# Bioinformatics

Group 4.1
Summerterm 2023
Molecular Biotechnology 4. Semester

Noemi Luana Tigrato, Lea Mrowiec, Paula Elisabeth Mostert & Manon Merle Mandernach

# Table of Contents

## I Introduction

Dengue virus is the leading cause of arthropod-borne viral disease in the world. According to the World Health Organization around 3.9 billion people are at risk of infection with dengue viruses all over the world with the global burden only rising in the last several years (WHO, 2023, Bhatt, 2013). Dengue viruses are single-stranded RNA viruses, of which 4 different serotypes exist: DENV 1-4. They are transmitted through the female *Aedes* mosquitos from one human to the next. An infection with dengue viruses leads to Dengue fever, the severity of which can range from mild or no symptoms to being fatal. Typical symptoms include high fever, a sore throat, nausea, muscle and joint pain and skin rashes (Schaefer, 2023).

The *Aedes* mosquitos are typically found in tropical or subtropical regions where environmental factors are favorable. The mosquitos can survive as long as the temperature remains above 4°C, however they require a temperature between 15-37°C to complete the life cycle which takes between 7-10 days from eggs to adult mosquitoes. The mosquitoes need stagnant water pools for the eggs to develop into larvae and adults (CDC, 2021). With this basic understanding of the vector, the complex relationship between weather and dengue transmission becomes clear. The transmission is strongly influenced by different environmental factors such as temperature, rainfall, and relative humidity. Relative Humidity can be explained as the water vapor present in the air divided by the maximum amount of water vapor the air can hold at a certain temperature before it starts to condensate. As well as factors related to the human host like population density and movement dynamics (Campbell et al., (2013)). An increase in temperature is associated with a decrease in incubation rate and increase in biting rate. A higher precipitation rate is associated with on the one hand more breeding possibilities but on the other hand if the rate is too high the larvae can be destroyed by the heavy rainfall. Higher altitudes might influence the transmission as well, as higher altitudes are associated with overall lower temperatures (Lai, 2018). In Thailand which is the main focus of this data analysis project the climatic variables are strongly influenced by the different monsoon seasons, the south-west monsoon and the north-west monsoon. The formation of these monsoon seasons is a result of seasonal temperature differences between the coast and the sea, which causes seasonal changes in wind-direction and characteristic weather during each season (SciJinks, 2023). With climate change creating more favorable conditions for the mosquito vectors. Investigating and understanding this complicated net of factors is an important step to understanding what scenarios cause especially high epidemic outbreaks. This is vital as the resources are limited and a forecasting system could greatly improve how epidemics are dealt with.

Therefore, in this data analysis project the main focus is the relationship between the relative humidity variable and dengue cases in Thailand. Can geographical and temporal clusters for Dengue cases and relative humidity be defined? And how do these clusters correlate with one another? A change in relative humidity is part of the characteristics of the different monsoon seasons in Thailand. To analyze this connection, the temporal and spatial clustering of the dengue cases and relative humidity is of interest. The yearly observations can give insight into stronger and cross season factors that may impact the number of dengue cases. The geographic impact in relation to relative humidity and dengue cases is analyzed. Dengue cases are also associated with population density and urban areas as the human to mosquito to human transmission is simplified by the closeness and number of humans that mosquitos can infect. Could this positive correlation between population density and dengue cases, as suggested in the papers mentioned above, be replicated?

# II Material and Methods:

## MATERIAL:

The foundation of the dataset of relative humidity used is the global atmospheric dataset ERA5. It is the fifth generation of a global reanalysis model combining model data with data based on observation from across the world resulting in a complete and consistent data set. In our case only the data of relative humidity between 2006 and 2020 was used consisting of monthly mean values for each province (ERA5, 2023). The data of the dengue cases was extracted from different reports and documents of the Thai government. This dataset provided monthly data of Dengue cases per province between the years of 2006 and 2020. This data was normalized with population data, which originate from the Statistical Yearbooks of the National Statistical Office of Thailand. To create geoplots for visualization of the variables of interest, relative humidity and Dengue cases, myltipolygon data was used forming the borders of each Thai province in the maps (provided by Supervisor (Dafka, 2023)). Moreover, to analyze the influence of population density on the number of dengue cases additional information by the National Statistics Office was consulted (N.S.O., 2021).

## METHODS:

### *Data Cleanup and Preparation:*

The Dengue data used in this analysis was first cleaned and prepared to create different data frames for the following analysis. The data cleaning was comprised of sorting the different excel tables for all the Dengue data frames from the years 2006 to 2020. The data frames were adjusted to one common format. The Dengue data was standardized to the population data and is measured in the unit Dengue cases per 100 000 persons.

To work with the data a combined data frame of the Dengue cases was used. The combined data frame holds the Dengue cases and as columns the “Reporting Area”, “Year”, “Month”, “Dengue Cases” and “Relative Humidity” in a wide format. For correlation tests this combined data was converted into a long format with the same columns. Reporting areas and Provinces can be used interchangeably here, as they contain the same variables. Moreover, another long format data frame with the columns “Province”, “Year”, “Seasons”, “Dengue Cases” and “Relative Humidity” was needed to look at the correlation between different monsoon seasons.

The four monsoon seasons were defined as follows: inter-monsoon season from December of the previous year to February of the current year, pre-monsoon season from March to May of the current year. The monsoon season from June to September of the current year and the post-monsoon season was defined from October to November of the current year. To sort the Dengue case data and the relative humidity data into the respective monsoon seasons the mean() function was used and NA’s were excluded in the calculation of the mean. For some parts of the analysis working with NA's turned out to be difficult so a separate data frame without NA's was constructed. The NA's in the original wide format data frame can be divided into two different types. NA's where the data for entire years in a certain province is missing and single NA's for one or up to three months within a year and province. For NA's of the first type, the missing values were simply replaced by the mean of the year and month. This method was applied to the province Kalasin from 2008 to 2009 and from 2011 to 2013 and for the province of Bungkan from 2006 to 2011. For NA's of the second type, the total cases for each year were given, and NA's were replaced by a simulated value. This process is further described in the Markdown document. This method might not accurately reflect the actual Dengue cases in all instances, but it can be seen as an estimation.

To be able to investigate the impact of population density one final data frame was which contains the columns "Province", "Year", "Dengue cases", "Population density" and "Relative humidity". As the population data is only available on a yearly basis and it is fair to assume that the population does not change significantly within a year, this data frame was based on yearly observations rather than monthly or seasonal. To calculate the population density the population of each year in each province was divided by the area of the province measured in square kilometers. The Dengue column contains the yearly total of the Dengue cases of each province and the humidity column contains the mean yearly relative humidity of each year and province. Data frames were adjusted as needed for different steps of the analysis.

### *Clustering:*

Clustering was used in the beginning of the analysis to break down the complex dimensions of both the Dengue data frame as well as the relative humidity data frame. It is a technique which is used to group similar data points together based on their similarities. It is useful to discover certain subgroups, patterns, or structures in the data. The R algorithm kmeans() randomly selects K points as initial cluster centroids to which each data point is assigned, based on the Euclidian distance. The algorithm then iteratively recalculates and reassigns the centroids of each cluster until the centroids no longer change significantly (Hermann, 2022). In a broader use, clustering can be used as a method to divide a data set in K numbers of Clusters, not depending on the amount of data points per cluster but the difference between the K Centroids. For this analysis, it was decided to focus on K= 3 to divide in "Low", "Medium" and "High" Dengue and humidity clusters.

### *Spatiotemporal Analysis:*

Spatiotemporal analysis can be defined as time-series analysis on a geographical level (Columbia, 2023). This kind of analysis is crucial to understand the emergence and spread of disease within a country as well as defining high-risk and low-risk areas via clustering. Moreover, by working with data combining temporal and geographical dependencies it is possible to understand the impact of seasonal weather events on the spread of a disease.

### *Geoplots:*

Mapping was used as an important and fundamental step of data visualization. By combining the geo data containing multipolygons defining the area of each province of Thailand with the disease or climate data different maps could be created. To do so, the function geom_sf() of the R package ggplot2 was used. This enabled clear visualization of characteristics and trends of Dengue cases and relative humidity. By using the parameter time on a monthly, yearly, and seasonal (pre-monsoon, monsoon, post-monsoon and inter-monsoon) bases different factors influencing the variables of interest (relative humidity and Dengue cases) can be examined (Columbia, 2023).

