically, it corresponds with the Poisson probability distribution. Thus, using Poisson regression seems to be an appropriate method. However, one of Poisson's prerequisites is that mean equals variance. As is shown in (Tab. 8 and Fig. 13), this prerequisite was not met, because for dengue data variance is much higher than mean. Therefore, over-dispersion of data, when conditional variance exceeds conditional mean[60], was visible and thus, it was not possible or statistically adequate to use the Poisson model.

One of possible solutions is a use of another mathematical model, which incorporates a parameter for explanation of the variance [60], such as the Negative Binomial model. Negative Binomial regression belongs among generalized linear models, thus, does not expect normal distribution of independent variable and under specific condition approximates Poisson distribution. With over-dispersed data, it is expected that with Negative Binomial regression the confidence intervals will become narrower, than when compared to the Poisson's confidence intervals.[60]

Tab. 8: Descriptive statistics for new dengue cases

| | N | Mean | Min. | Max. | Lower quartile | Upper quartile | Variance | St. Dev. |
|---|---|---|---|---|---|---|---|---|
| **New cases descriptive statistics** | 2,273 | 133,5 | 0,0 | 2,023 | 19,0 | 133,0 | 49,773,9 | 223,1 |

Tab. 8 shows descriptive statistics for new dengue cases. It includes number of observations (N), mean, minimum, maximum, lower quartile, upper quartile, variance and standard deviation. As described above, it is visible from the resulting numbers that variance is significantly much greater than mean and thus, the Poisson prerequisite is not met.

Fig. 13 shows the distribution of regional dengue cases histogram. It is clear that the data distribution is Poisson like. However, it is graphically visible that the over-dispersion of data is present, as there are some outlying extreme number of cases. Even if the log of values was taken, it would help only to slightly improve the asymmetry, however, the variance would remain.

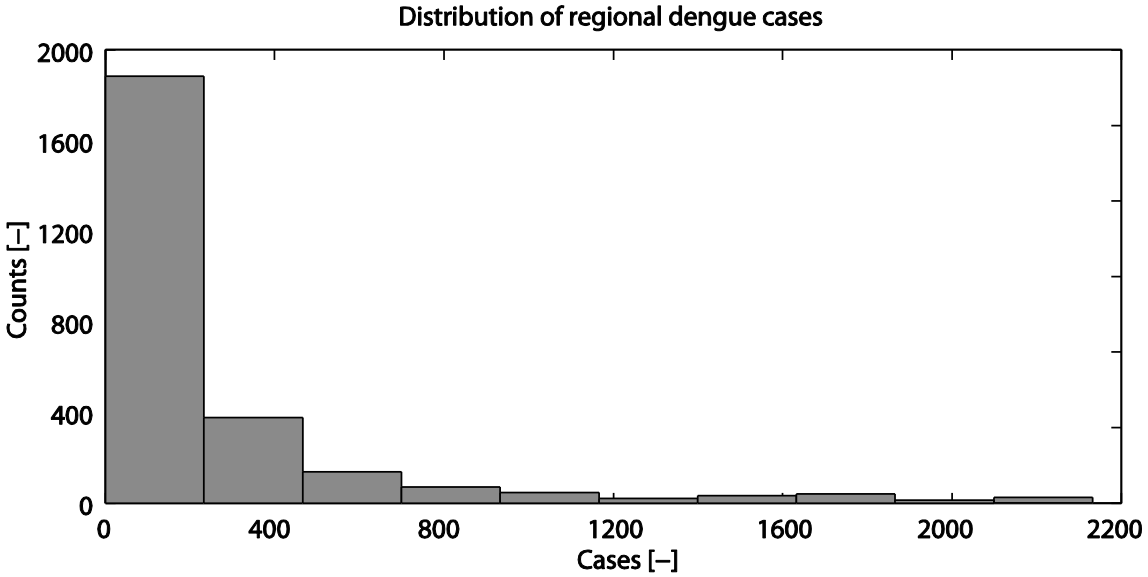


Fig. 13: Distribution of regional dengue cases.

Negative Binomial regression can use as a linkage function natural logarithm (*ln*), exponential function or identity function. With respect to data nature, *ln* function is the most appropriate. The formula for the dengue cases model corresponds and is derived from the one of Poisson regression, already presented in equation (10).

For the natural logarithm linkage function, relationship between Incidence rate ($IR$) and dengue cases ($Y_c$) can be defined as follows:

$$ln(IR) = ln\left(\frac{Y_c}{P_{mean}}\right) = \ln(Y_c) - \ln(P_{mean}), \quad (15)$$

where the incidence rate is equal to dengue cases divided by mean year population ($P_{mean}$). With respect to logarithm rules, logarithm of fraction is equal to difference of logarithms of each variable. Therefore, the equation of the final model is defined as:

$$ln(Y_c) = \beta_0 + \beta_1 T + \beta_2 ln(PRC_{2w} + 1) + ln(P_{mean}). \quad (16)$$

Where, the left side of the equation is expressed as linkage function natural logarithm $ln$ and output dependent cases variable = point estimate $Y_c$. The right side of the equation is expressed as fitted coefficients ($\beta_0$, $\beta_1$, $\beta_2$), independent variable mean temperature $T$, natural logarithm of cumulative precipitation independent variable $ln(PRC_{2w} + 1)$, natural logarithm of mean population independent variable $ln(P_{mean})$.

