

Reflection Master Thesis Bart Roelofs

This reflection report is constructed as part of my Master's Thesis. I have chosen to write my thesis in the form of a paper which eventually may be transformed in a peer reviewed article. A journal in which I would be very happy to publish would be *International Journal of Infectious Diseases*, as this is an open access journal, which is something I am in favour of. Additionally, they state to “*place a particular emphasis placed on those diseases that are most common in under-resourced countries*” (IJID, 2020), which I think is something that aligns with this research. Because of the limitations that this journal places on its articles, for example a word count of maximum 4000 words, this reflection report serves the purpose to inform the reader on the reasoning behind the decisions that were made during the writing of the paper. A research paper is bound to more or less set academic writing rules, which makes it difficult to reflect on personal thoughts and considerations. I have chosen to write this reflection a bit more light-heartedly, to better inform the reader about my experiences during the process of writing this thesis.

This master thesis originates from two of my largest research interests: health geography and GIS. During the last year of my Master's I have been searching for a topic which embodies the combination of the two. The spread of diseases throughout a region is something I soon found to be of great interest, as it a combination of geography and health. Through contact with my supervisor from the FSS I got in contact with a researcher from the UMCG (later to become my external supervisor), with experience with geographical research in health geography, or specifically spatial epidemiology. I developed a research proposal during the course Research Process and Proposal Writing, which was a great starting point for this thesis.

This report will start with a reflection on the specific contents of the research paper and will end with a reflection on ethical issues and the research process.

Research Problem and Research questions:

Over 40% of the world population is currently at risk for dengue fever and the WHO has stated dengue to be one of the 10 largest threats to global health (WHO, 2019). Curaçao is in the ‘hot-zone’ of dengue infections, with chances of getting infected throughout the whole year. There is no cure for dengue, which makes the combatting of this disease difficult and reliant on awareness and vector control (Elsinga, 2018). As vector control is the most profound way of combatting dengue, having knowledge on the geographical and temporal hot spots is of great importance. This study can adequately inform institutions on the locations and periods in which dengue has been present over the past decades, which can help them in making decisions on where and when to apply the vector control.

There are multiple studies reporting on the geographical spread of dengue (Schmidt, 2011; Yue, 2018) and there appears to be large differences in the number of dengue infections and the causes of the number of infections (Hernandez-Gaitan, 2017; Liu, 2019). These large differences make it difficult to have an adequate knowledge transfer from one place to another, as there are multiple variables which can affect the outcomes in other areas. This effect of geography makes it important to study areas with high dengue occurrence individually. This has been the case for countries such as Malaysia (i.e. Majid, 2019) and Venezuela (i.e. Vincenti-Gonzalez), but the research on dengue on Curaçao is limited. During the exploration of the literature on the topic, I could only find a handful of papers on dengue on Curaçao, with none specifically researching the spatial trends and only one studying the temporal trends (Limper, 2014). Searching the WorldCat article library, four articles which included ‘dengue’ and ‘Curaçao’ in their title were returned. A similar search on PubMed returned five articles. The small number of studies into the spatial and temporal trends of dengue on

Curaçao limit our understanding of the underlying mechanisms that are taking place on Curaçao. Therefore, a better understanding of spatial and temporal trends is required to be able to inform relevant actors about the spread and patterns of the dengue virus.

Throughout my studies the importance of starting a research from a question or a theory has been taught to me by many teachers. As much as I agree on this and clearly see the strength of this, I feel it is not uncommon to take a different approach. In my case, I was given access to a dataset which allowed me to combine my two research interests. Upon exploring the data, a multitude of questions and research ideas arose, which in turn were the foundation of this research. This does mean that the research questions are based on the data, instead of searching or collecting data to answer your research questions. I think this is not a problem, as long as you, as a researcher, are aware of this fact. Doing research that originates from data can result in mis specified research questions or false assumptions, as the questions can be tailored to the expectations of the results. By being aware of this, I tried to create open ended research questions to prevent these problems from occurring. This resulted in the following research question:

What are the spatial and temporal trends of dengue virus infections on Curaçao in the period from 1995-2016?

I created sub-questions to be able to more adequately indicate my purposes of this research. These questions are specified below:

- What is the extent of spatial heterogeneity for the dengue virus on Curaçao?
- What are the geographical patterns of dengue infections on Curaçao?
- What are the temporal patterns of dengue infections on Curaçao?
- To what other factors can potential geographical and temporal patterns of dengue infections be related?