*Correlation test and linear models:*

For the correlation tests a quick overview of the variables was performed. The pearson correlation coefficient r was calculated with the cor() function for non-lagged relationships between the variables and lags of a month up to three months and for the seasonal data up to one season. The lag was assumed to adjust for the time the mosquitoes need to develop. Linear models are used to model the relationship between a dependent variable and one or more independent variables. It can help to understand how changes in the independent variables are associated with changes in the dependent variable. The dengue cases were seen as the dependent variable while the relative humidity and population density were seen as independent variables. A linear model was chosen as the normalized dengue cases were no longer a discrete variable but rather a continuous one.

*ARIMA:*

Autoregressive Integrated Moving Average (ARIMA) is the technique used for analyzing the dengue time series data. It is a tool that combines autoregressive (AR), integrated (I), and moving average (MA) components to capture trends and patterns in the data. Also, the dependent variable will be referred to as X and the predictor variable as Y.

ARIMA models learn from past patterns and predict future outcomes based on those patterns. In this analysis ARIMA was used to predict the Dengue fever cases for a period from January 2021 until December 2023. The AR (p) component considers the relationship between an observation and its past values. Therefore, p refers to the number of lags of Y used as predictors. The Integrated (d) component makes the data stationary by removing trends and d is the number of differencing required to make the data stationary. The MA (q) component captures short-term fluctuations and noise in the data and therefore q is the number of lagged forecast errors that should go into the ARIMA model (Shumway, 2017).

The ARIMA Model was chosen to complete the analysis, since it is specifically designed for a time series analysis. A GAM model was also considered however ARIMA provides an effective method to capture seasonal pattern present in the data. The use of ARIMA for modelling infectious disease cases such as dengue fever, as in the analysis Nayak et al (Nakay et al., 2019). The data used for the ARIMA model were the Dengue cases between the years 2006 to 2020 in Thailand and an analysis on the total Dengue cases was performed. Therefore, the monthly total values were calculated. The data frame then was converted into a time series.

# III Results

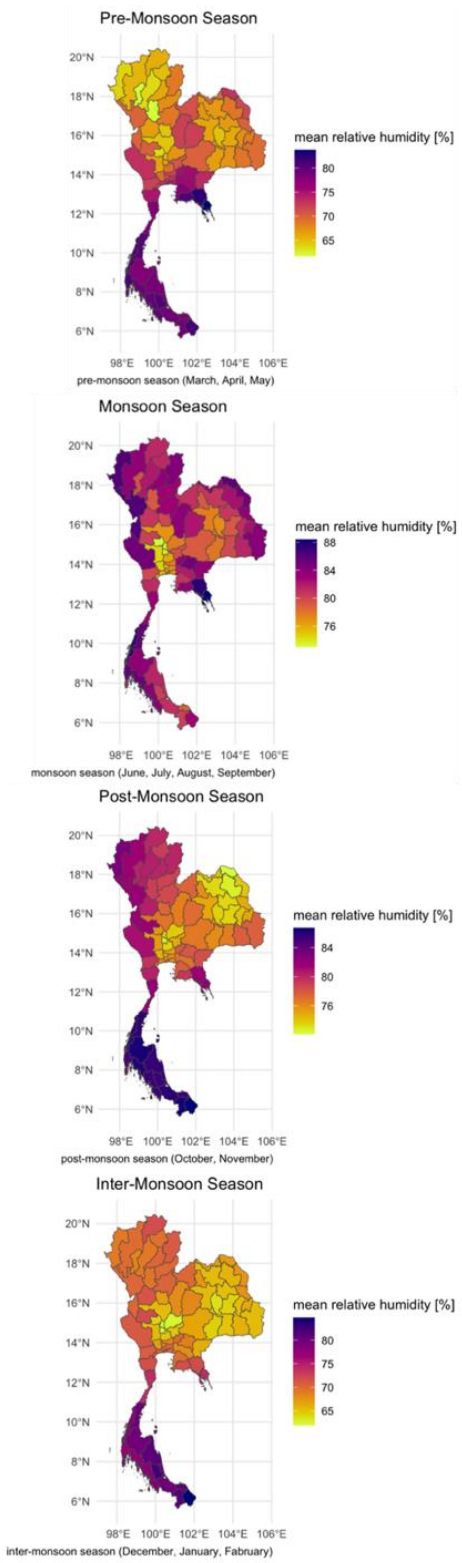



*Figure 1: Mean relative humidity over all analyzed years (2006 to 2020) in each season*

## Clustering

### *Monsoon Seasons*

To find relations between relative humidity and dengue cases on a geographical level during each season, characteristic patterns of high/low relative humidity and Dengue cases per 100 000 persons had to be identified. This unit applies to all further mentions of Dengue.

### Relative Humidity

During the pre-monsoon season, there was a noticeable increase in mean relative humidity from north to south. Low humidity spots were found in the mid-north, mid-east, and centre of the country. The entire country experienced a rise in relative humidity during the monsoon season, reaching around 89%. The lowest relative humidity, around 71%, was observed in the centre of the country. In the post-monsoon season, there was a general decline in relative humidity from west to east, with high values of up to approximately 87% in the south. The inter-monsoon season showed a similar distribution of high and low relative humidity as the previous season. However, there was a relative humidity plateau during the monsoon season, with declining values in the post-monsoon, pre-monsoon, and inter-monsoon seasons.

### *Dengue cases*

In general, the identification of clear patterns concerning the amount of Dengue cases (Figure.2) in the whole country was not as straight-forward as with the data of relative humidity because the maps showed a more spotted pattern than the plots of relative humidity. The most significant increase in dengue cases could be located in the north from less than four cases per month during **inter-monsoon season** to up to 40 total cases and month during **monsoon-season**. In comparison to that, a more subtle change of the number of cases could be seen in the South of the country. The most even distribution of the amount of Dengue cases could be seen in pre-monsoon season.

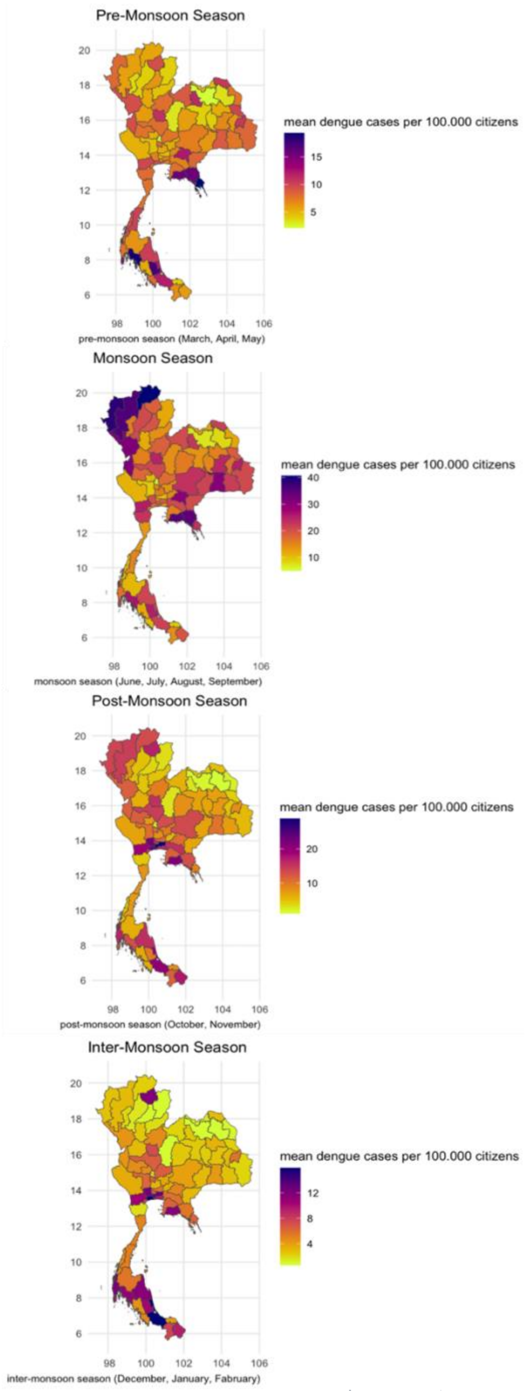



*Figure 2: Mean Dengue cases for each province per 100 000 citizens per month (2006 to 2020)*

### *Geographical*

K-means clustering was used to assign mean humidity and mean Dengue cases to three clusters: "Low," "Medium," and "High." To visualize the cluster assignments and changes over the years, ggPlots were created, showing the yearly centers of the clusters. The centers for each cluster shifted annually, indicating fluctuations. Notably, there was a significant increase in the highest-clustered Dengue cases in 2013 and 2016 (see VI Appendix Figure 4b). Maps were also created to assign provinces to the "Low," "Medium," and "High" clusters for both Dengue and relative humidity. The maps demonstrate how provinces shifted clusters over the years (see VI Appendix Figure 3).