The point estimate of dengue cases for a certain moment of time can be derived from equation (16) using inverse function to the natural logarithm:

$$Y_c = e^{(\beta_0 + \beta_1 T + \beta_2 ln(PRC_{2w}+1) + ln(P_{mean}))}, \quad (17)$$

where, $Y_c$ is the output variable of dengue cases and $e^x$ is the inverse function to the natural logarithm and $x = \beta_0 + \beta_1 T + \beta_2 ln(PRC_{2w} + 1) + ln(P_{mean})$ is linear predictor.

## 7.5 Model evaluation

The parameter of variance was experimentally set to 1.2 for the model. In order to obtain adequate model statistics described in Tab. 9 and

Tab. 10, different values of variance parameter were experimentally used. However, the most adequate result for the variance parameter was 1.2. When higher values of variance were used, the total model residual deviance was lower, however then, the $Chi^2$ statistics did not result adequately, for example; value/degrees of freedom were then equal to about 0.5 which is statistically incorrect. This can be either a result of under-dispersion, over-dispersion of data or the fact that another significant, explaining variable was not included into the model or a possible existence of a relation between the used variables, which the model is not able to explain.

Basic goodness of fit statistics is shown in Tab. 9 for the whole model. If the expected value of $Chi^2$ distribution is equal to its degree of freedom, the residual deviance should approximate Pearson $Chi^2$ distribution of *(n-p)* degrees of freedom, where $p$ is count of $\beta$

coefficients within the model. Therefore, in a well fitted model the residual deviance should be approximately equal to its degrees of freedom. Then, the mean deviance should be close to one.

Statistical significance level of each independent variable is shown in Tab. 10. It is clear that both selected variables are able to explain (p=0.00), to some extent, observed dengue cases. Tab. 11 shows the estimated values of $\beta$ coefficients used in the model.

Tab. 9: Goodness-of-fit statistics test for Negative Binomial regression model with linkage function *ln.*

| | **Degrees of freedom** | **Statistics value** | **Value/Degrees of freedom** |
|---|---|---|---|
| **Deviance** | 2270 | 2304.737177 | 1.015303 |
| **Pearson Chi2** | 2270 | 2347.408046 | 1.034100 |

Tab. 10: Goodness-of-fit statistics test for Negative Binomial regression for each variable with linkage function *ln*.

| | **Wald Statistics** | **p** |
|---|---|---|
| $\boldsymbol{\beta_0}$ | 1381.517 | 0.00 |
| **ln($\boldsymbol{PRC_{2w}}$)** | 171.736 | 0.00 |
| **ln(Popul)** | 706.1224 | 0.00 |
| **Mean_T** | 213.343 | 0.00 |

Tab. 11: Model estimated $\beta$ coefficients.

| | $\boldsymbol{\beta_0}$ | $\boldsymbol{\beta_1}$ | $\boldsymbol{\beta_2}$ |
|---|---|---|---|
| **Estimated coefficient** | -20.1261 | 0.2972 | 0.3266 |

# 8 RESULTS

## 8.1 Regional coupling between dengue incidence and climate variables

In order to show a possible relation between the explaining ability of climate variables towards dengue incidence and the diverse climate types in the Philippines, a stratification of the regions into groups, according to the measure of their mutual relations, was executed.

The Spearman correlation coefficient was calculated for dengue incidence rate (IR) and adjusted mean temperature ($TA_{9w}$) and cumulative precipitation ($PRC_{2w}$). This was to show the degree of correlation between these variables within all 15 remaining regions. Then, the non-hierarchical K-means clustering method was used to group regions with similar relationship between dependent and independet variables. Number of clusters *k* was given in accordance to the number of climatic types (*k*=4). Therefore, it was possible to explain correlation or no correlation with diverse climatic types. This was applied to three cases; one, for degree of correlation between $TA_{9w}$ and incidence rate; two, for degree of correlation between $PRC_{2w}$ and incidence rate and three, for both variables $TA_{9w}$ and $PRC_{2w}$ and incidence rate.

Fig. 14 shows groups with similar relationship between the dengue incidence rate and a) $TA_{9w}$, b) $PRC_{2w}$ and c) both independent variables; $TA_{9w}$ and $PRC_{2w}$. Region CAR and AMMR were not considered for the calculations and so are depicted in grey color. The reason for their exclusion was explained in previous text.

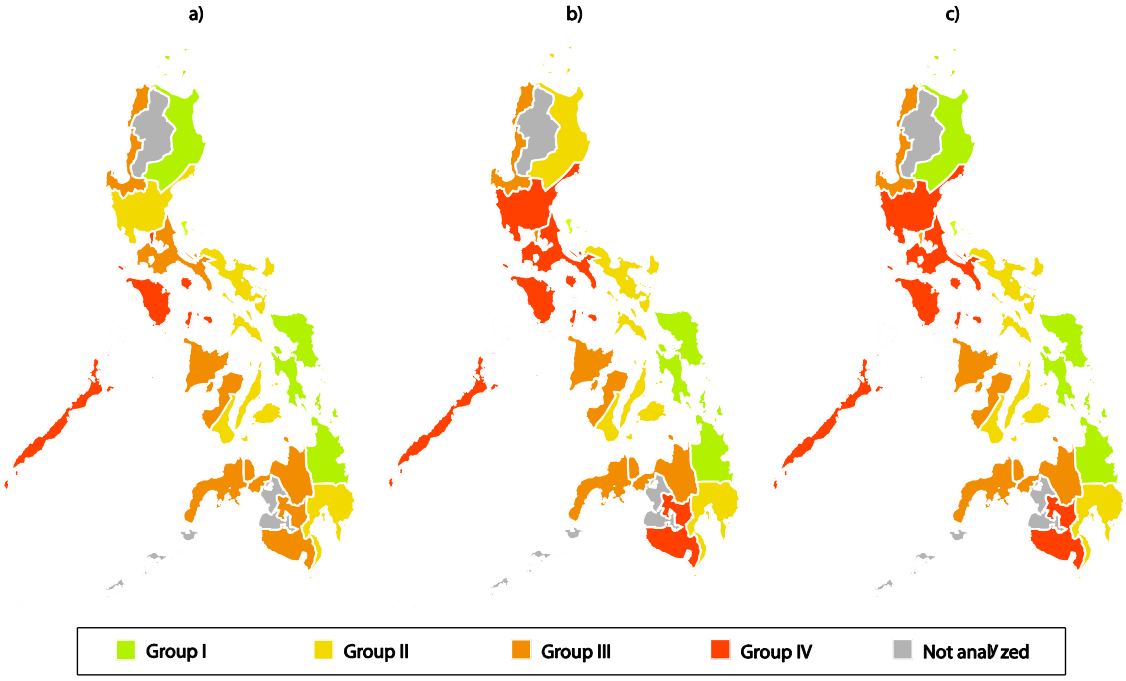


Fig. 14: Philippines' regions grouped by degree of correlation between; a) dengue incidence rate and $TA_{9w}$, b) dengue incidence rate and $PRC_{2w}$, c) dengue incidence rate and both $TA_{9w}$ and $PRC_{2w}$ .

Group I, depicted in green color shows regions with very low correlation. Group II depicts regions with low but higher correlation than in Group I. Group III shows regions with slightly higher correlation and Group IV regions with high correlation. The minimum correlation was 0.01 (typically Group I) and the highest correlation was equal 0.6 (typically Group IV).

The coherence of Group I-IV with climatic types; in a), b) and c) for very low correlation (Group I) corresponds with climate type II, where there is no dry season and continuous rainfall over the year. Therefore, it is obvious that there are other influences which should be taken into account, when predicting incidence for this geographical area.

In all cases; a), b) and c) low correlation (Group II) corresponds to the climatic type IV, where there is no dry season through thou the year and the rainfall is quite evenly distributed. Even that the correlation is statistically significant, there are other influences which need to be taken into account when analyzing the relationship with dengue incidence.

For a) the correlations were for both; higher (Group III) and high (Group IV) very similar. They both partially correspond to climate type I and III. Where it is dry for part of the year (November to April) and wet for the rest of the year. This means that these two variables are statistically more significant in this type of climate within the given geographical area. And thus, are able to define the incidence better.

For b) and c) both higher (Group III) and high (Group IV) correlations evenly correspond to climatic type I and III. This represents the importance of analysis of both variables; $TA_{9w}$ and $PRC_{2w}$ in these geographical areas, in all cases a), b) and c).

In general, where there was small seasonal component the correlations among the two variables and incidence rate were quite low. Compared to areas, where more seasonal divergence is present, the correlations are higher and thus statistically more significant. Therefore, it is visible that the seasonal component increases and decreases the predictive capacity of the climate indicators. Thus, for any future predictions it is rather to analyze the variables and incidence rate for each region separately, to be able to adjust the model for different seasonal diversities and other possible influencing factors.

## 8.2 Observed and predicted dengue cases

Finally, the comparison of the initial recorded dengue cases data and the modelled predicted dengue cases is shown in the following figures. It is visible how the modeled trend of the dengue cases corresponds with the recorded trend and how it develops. For the presentation of the model results four distinct regions were chosen. Each region represents one correlation group for dengue incidence and both climatic variables, presented in Fig. 14. The regions also represent several climatic types, present in the Philippines, in order to be able to compare the degree of correlation and seasonal element within the results as well.

**Region IVa**

Fig. 15 shows the comparison of recorded dengue cases and predicted dengue cases obtained from the predictive model for Region IVa. The region represents group with the highest degree of correlation between dengue incidence and $TA_{9w}$ and $PRC_{2w}$. In terms of climatic type it corresponds with I, III and partially with IV. It is visible that the predicted number of cases trend correspond well with the recorded number of cases trend. Concurrently, as there is the highest correlation the independent variables within the model are best able to explain the dengue cases within a given time point. Colored rectangles within the figure depict those weeks, where there were no reported dengue cases data. Spots, where independent variables were not fully able to explain the variability in dengue cases, are visible at the point, where the predicted values do not follow the recorded values. Large dispersion between predicted and recorded dengue cases are visible at that point.