## Geographical Patterns of Relative Humidity and Dengue Cases

### *Different levels above sea level*

To investigate whether the geographical characteristics of different provinces influence the dengue cases, different exemplary provinces were chosen: Maha-Sarakham (northeast, 100-200 masl), Chaing Mai (north to northwest, 1000-1500 masl), Chumphon (south, 0-100 masl) and Songkhla (south, 100-200 masl).
Generally, a similar trend of relative humidity throughout the year was visible, characterized by a peak in September and October followed by a steep decline towards the end of the year. Nevertheless, the mean relative humidity of those provinces differed. Regarding the Dengue cases, a general peak was identified between May and August. (Appendix Figure. 13)

### *Provinces with Many Dengue Cases (Highest Cluster)*

Provinces with a high amount of Dengue cases from the highest cluster were mostly located near the Gulf of Thailand. In those provinces a greater difference between the highest and lowest relative humidity could be observed (e.g. in Surat Thani and Songkhla).(see VI Appendix Figure.14)
With further analysis, a decline of Dengue cases after relative humidity reachhed about 80% could be observed in Kalasin (see Appendix Figure 5).

## Correlation of Dengue and Humidity

To visualize and analyse the relation between Dengue and humidity in specific geographic provinces, plots with both monthly means of humidity in % and Dengue cases were plotted for different provinces overall years from 2006 to 2020, which are representatives for different levels above sea level or which are influenced by the proximity to the gulf of Thailand. The correlation between yearly Dengue and humidity was calculated and show significantly high results.(see VI Appendix Figure 5)

The correlation coefficient between the relative humidity and dengue cases over all years is 0.242 while the r-squared value returned a value of around 0.06 indicating that only around 6% of the total variance in the Dengue cases can be explained by the relative humidity. If the Dengue data is lagged compared to the relative humidity (January humidity is compared to February dengue cases) an overall negative trend in the correlation could be observed. The higher the lag the lower the calculated r-value and r-squared. In tune with these findings the correlation between the relative humidity and Dengue cases in monsoon seasons returned an r value of 0.21 and an r-squared value of 0.043 indicating a weak positive correlation between the two variables even sorted into the monsoon seasons, however the linear model could only explain around 4.3% of the total variance of the dengue cases. The lagged correlation of one season returned a smaller, but also negative r-value of -0.131 and the linear model could only explain 1.7% of the variance.
With r = 0.062 a weak positive relationship between the population density and the dengue cases could be observed. However, the corresponding linear model was only able to explain less than 1% of the variance in the data set. Even when applying a multiple linear regression model to look at the influence of population density and relative humidity on the dengue cases at the same time the model was not able to explain more than 1% of the variance.
It is important to note that for all the linear models the p-value was consistently lower than 0.05, indicating that however small the correlation might be, it is still significantly present.

## ARIMA

The first step to get a better understanding for the data was to perform a decomposition of a periodic time series (see VI Appendix Figure.10 for complete figure). An upward trend was identified with peaks in the years 2013 and 2015. In Fig.3 the ARIMA model was applied in Thailand for forecasting of monthly total dengue fever cases not normalized. The forecasted values include the years 2021 to 2022.
The process of ARIMA modeling can be divided into four steps, data preprocessing, model identification, estimation of values and model evaluation. The data preprocessing step is already described in III Methods p.(6).

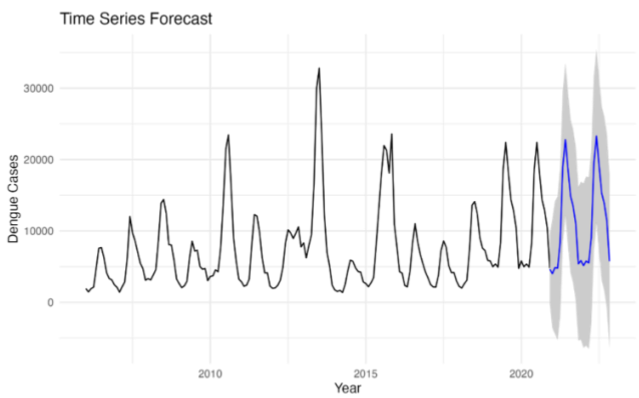



*Figure 3: ARIMA model of Dengue cases in Thailand between 2006 and 2020. Forecast of Dengue cases for the years 2021 and 2022*

## Model Specification:

In the identification step, the optimal values of the AR, I, and MA parameters were determined using the Augmented Dickey-Fuller test (ADF), autocorrelation (ACF) and partial autocorrelation (PACF) plots.
(A) Ensuring the stationarity – Integrated term (d)
The ADF test was applied to test the monthly total dengue cases between the years 2006 – 2020 for stationarity. The order of differencing is selected so that the standard deviation is minimalized. The test indicated the data to be stationary. Therefore, no differencing was needed for the model.
(B) Identification of AR(p) and MA(q) terms

The ACF and PACF were used to determine the order of the AR (p) and MA (q) terms (see VI Appendix Figure.11, Figure.12 for complete figure). The ACF plot shows the correlation coefficient between a time series and lags of itself, and the PACF plot depicts the partial correlation coefficient. It was found that the ACF plot a showed seasonal pattern with a frequency of twelve months. The plot was not cut-off sharply however the values were not consistent within the limit (0.2). Therefore, an AR term was added. The PACF plot, however, showed a sharp cutoff and was also negative. Hence, an MA term was added.

### Estimation of values:

The model was then estimated by fitting it to the historical data and estimating the coefficients. The Auto-ARIMA function was mainly used to create a model, but also to confirm the beforehand identified model specifications. Due to the seasonal pattern in the dengue cases data, the seasonal parameter was set on TRUE, to check for the best seasonal ARIMA (SARIMA) model specifications.

The best fitted seasonal ARIMA model for dengue cases is found to be ARIMA (1,0,2) $(1,1,0)_{12}$ with drift. This confirms our beforehand predicted order of the Model. There is one autoregressive term (p=1), no differencing term (d=0) and two moving average terms (q=2). To capture the seasonal pattern the order of the SARIMA was one autoregressive term (P=1), one differencing term (D=1) and no moving average term (Q=0) with a seasonal period of twelve. The drift term is needed to account for the constant trend in the data.

### Model Evaluation:

The parameters sigma$^2$ and log likelihood were used to evaluate the ARIMA model. The value of sigma$^2$ was 596 2125, this represents the variability of the residuals. The residuals are defined as the differences between the predicted and original values. The log likelihood measure indicates how well the ARIMA model fits the observed data. The value of the log likelihood was -1548.86. However, high log likelihood value indicates a better fit. Yet, the parameters only give an insight into the goodness of the model and are not significant enough without comparison to a second model. Thus, a statistical and visual analysis of the residuals was also performed. The model’s performance was validated by comparing the predicted values for the years 2019 and 2020 with the original values. On average the forecasted values deviate from the actual case number by approximately 1324.6 cases. The typical magnitude of the forecast errors was 1412.95 cases. And the average difference between the predicted values and the actual values was 12.44 %.

# IV Discussion:

Having analysed Dengue cases and relative humidity in Thailand on a temporal and geographical level, the following findings can be reported: The geographic shift of clusters was not as clear visible as expected. Furthermore, the visual correlation between the clusters could not be confirmed. Dengue cases and humidity seem to be influenced by their geographical characteristics. Additionally, a slight but significant positive correlation could be seen between Dengue and relative humidity overall. ARIMA modelling was able to model the Dengue cases accurately to an extent.