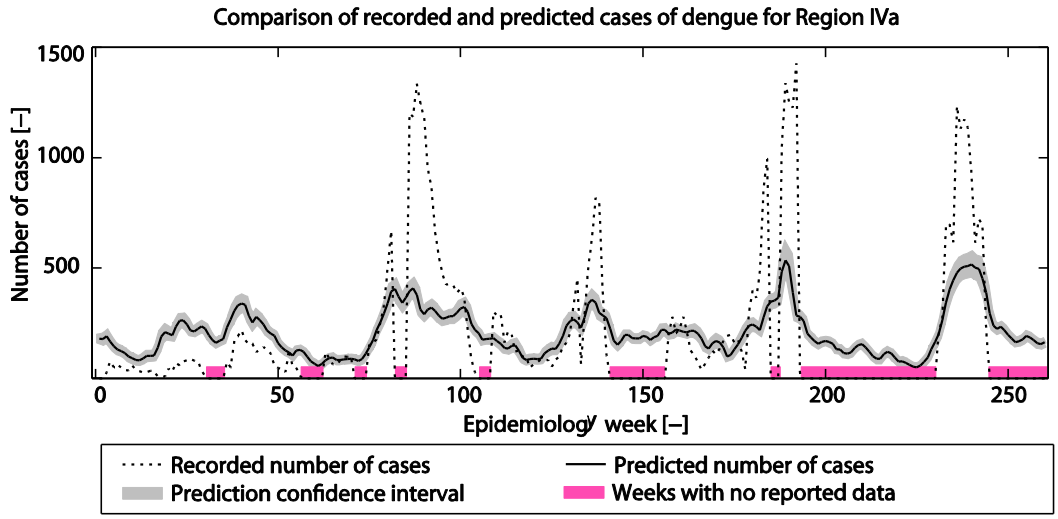


Fig. 15: Comparison of recorder and predicted dengue cases for Region IVa.

**Region VI**

Fig. 16 shows the comparison of recorded and predicted dengue cases for Region VI. This region represents correlation group III, also with higher degree of correlation. This region belongs to I and III climatic type with clearly defined dry and wet season. It is likely, that with more diverse seasons in a given area, the independent variables are able to explain the dengue cases better, than when there is one or two similar seasons. This will be also visible in the following regions. Simultaneously, there are some time points, where the variables were still not fully able to explain variability in dengue. An explanation could be, that another factor, which was not included in the model, was influencing the dengue at that point of time. It can be also interpreted, that the vector thrives rather in areas, where there are both dry and wet seasons, when compared to those areas, where there is rather wet climate through thou the year.

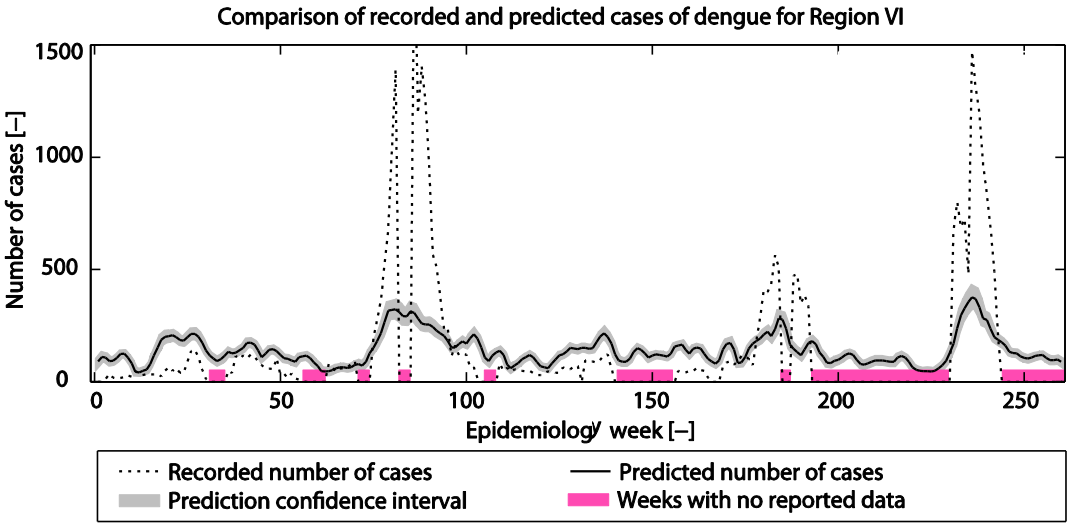


Fig. 16: Comparison of recorded and predicted dengue cases for Region VI.

**Region XIII**

Fig. 17 shows the comparison of recorded and predicted cases for Region XIII. This region represents correlation group I, with the lowest degree of correlation among the variables. The region stands for an equal distribution of climate type II and IV. These are wet seasons for the overall year. This would support the presumption defined in Fig. 16 interpretation, that the more monotonous season within a given area, especially in case of wet season, the lower ability of the variables to explain dengue cases. The predicted trend does correspond with reported trend very little in this case. Another explanation could be that within this geographical area, there is another stronger factor or factors, which would explain the dengue cases better either, than the two selected climate variables or, when incorporated into the model with them.