## Clustering

According to Singhrattna et.al yearly variations of relative humidity could be caused by the Walker-Circulation and ENSO due to its quantitatively negative impact on rainfall etc (Singhrattna, 2005). Moreover, the presence of monsoon-seasons could be reconstructed by using relative humidity defining characteristic patterns for each season. Also, as Xu et.al have shown, high risk clusters of Dengue vary annually and peak during monsoon-season predominantly in the northeast of Thailand (Xu et al., 2019). The temporal connection of overall higher Dengue cases and relative humidity during monsoon-season can be explained by the already observed connection of higher relative humidity and an increased daily biting rate of *Aedes aegypty* (Khedari, 2002).

## Geographical Patterns of Relative Humidity and Dengue Cases

The geography of Chiang Mai, despite its elevation of 1000 to 1500 meters above sea level, features a valley that leads to higher humidity due to the convergence of moist air masses from the Gulf of Thailand and the Indian Ocean with the surrounding mountains. This phenomenon results in heavy rainfall during the monsoon season, which contributes to Chiang Mai's relatively higher Dengue cases compared to the national average. Contrary to the assumption that provinces at higher elevations would have lower humidity, Chiang Mai's unique geography defies this expectation. Provinces situated along Thailand's southeastern coast, such as Chumphon and Nakhon Si Thammarat (0-100 meters above sea level), also experience high levels of humidity. The peak in Dengue cases in these areas can be attributed to the proliferation of breeding sites for Aedes aegypti mosquitoes, known to thrive in stagnant water (Getachew, 2015). However, research by M. Osmana et.al suggests that severe sequential flooding and drying can decrease the number of Dengue cases by reducing the survival rate of mosquito larvae (Seidahmed, 2016). This pattern holds true for provinces along the northeast coast, influenced by the Gulf of Thailand.Especially high dengue cases in Thailand overall and in Provinces in the North (Chiang Mai and Maha Sarakham) could be observed in 2013 as shown in the graphs. One Reason how this high outbreak of Dengue cases can be explained, is that Thailand has suffers severe flooding in 2013 (Davies, 2013) and the proximity of the provinces to the gulf of Thailand.

In general, provinces with higher humidity often showed higher Dengue cases (Worldbank, 2023). However, one can observe a decline of Dengue cases once relative humidity has reached around 80%, which might be due to the interaction of the mosquito survival and different climatic factors like relative humidity (Polwiang et al., 2020). An example for this observation is the province Kalasin in the northeast of Thailand. With a mean of relative humidity of 75% Thailand, which is below maximum survival threshold, overall provides favourable conditions for the Dengue vector. When comparing the overall mean relative humidity of Thailand of about 75% to the mentioned value of 80% one can conclude that relative humidity in Thailand is favourable to the mosquitos and distribution of dengue cases.

## Correlation of Dengue and Humidity

Overall, a weak positive relationship between relative humidity and dengue cases could be observed over the years which is in tune with previous findings (Abdullah, 2022). However, the lagged correlation did not improve the relationship as expected. This might be because the lag even at one month is too long to accurately include the life cycle of the mosquitoes, as previous studies have used a lag of only one week on the relative humidity data (Gómez, 2022). Furthermore, the fact that the relative humidity variable depends on both the precipitation and temperature might make it harder to accurately predict its influence. Two very different circumstances can have the same relative humidity, for example a cool but rainy day and a warm and moist day might record the same relative humidity, however for the vectors, these are very different conditions. Higher temperatures are associated with a shorter intrinsic incubation period and therefore more dengue cases, but only up to a certain breaking point where the mosquitoes are no longer able to survive. And precipitation also has a positive correlation but only as long as the rain is not too strong and destroys the eggs and larvae (Abdullah, 2022). The variance in the correlation that could be seen when looking at individual provinces further suggests what the linear regression model indicates as well; looking at only one of these factors is not sufficient to accurately estimate the dengue cases.

The population density had a very weak positive correlation to the number of dengue cases. One reason for this might be that dengue viruses cannot be directly transmitted from one human to the next. Therefore, this correlation would probably be more clearly observed in diseases transmitted via aerosols. Previous studies also suggest that the dengue risk is higher in rural and pre-urban areas as cities usually supply their water through pipes and therefore avoid stagnant water pools whereas they are more common in rural areas. (Schmidt, 2011)

## ARIMA

Even though the sigma $^2$ and the log likelihood value indicated that the goodness of the model is not sufficient. A complete evaluation of the model would only be possible with a comparison to another model and because of the low average difference, we decided to proceed with a forecast of the Dengue cases for the years 2021 to 2022. In conclusion ARIMA modeling does not account for the influence of other variables on Dengue. Therefore, the influence of unforeseen extreme weather conditions or population fluctuations are not covered by ARIMA. This limitation could be alleviated by using a nonlinear multiple regression analysis and forecasting.

## Conclusion:

Due to contradictory observations regarding the connection of relative humidity and dengue cases as mentioned in the results, one cannot conclude that higher humidity automatically leads to higher dengue cases. Nevertheless, it can be concluded that the relationship between dengue and humidity seems to be more complex and cannot be understood by the variable relative humidity alone. Other factors like human movement, infrastructure, wealth index and sanitation are also of importance (Polwiang, 2020). Hence, to understand how to accurately predict and prevent dengue outbreaks, more variables must be included and understood.

# V References:

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. Nature 496, 504-507.

Campbell, K.M., Lin, C.D., Iamsirithaworn, S., and Scott, T.W. (2013). The Complex Relationship between Weather and Dengue Virus Transmission in Thailand. Am J Trop Med Hyg 89, 1066-1080.

CDC (2021). Life Cycle of Aedes aegypti and Ae. albopictus Mosquitoes. https://www.cdc.gov/mosquitoes/about/life-cycles/aedes.html.

Columbia (2023). Spatiotemporal Analysis https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis.

Dafka, S. (2023). Geodata.

Davies, R. (2013). Thailand Floods Update September 2013. https://floodlist.com/asia/thailand-floods-september-2013.

Getachew, D., Tekie, H., Gebre-Michael, T., Balkew, M., and Mesfin, A. (2015). Breeding Sites of Aedes aegypti: Potential Dengue Vectors in Dire Dawa, East Ethiopia. Interdiscip Perspect Infect Dis 2015, 706276.

Hermann, C. (2022). Vorlesung: Einführung in die Bioinformatik. Hermann Carl.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023). ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

Khedari, J A.S., Hirunlabh, J. (2002). Thailand climatic zones. Renewable Energy 25, 267-280.

Lai, Y.-H. (2018). The climatic factors affecting dengue fever outbreaks in southern Taiwan: an application of symbolic data analysis. BioMedical Engineering OnLine 17, 148.

Nayak, M.S.D.P., and Narayan, K. (2019). Forecasting Dengue Fever Incidence Using ARIMA Analysis. International Journal of Collaborative Research on Internal Medicine & Public Health 11, 924-932.

Schaefer TJ, P.P., Wolford RW (2023). Dengue Fever. StatPearls. StatPearls Publishing.

Schmidt, W.P., Suzuki, M., Thiem, V.D., White, R.G., Tsuzuki, A., Yoshida, L.M., Yanai, H., Haque, U., Tho le, H., Anh, D.D., and Ariyoshi, K. (2011). Population Density, Water Supply, and the Risk of Dengue Fever in Vietnam: Cohort Study and Spatial Analysis. PLoS Med 8, e1001082.

SciJinks (2023). What Is a Monsoon? https://scijinks.gov/what-is-a-monsoon/#:~:text=Warm%20water%20in%20the%20ocean,that%20can%20last%20for%20months.

Seidahmed, O.M.E., and Eltahir, E.A.B. (2016). A Sequence of Flushing and Drying of Breeding Habitats of Aedes aegypti (L.) Prior to the Low Dengue Season in Singapore. PLOS Neglected Tropical Diseases 10, e0004842.

Shumway, R.H., and Stoffer, D.S. (2017). ARIMA Models. In Time Series Analysis and Its Applications: With R Examples, (Cham: Springer International Publishing), pp. 75-163.

Singhrattna, N., Rajagopalan, B., Kumar, K.K., and Clark, M. (2005). Interannual and Interdecadal Variability of Thailand Summer Monsoon Season. Journal of Climate 18, 1697-1708.

Thailand, N.S.O. (2021). Number of Area by Region and Province: 2012-2021. National Statistical Office Thailand.

Thailand, N.S.O. (2021). Number of Population from Registration by Age Group, Region and Province: 2012 - 2022. National Statistical Office Thailand.

WHO (2023). Dengue and severe dengue. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue#:~:text=About%20half%20of%20the%20world's,urban%20and%20semi%2Durban%20areas.

Xu, Z., Bambrick, H., Yakob, L., Devine, G., Lu, J., Frentiu, F.D., Yang, W., Williams, G., and Hu, W. (2019). Spatiotemporal patterns and climatic drivers of severe dengue in Thailand. Science of The Total Environment 656, 889-901.

# VI Appendix

## 1. Yearly Centers of Humidity and Dengue Clusters

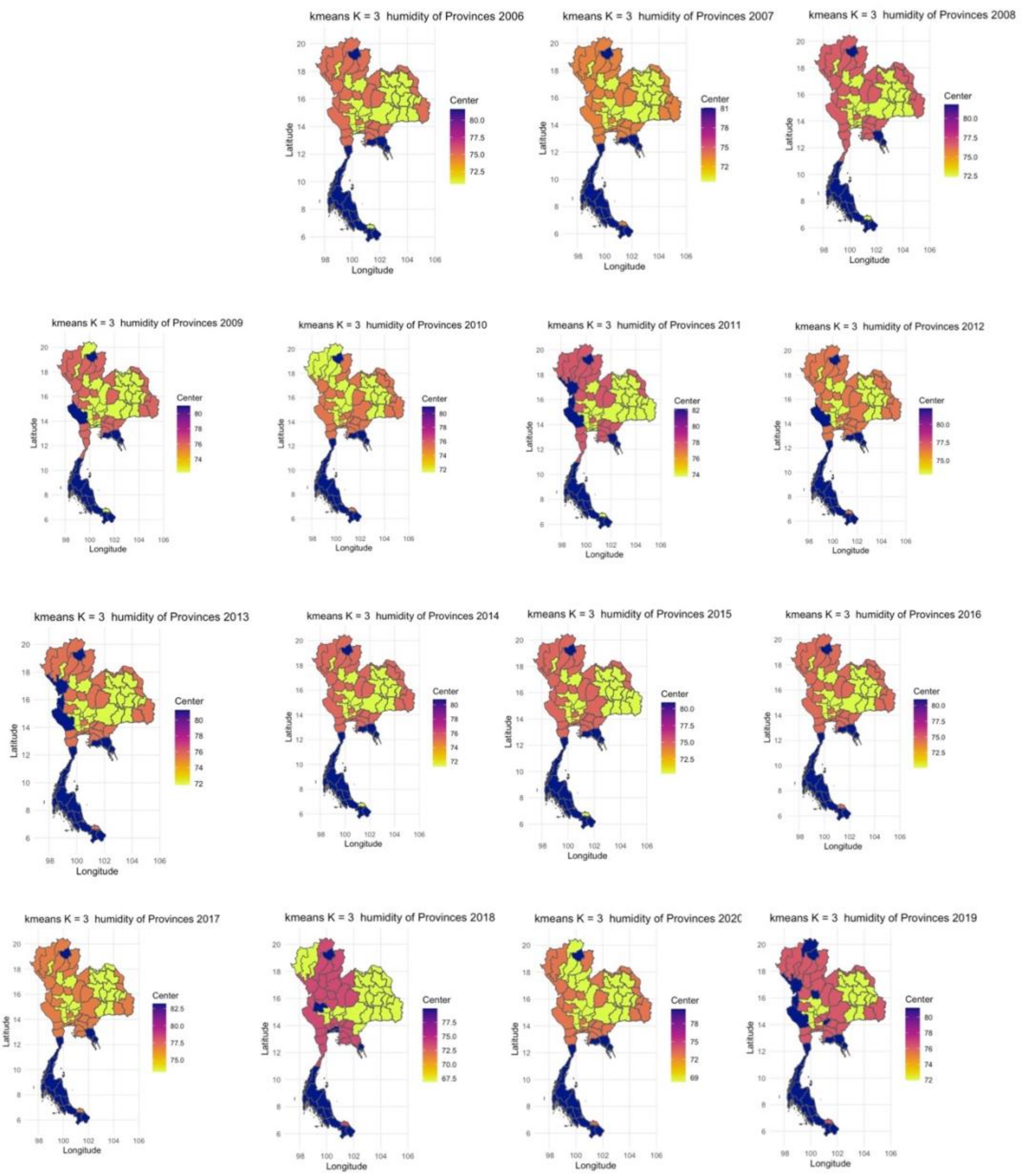



*Figure 3: Maps of the yearly centroids "Low", "Medium" and "High" humidity from 2006 to 2020*

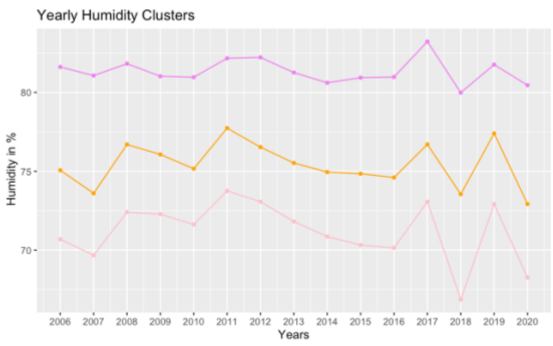



*Figure 4a: yearly relative humidity of the different clusters*

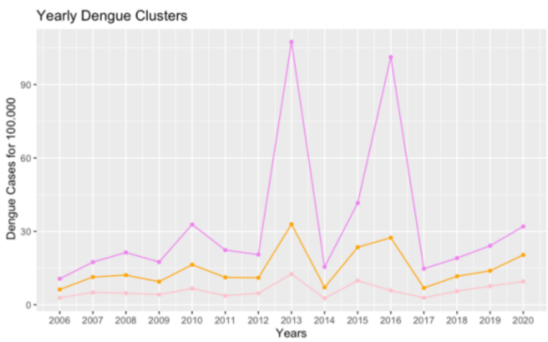



*Figure 4b: yearly dengue cases in the different clusters*

2. **Decrease of dengue cases when Humidity higher than 80% : Example Province Kalasin**