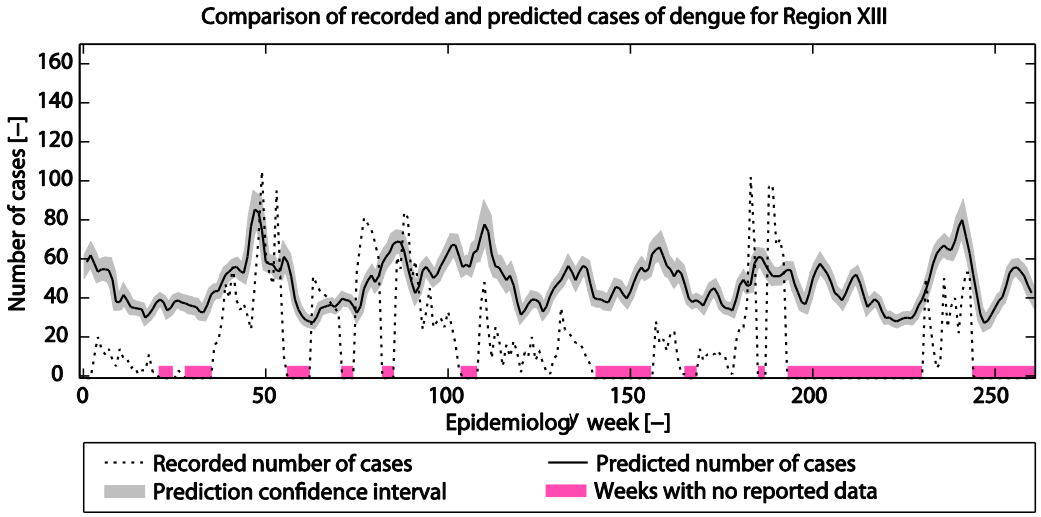


Fig. 17: Comparison of recorded and predicted dengue cases for Region XIII.

**Region V**

Fig. 18 shows the comparison of both recorded and predicted dengue cases in Region V. The region falls into group II in terms of the degree of correlation between incidence rate and the two independent variables. This degree of correlation is rather low, even though statistically significant. It represents season type II and III and IV. Seasons II and IV are more or less representing wet season for the whole year, but partially the central part of the region incorporates season III with both dry and wet season. The predicted cases trend corresponds with the recorded cases trend more accurately than in case of the previous region. However, the predicted cases baseline quite significantly exceed the registered cases. Even though, it seems as if the variables were completely independent towards the dengue cases. Still, when the baseline of the recorded cases would shift higher, some parts of the predicted cases correspond to those recorded. Insufficient reporting within the region could also have an impact on the overall recorded cases baseline.

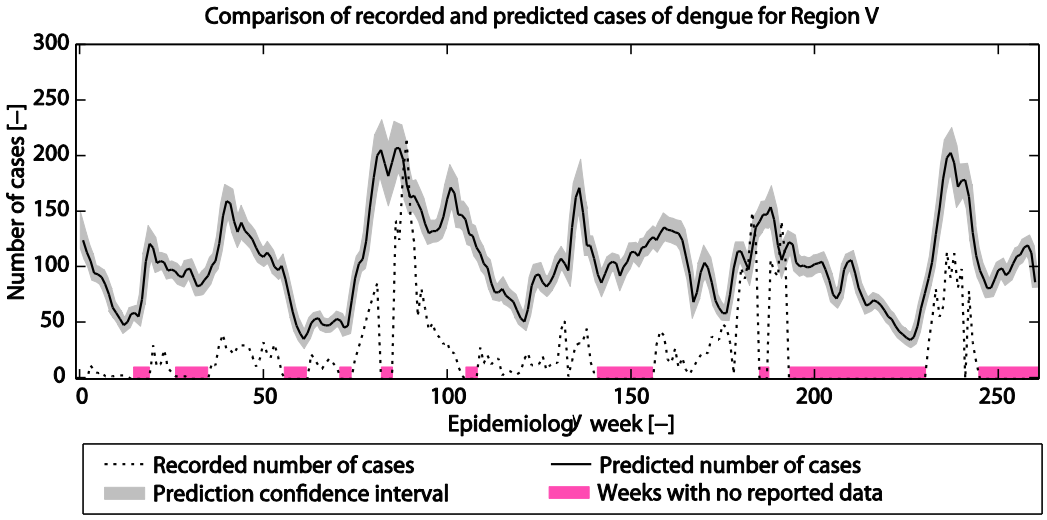


Fig. 18: Comparison of recorded and predicted dengue cases for Region V.

# 9 DISCUSSION

Both $TA_{9w}$ and $PRC_{2w}$ partially do explain the dengue incidence within the 15 regions in the Philippines. Although, the relationship of both climatic variables and dengue incidence is explained, based on the statistics of the model, significantly. Concurrently, there is quite significant randomness in how a given variable explains incidence in a given time point. Still, their ability to fully explain any variability between the dependent and independent variables is not always visible. This ability varies also from region to region. However, in overall results for the Philippines as a whole, the ability is rather low. The correlation among the variables is not very high, even though, statistically it was interpreted as significant within the model.

There can be several explanations of why within the model, the variables were not fully able to explain the dengue incidence and the variability among the variables. For one, the lower correlation coefficient between the variables could be seen as a partial explanation. Two, it is possible that the utilization of only two, independent variables was not sufficient. And thus, only these two variables were not able to explain or reason the variability of the modeled cases. Three, it is rather possible, that one of model prerequisites was not met. As the model was based on counts, where the independence of the events, in this case dengue incidence, is necessary. It is likely that there is a dependency among the dengue incidence events. Also, an exponential element could occur, thus, causing a so called chain reaction, then, one event would influence the following and therefore, it would confirm event dependency. Four, another likely explanation is that there exist other variable or variables, which do influence the incidence significantly and therefore, should have been incorporated within the model. An example of such variables is; any of the socio-environmental or economic factors e.g.; number of water containers within a household or migration. This suggests fifth explanation that the model does not take into account the vector-host relationship e.g.; the vector life cycle and the dengue transmission cycle. Sixth explanation could be the inability of the model to reflect the diverse climatic types within regions, especially within regions, where more climatic types are present. Seventh, it is possible that when the initial raw data were being adjusted in terms of the model prerequisites e.g.; data interpolation, estimation of representative years for each variable etc. the data could lose some of their explanation ability. Finally, it is possible that the selected model is not adequate, as the incidence and the variables are rather not multiplicative and have a different kind of dependency.

Further, the inadequate, initial availability of reported data for either population dataset, climatic dataset or dengue cases dataset was of an influence towards the final model as well. Some of the missing data were substituted by interpolation or averaging, but still, it might be assumed that having the initial raw data would have been more adequate for the model.

# 10 CONCLUSION

Dengue is a rapidly expanding disease both globally, regionally and locally. Both in terms of Southeast Asia region and the Philippines it has a crucial impact on public health and represents a major vector-borne disease. Therefore, it is essential to focus on studying the complex interactions of diverse risk factors which affect dengue incidence and transmission. Among the most commonly known factors belong; temperature, humidity, precipitation, travel and transport, migration, housing condition, human behavior or urbanization. These factors influence dengue incidence through direct or indirect impact on both; the *aedes aegypti* mosquito life cycle and the disease transmission cycle.

The results achieved from the model within this work have shown that $TA_{9w}$ and $PRC_{2w}$ do have an impact on dengue incidence. However, the relationship between increasing and decreasing mean temperature, cumulative precipitation and dengue incidence has shown to be very complex and dynamic. The results show the independent climatic variables are able to explain the dengue incidence only to some extent. Where the two variables used in the model were not able to explain the relationship, it was assumed that other unspecified factors do influence the incidence. It was also realized that for any further modeling, it would be rather effective to model the relationship of any variables for regions separately, rather than for Philippines as a whole. This is due to factors such as; the significant geographical and climatic diversities among the regions, different level and quality of data reporting within each region etc. The quality and level of data reporting within the country were diverse, as some regions do report reliably and others not to such an extent. Also possibly the country's authorities should focus on unification of the reporting systems. Many dengue cases do not get reported, because reporting works only within hospitals and health centers. Without quality data the epidemiologic modeling of the disease does not reflect the real status and therefore, any possible predictions will not be reliable. Within the model used in this work, some of the errors which arose during the analysis could have been influenced by low quality data. Because of none or inadequate reporting two of the overall 17 regions had to be excluded through thou the modeling process, as those would influence the results negatively. The results also suggest that inclusion of another factor would likely improve the relationship and increase the explaining power of the model. This is also supported by the fact that only a little amount of studies could be found focusing on impact of other e.g.; socio-environmental, factors on dengue transmission and incidence.

Therefore, for any further research and modeling of relationship between dengue incidence and possible risk factors, it is essential to focus on more complex connection and utilize different sets of risk factors, in order to receive a model which would be able to fully explain the

relationship and impacts of the factors on incidence and also provide reliable predictions. These could be further used for disease control and elimination programs. Mathematical models are an effective instrument for disease control, specifically, dengue control. Therefore, results obtained from such models will provide an important base for future creation of surveillance, control and elimination programs, which are essential for disease elimination and possibly eradication. To be able to provide such modeling, the raw data need to be of high quality and therefore, reporting systems at global, regional and local level must unify and improve their reliability. The quality of data is essential for dengue cases but also for any data in connection to risk factors e.g.; meteorological reporting, population census etc.

# LIST OF REFERENCES


1. World Health Organization. *First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases*. *World Health Organization* (2010). doi:10.1177/1757913912449575

2. Feasey, N., Wansbrough-Jones, M., Mabey, D. C. W. & Solomon, a. W. Neglected tropical diseases. *Br. Med. Bull.* **93,** 179–200 (2010).

3. Hotez, P. J. *et al.* Neglected Tropical Diseases among the Association of Southeast Asian Nations (ASEAN): Overview and Update. *PLoS Negl. Trop. Dis.* **9,** 1–15 (2015).

4. Malavige, G. N., Fernando, S., Fernando, D. J. & Seneviratne, S. L. Dengue viral infections. *Postgrad. Med. J.* **80,** 588–601 (2004).

5. Cossio, M. L. T. *et al. A brief guide to emerging infectious diseases and zoonoses*. *World Health Organization* (2014). doi:10.1007/s13398-014-0173-7.2

6. Rajiah, K. *et al.* Dengue in Asean countries : A simple review on prevalence and current approches. *Indo Am. J. Pharm. Res.* **4,** 1857–1866 (2014).

7. Schoenbach, V. J. & Rosamond, W. D. *Understanding the Fundamentals of Epidemiology: An Evolving Text*. (University of North Carolina, 2000).