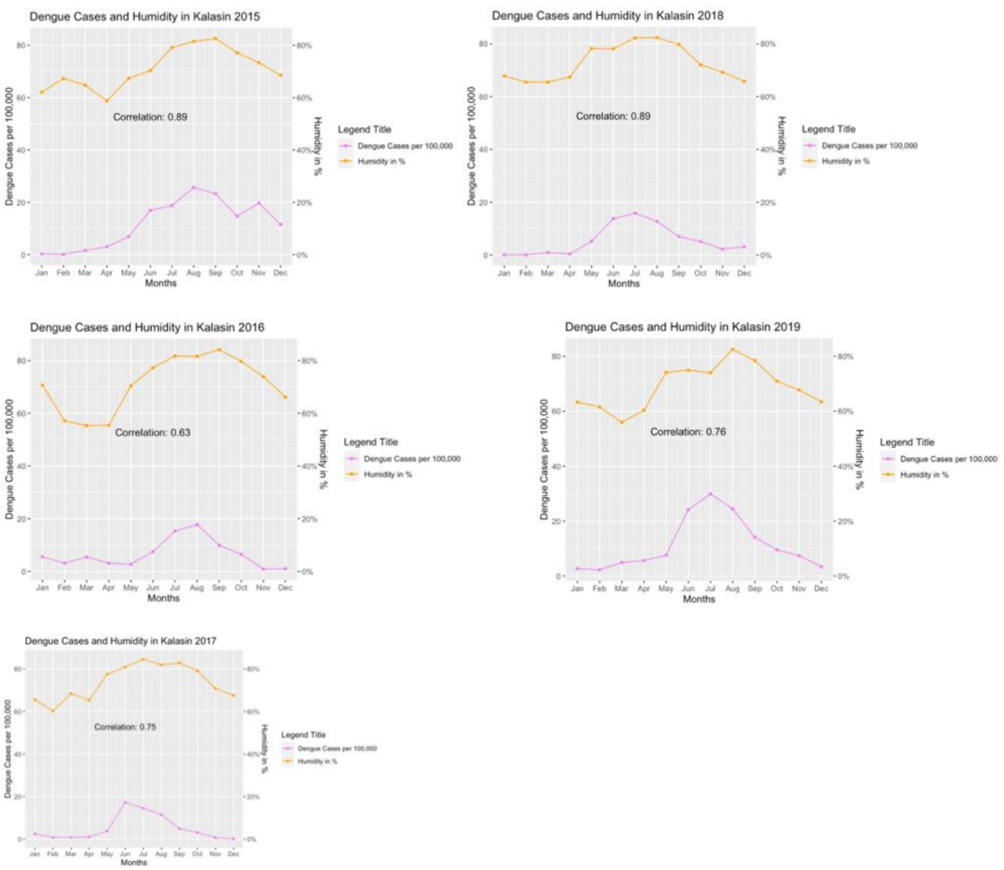



*Figure 5: comparison of Dengue cases in Kalasin in different years*

## 3. Peak of Dengue in 2013 in different Provinces

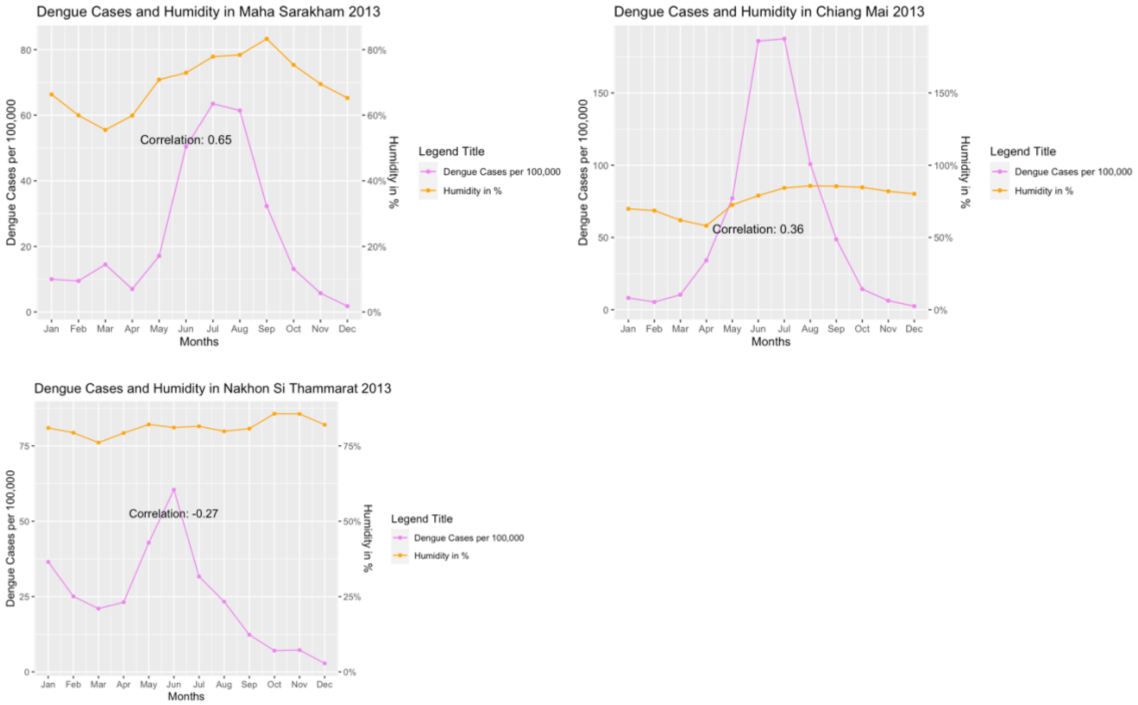



*Figure 6: comparison of Dengue cases in Nakhon Si Thammarat in different years*

## 4. Geomaps of Relative Humidity in the Course of a Year

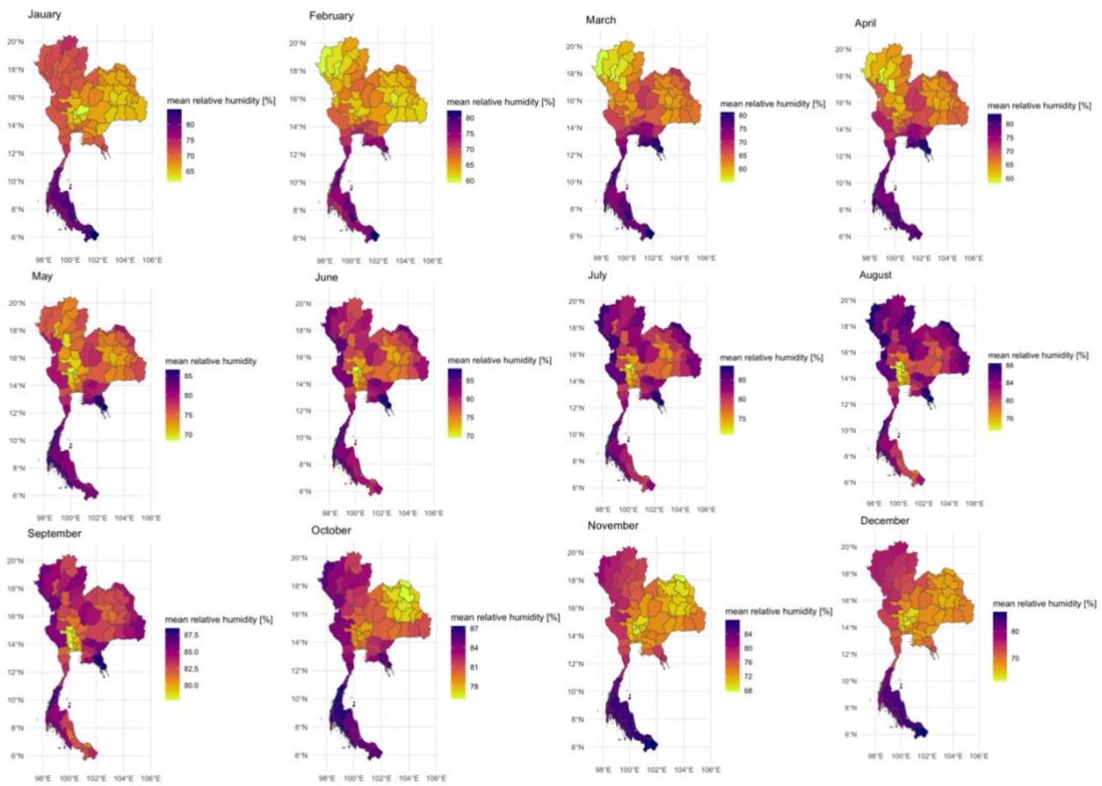



*Figure 7: geoplots of each month of the mean relative humidity*

## 5. Geomaps of Dengue Cases 2006 to 2020

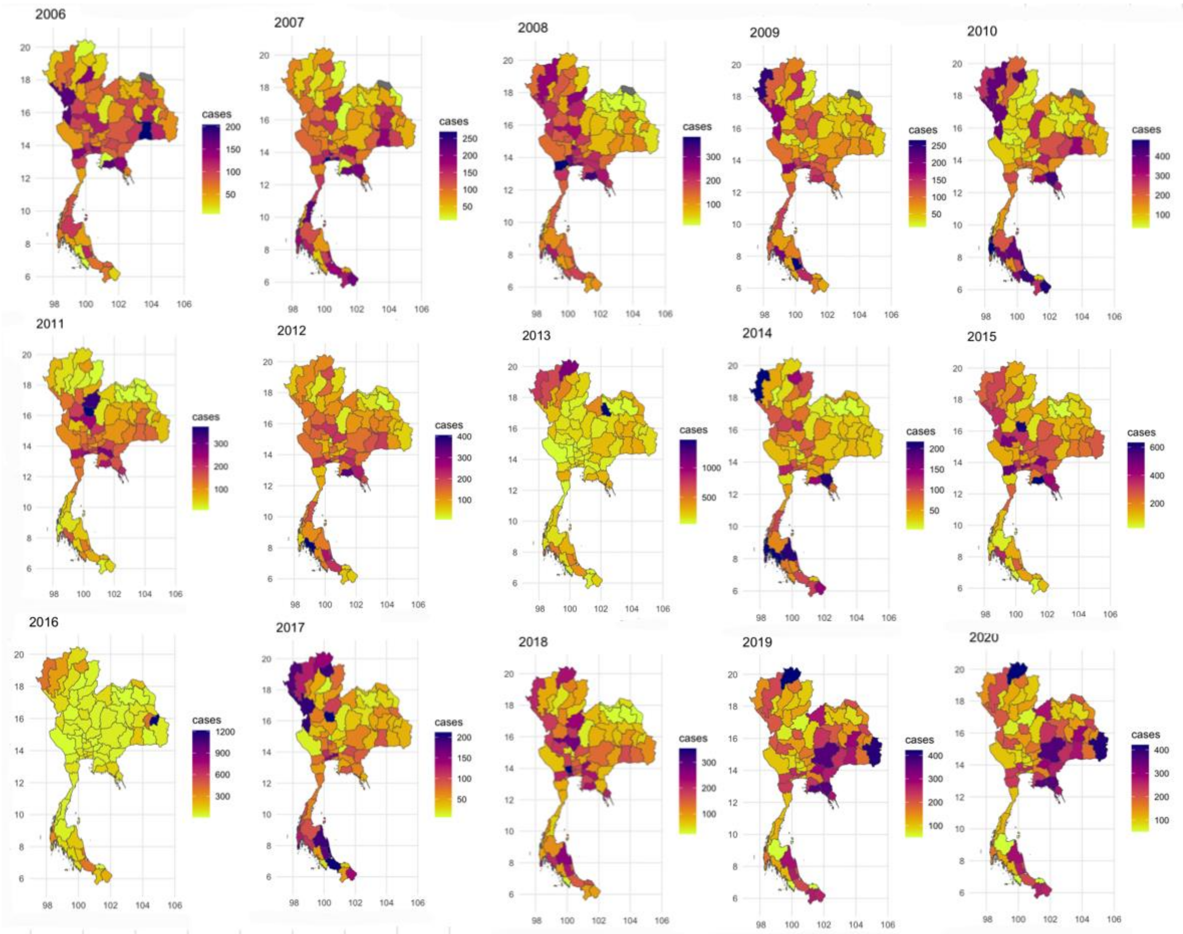
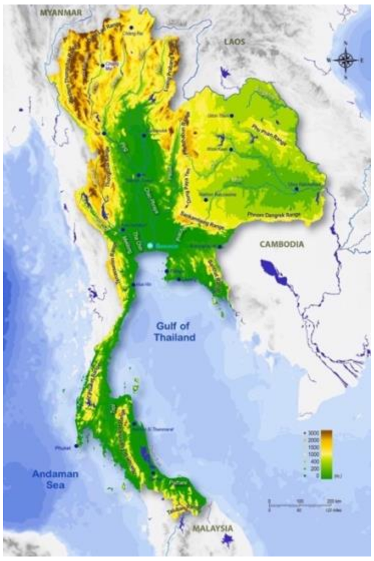



*Figure 7: geoplots of mean monthly Dengue cases througout the timeframe of interest*

## 6. Geographical Map of Thailand and Map Showing the Provinces



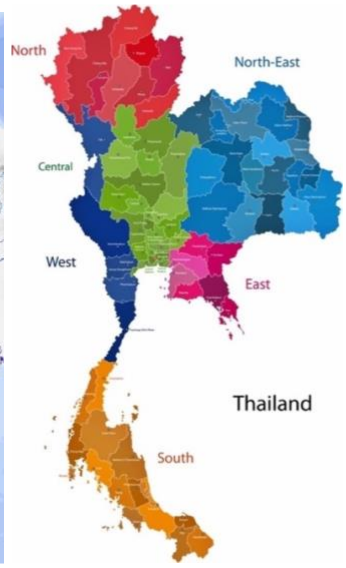




*Figure 8: geographical map of Thailand showing the level of the land above sea level*

*Figure 9: map of Thailand showing the provinces and regions*

## 7. Dengue cases time series – decomposition of periodic time series

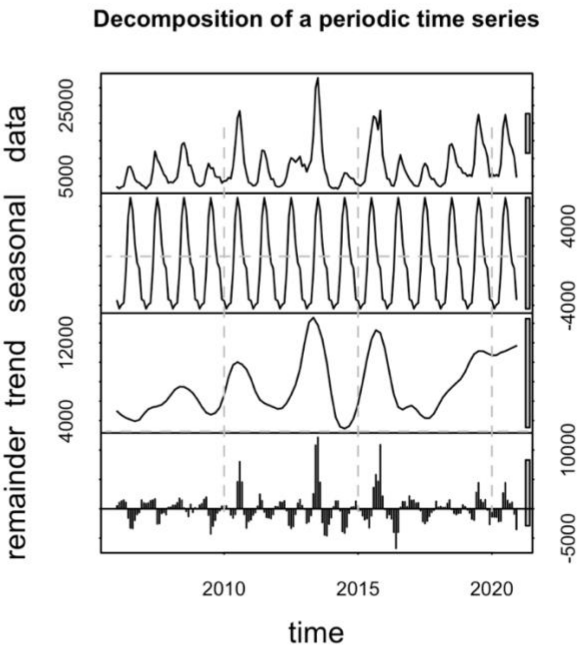



*Figure 10: Decomposition of the periodic time series of dengue cases between the years 2006 and 2020 in Thailand.*

## 8. ACF and PCAF plots of dengue cases time series

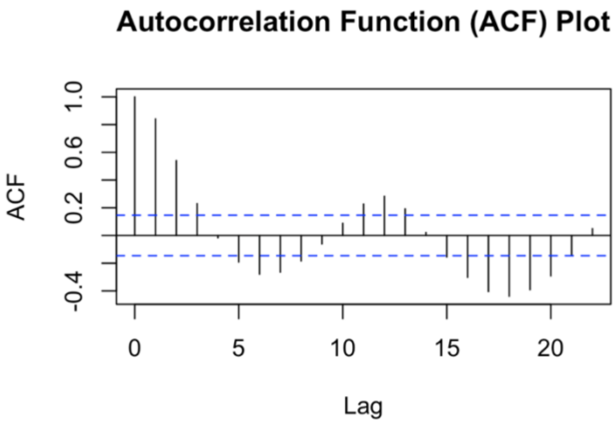



Figure 11: ACF plot of the dengue cases time series between the years 2006 and 2020 in Thailand. The correlation of the time series with itself in regard of a time lag.

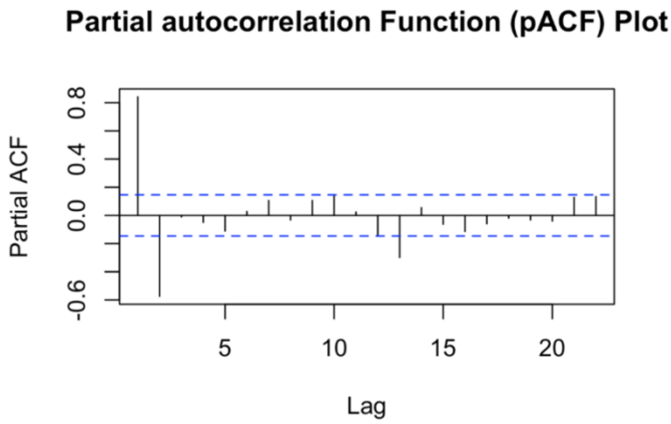



Figure 12: PACF plot of the dengue cases time series between the years 2006 and 2020 in Thailand.