8. Kirkwood, B. R. & Sterne, J. A. C. *Essential Medical Statistics*. (Blackwell Publishing, 2003). doi:10.1007/s13398-014-0173-7.2

9. Guha-Sapir, D. & & Schimmer, B. Dengue fever: new paradigms for a changing epidemiology. *Emerg. Themes Epidemiol.* **2,** (2005).

10. Banu, S., Hu, W., Hurst, C. & Tong, S. Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors. *Trop. Med. Int. Heal. TM IH* **16,** 598–607 (2011).

11. Naish, S. *et al.* Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infect. Dis.* **14,** 167 (2014).

12. World Health Organization. *Sustaining the drive to overcome the global impact of neglected tropical diseases*. *Second WHO report on neglected tropical diseases* **1,** (2013).

13. Utzinger, J. *et al.* Neglected tropical diseases: Diagnosis, clinical management, treatment and control. *Swiss Med. Wkly.* **142,** 1–16 (2012).

14. Adams, J., Gurney, K. A. & Pendlebury, D. *Thomson Reuters Global Research Report - Neglected Tropical Diseases*. (2012). at <http://www.apjtcm.com/news/Neglected Tropical Diseases.pdf>

15. IOM (Institute of Medicine). *The Causes and Impacts of Neglected Tropical and Zoonotic Diseases: Opportunities for Integrated Intervention Strategies: Workshop Summary*. *Nature* (The National Acedemies Press, 2011). doi:10.17226/13087

16. Daumerie, D. & Kindhauser, M. K. *Intensified Control of Neglected Diseases. Report of an International Workshop*. *WHO Communicable Diseases* (2004). at <http://apps.who.int/iris/bitstream/10665/68529/1/WHO_CDS_CPE_CEE_2004.45.pdf>

17. World Health Organization. *Accelerating Work To Overcome the Global Impact of Neglected Tropical Diseases – A roadmap for implementation*. (2012). at <http://apps.who.int/iris/bitstream/10665/70809/1/WHO_HTM_NTD_2012.1_eng.pdf>

18. Molyneux, D. H. & Malecela, M. N. Neglected Tropical Diseases and the Millennium Development Goals-why the 'other diseases' matter: reality versus rhetoric. *Parasit. Vectors* **4,** 234 (2011).

19. World Health Organization. *Investing to overcome the global impact of neglected tropical diseases - Third WHO report on neglected tropical diseases*. (2015). at <http://apps.who.int/iris/bitstream/10665/152781/1/9789241564861_eng.pdf?ua=1>

20. Narain, J. P. *et al.* Elimination of neglected tropical diseases in the South-East Asia Region of the World Health Organization. *Bull. World Health Organ.* **88,** 206–210 (2010).

21. Hirschman, C. & Bonaparte, S. in *Demographic change in Southeast Asia: Recent histories and future directions.* (eds. Williams, L. & Guest, M. P.) **9,** 1–37 (Cornell Southeast Asia Program Publications, 2012).

22. World Health Organization. *Epidemiological Review of Leprosy in the Western Pacific Region, 2008-2010*. *World Health Organization* (2011). doi:10.1073/pnas.0703993104

23. Department of Health Statistics and Information Systems WHO. *WHO methods and data sources for global burden of disease estimates 2000-2011*. *Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2013.4* (2013). at <http://www.who.int/healthinfo/statistics/GlobalDALYmethods_2000_2011.pdf?ua=1>

24. World Health Organization. Disability-adjusted life years (DALYs), 2000-2012, South-East Asia region. *Global health estimates* (2014). at <http://apps.who.int/gho/data/view.main.DALYRATEREG6SEARV?lang=en>

25. World Health Organization and Asian Development Bank. *Addressing Diseases of Poverty: An Initiative to Reduce the Unacceptable Burden of Neglected Tropical Diseases in the Asia Pacific Region*. (2014). at <http://www.wpro.who.int/mvp/documents/docs/NTD_case_for_support_final1.pdf>

26. Huy, N. T. *et al.* Factors Associated with Dengue Shock Syndrome: A Systematic Review and Meta-Analysis. *PLoS Negl. Trop. Dis.* **7,** 1–15 (2013).

27. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).

28. Guzman, M. G. *et al.* Dengue: a continuing global threat. *Nat. Rev. Microbiol.* **8,** 7–16 (2010).

29. Centers for Disease Control and Prevention. Dengue: Mosquito life cycle. *Dengue* (2012). at <http://www.cdc.gov/dengue/resources/factSheets/MosquitoLifecycleFINAL.pdf>

30. Williams, R. E., Bennett, G. W., Hill, C. & Sinsko, M. J. *Mosquitos In And Around The Home*. (2013). at <http://extension.entm.purdue.edu/publications/E-26.pdf>

31. Nature Publishing Group. Dengue Transmission. *Nature Education* (2014). at <http://www.nature.com/scitable/topicpage/dengue-transmission-22399758>

32. World Health Organization. *Report on Dengue*. *Report of the Scientific Working Group meeting on Dengue* (2007). doi:10.1111/j.1556-4029.2009.01139.x

33. Meloun, M. & Militký, J. *Statistická analýza experimentálních dat*. (Academia, 2004). at <http://meloun.upce.cz/docs/books/ucebnice-sken.pdf>

34. Broemeling, L. D. *Bayesian Methods in Epidemiology*. *Bayesian Methods in Epidemiology* (CRC Press, 2014). doi:10.1007/978-0-387-09834-0

35. Lawson, A. B. *Statistical Methods in Spatial Epidemiology*. (John Wiley & Sons, 2013). at < https://goo.gl/gNFRwo>

36. Chatfield, C. *The Analysis of Time Series: An Introduction*. (CRC Press, 2004). doi:10.1007/978-1-4899-2921-1

37. Derouich, M., Boutayeb, A. & Twizell, E. H. A model of dengue fever. *Biomed. Eng. Online* **2,** 10 (2003).

38. Shepard, D. S., Undurraga, E. A. & Halasa, Y. A. Economic and Disease Burden of Dengue in Southeast Asia. *PLoS Negl. Trop. Dis.* **7,** 1–12 (2013).

39. Chang, A. Y., Fuller, D. O., Carrasquillo, O. & Beier, J. C. Social Justice, Climate Change, and Dengue. *Health Hum. Rights* **16,** 93–104 (2014).

40. Ooi, E. E. & Gubler, D. J. Dengue in Southeast Asia: epidemiological characteristics and strategic challenges in disease prevention. *Cad. Saude Publica* **25,** 115–124 (2009).

41. Morin, C. W., Comrie, A. C. & Ernst, K. Climate and dengue transmission: Evidence and implications. *Environ. Health Perspect.* **121,** 1264–1272 (2013).

42. *Viral Infections and Global change*. (John Wiley & Sons Press, 2014). doi:10.1002/9781118297469

43. Hagenlocher, M., Delmelle, E., Casas, I. & Kienberger, S. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. *Int. J. Health Geogr.* **12,** 36 (2013).

44. Arcari, P., Tapper, N. & Pfueller, S. Regional variability in relationships between climate and dengue/DHF in Indonesia. *Singap. J. Trop. Geogr.* **28,** 251–272 (2007).

45. Tipayamongkholgul, M., Fang, C.-T., Klinchan, S., Liu, C.-M. & King, C.-C. Effects of the El Niño-southern oscillation on dengue epidemics in Thailand, 1996-2005. *BMC Public Health* **9,** 1–15 (2009).

46. Wu, P. *et al.* Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Sci. Total Environ.* **407,** 2224–2233 (2009).

47. Hii, Y. L. *et al.* Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Glob. Health Action* **2,** (2009).

48. Cummings, D. a T. *et al.* The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS Med.* **6,** 1–10 (2009).

49. Wu, P. C., Guo, H. R., Lung, S. C., Lin, C. Y. & Su, H. J. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop.* **103,** 50–57 (2007).

50. Yu, H., Yang, S., Yen, H. & Christakos, G. A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. *Stoch Env. Res Risk Assess* **25,** 485–494 (2011).

51. Xuan, L. T. T., Hau, P. Van, Thu, D. T. & Toan, T. D. Estimates of meteorological variability in association with dengue cases in a coastal city in northern Vietnam: an ecological study. *Glob. Health Action* **7,** 1–7 (2014).

52. Pham, D. N. *et al.* An Efficient Method To Predict Dengue Ooutbreaks In Kuala Lumpur. in *Proceeding of the 3rd International Conference on Artificial Intelligence and Computer Science (AICS2015)* (eds. Basri, M., Ahmad, S. B. & Rahman, K.) 169–178 (World Conference Resources, 2015). at <http://goo.gl/hG8psW>

53. Bravo, L., Roque, V. G., Brett, J., Dizon, R. & L'Azou, M. Epidemiology of Dengue Disease in the Philippines (2000-2011): A Systematic Literature Review. *PLoS Negl. Trop. Dis.* **8,** 1–11 (2014).

54. Philippine Athmospheric Geophysical and Astronomical Services Administration. Climate of the Philippines. (2015). at <https://web.pagasa.dost.gov.ph/index.php/climate-of-the-philippines>

55. UN Commission on Sustainable Development. *Indicators of Sustainable Development: Guidelines and Methodologies*. (2000). at <http://goo.gl/nO9Uv8>

56. Philippine Statistics Authority. *Population and Annual Growth Rates for The Philippines and Its Regions, Provinces, and Highly Urbanized Cities Based on 1990, 2000, and 2010 Censuses*. *2010 Population and Housing Census* (2010). at <https://goo.gl/KAsCvk>

57. US Department of Commerce, Administration National Oceanic and Atmosperic & Service National Weather. Dew Point. *Glossary* (2009). at <http://goo.gl/jiTm5K>

58. Alduchov, O. A. & Eskridge, R. E. Improved Magnus Form Approximation of Saturation Vapor Pressure. *J. Appl. Meteorol.* **35,** 601–609 (1996).

59. Chen, Z., Haykin, S., Eggermont, J. J. & Becker, S. *Correlative Learning: A Basis for Brain and Adaptive Systems*. (John Wiley & Sons, Inc, 2007). doi:10.1002/9780470171455

60. UCLA: Statistical Consulting Group. R Data Analysis Examples: Poisson Regression. *Institute for Digital Research and Education* (2015). at <http://goo.gl/o026hf>

# LIST OF FIGURES

# LIST OF TABLES