## 9. Exemplary Provinces

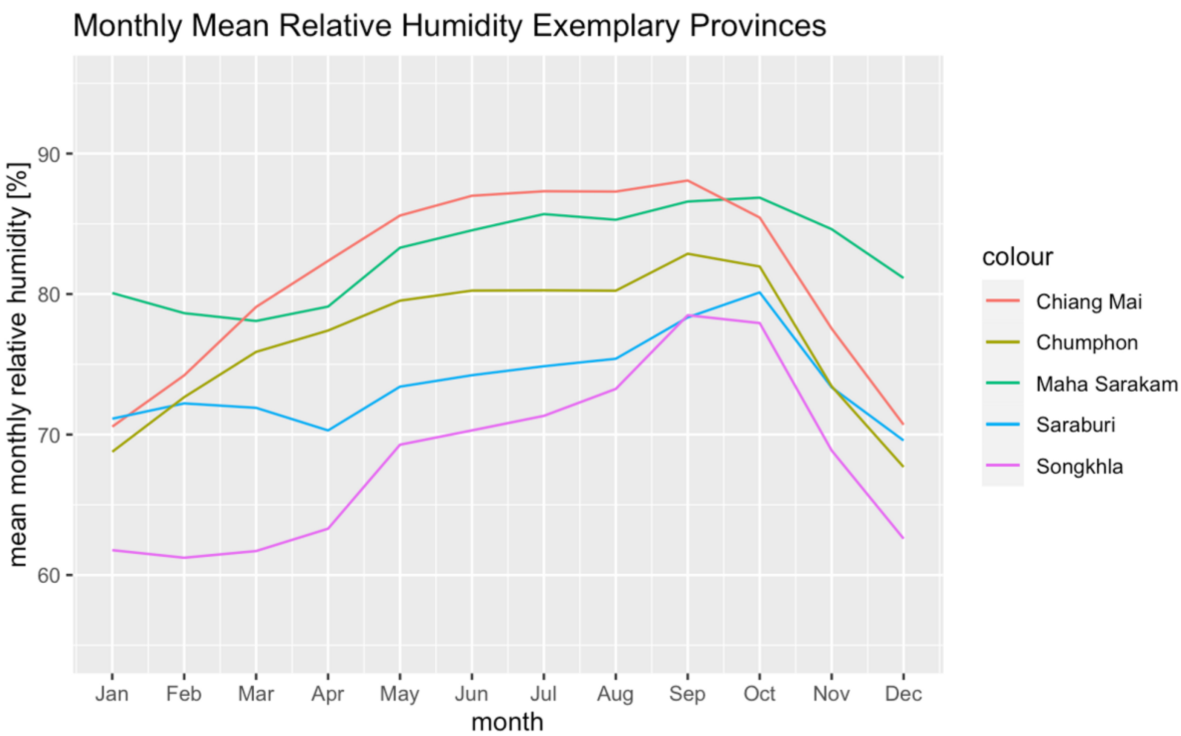



Figure 13: monthly mean relative humidity of exemplary provinces

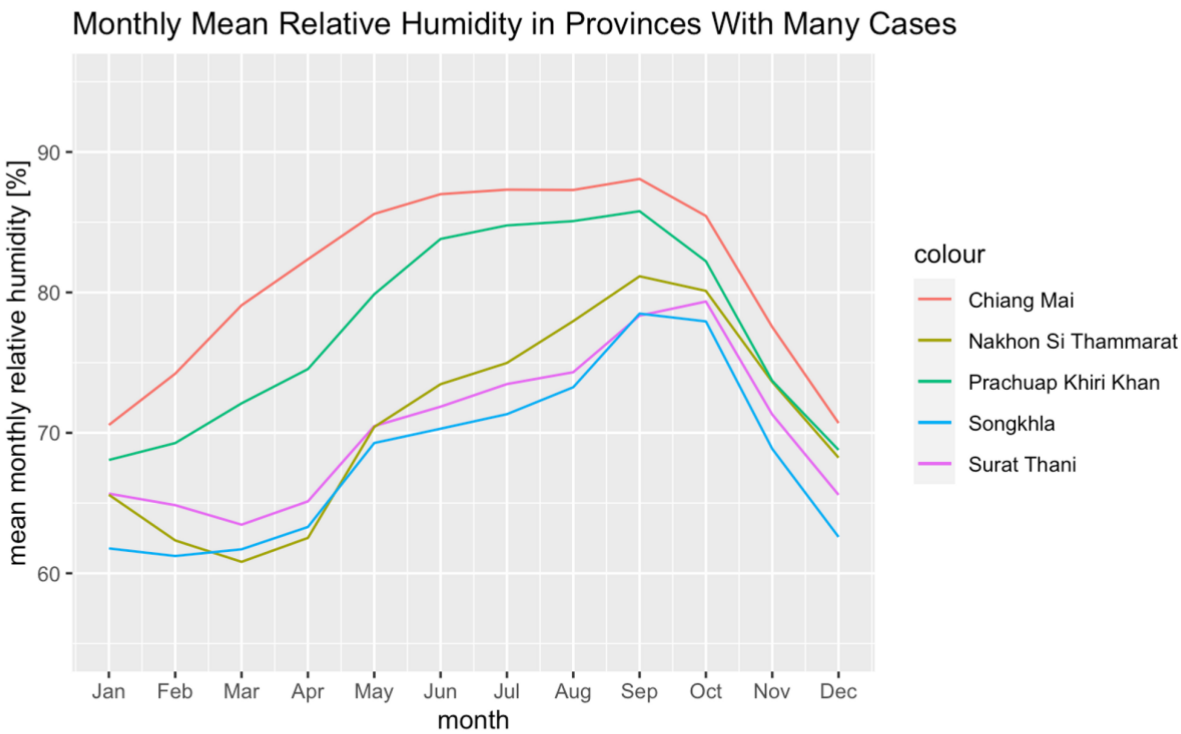



*Figure 14: monthly mean relative humidity in provinces with high Dengue cases*

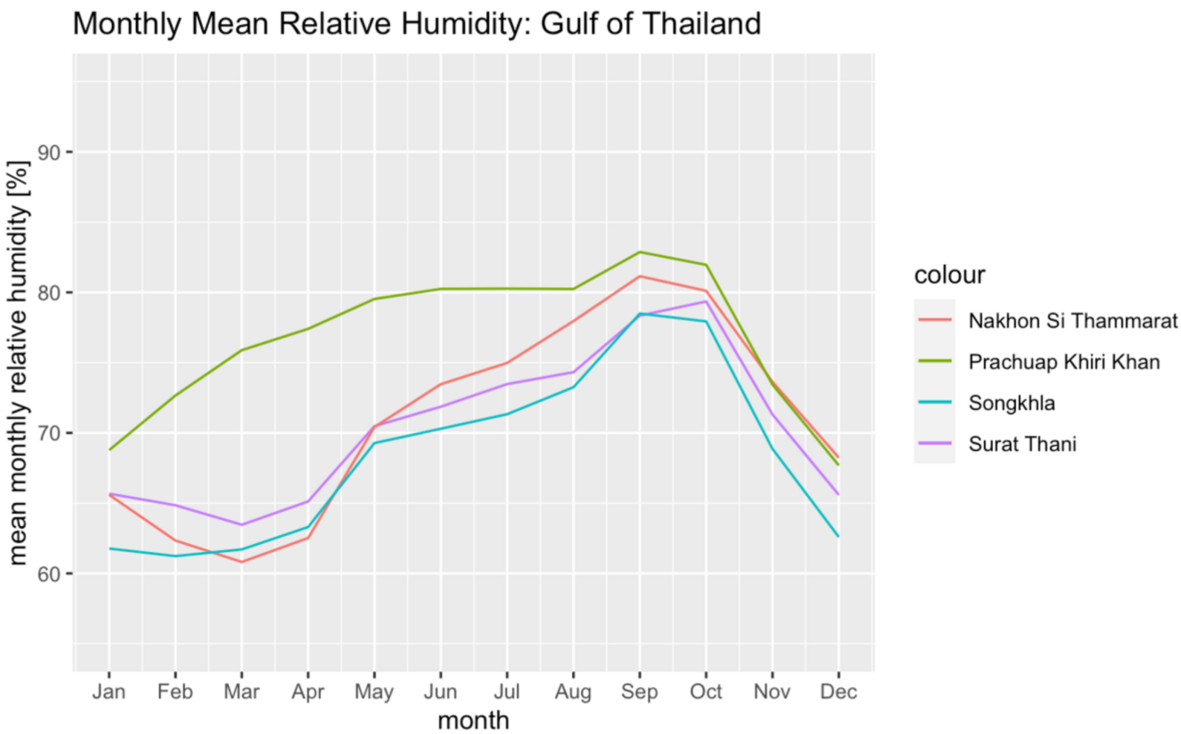



*Abbildung 15: monthly mean relative humidity in provinces near the Gulf of Thailand*