These research questions have their origins in different theories and earlier findings. The first sub question specifically questions the spatial heterogeneity. Spatial heterogeneity in a point pattern setting is explained by Dutilleul et al. (1993) as the distribution of individuals or objects through space and their corresponding variation in density as opposed to a randomly distributed variation in density. Questioning this is necessary, as it provides a statistical proof for the patterns observed in the distribution of the dengue cases.

The second sub-question originates from the data exploration, in which a non uniform spread of the dengue infections was visible, as well as from other research on dengue virus. Research mentioned in the paper, such as the studies by Bisanzio (2018), Yue (2018) and Vincenti-Gonzalez (2018), all state the geographical differences in dengue infections. As this has not been investigated on Curaçao, it is important to include this in the research.

The third sub-question is created similarly to the second sub-question. By having an open-ended sub question which aims to explore the patterns, this allows me to apply a multitude of analysis methods. Research by Limper et al. (2014) identified temporal patterns related to climatic variables which could be investigated even more using the data available in this research.

The fourth-sub question originates from the research by Elsinga (2018), Yue (2018) and Liu (2019), in which explanatory variables are explored which are different to population and climate. This can prove insight in the explanation of certain geographical or temporal trends, which are difficult to answer with just pattern exploration. Additionally, the findings by Vincenti-Gonzalez et al. (2018) on

the effect of the El Niño Southern Oscillation, resulted in the idea of exploring temporal and climatic trends.

Finally, I have created some hypotheses, based on my knowledge of the subject and other literature. I think listing hypotheses is helpful in a data-driven research where the questions are deliberately open ended. The hypotheses that are listed below were developed to help structure this research and I will return to these hypotheses in the final parts of this reflection.

- There is a relationship between population density and dengue cases
- There is a relationship between socio-economic variables and dengue cases
- There is a relationship between time, specifically seasonality, and dengue cases
- Rainfall or temperatures changes affect dengue cases after a certain time period has passed.

Theoretical Framework

The viewpoint of this research originates from the positivist paradigm; making sense of the world through observation and experimentation (Babbie, 2010). Conducting a study with a large dataset, which represents the whole population, allows a researcher to discover patterns and make generalizations on the whole study period. Research on the community, not the individual level, is another trait of positivist research. In this study, outcomes on the community level are of primary interest, making the positivist paradigm a logical starting point for this research.

Although the strengths of the positivist paradigm in this research have been stated, it is helpful to reflect on the missed opportunities of other paradigms. The usage of grounded theory could have increased understanding of the processes on an individual level. Grounded theory attempts to derive knowledge and understanding from common themes and patterns in observational data. In a sense, this is very similar as the approach in this study, albeit grounded theory often includes qualitative data collection and analysis. In this study, qualitative research could have brought insight in the experiences of people with dengue; Are they aware of the dangers and protection methods? What patterns do patients of dengue see? These questions could provide insights in things out of reach of the current study design. There are however, practical issues in conducting this study in such a way, with the most obvious one being the difficulty in data collection. Additionally, with the starting point being the exploration of the topic with GIS analysis, the positivist paradigm felt most suited.

A key theory behind this research as well as many other geographical research and analysis, is the famous “First law of Geography” by Waldo Tobler. This “law” states that: “everything is related to everything else, but close things are more related than distant things”. The word law is purposefully noted between brackets, as there has been a quite extensive debate by five scholars on the validity of this Law and its terminology (i.e. Miller, 2004; Barnes, 2004; Tobler, 2004). Sui (2004) concluded in his synopsis of the debate between the five scholars that: even though there might not be a consensus of the validity of the law, the spirit of the law has certainly become encased in the modern understanding of geography. It is important to think about the validity and implications of this law in this research, as it is embedded in many analysis methods available in GIS.

This law is an important driver behind this research, because many of the analysis methods are based on this assumption. I.e. the optimized hot spot analysis or the Kuldorff’s scan identify clusters based on the intensity of the phenomena in the surrounding areas. Having a lot of cases in the vicinity increases the chances of a location being a hot spot, while lots of cases further away have less influence. Even during the investigation of the socio-economic variables, for example the GDP per capita, the underlying assumption is based on geography and in a sense on Tobler’s Law. Certain areas are not expected to have more dengue cases because the people living in it are poorer, but because the area in which they live is poor. A single household with a low GDP in a high income

neighbourhood is not expected to increase the number of dengue cases in a neighbourhood, but a high income household in a lower income neighbourhood is expected to have more chances of dengue infections. Additionally, when investigating temporal variables such as climate there is a geographical aspect in which Tobler's law is present as well. Areas with more precipitation are more likely to have a higher mosquito presence and areas close to these areas are more affected than areas further away.

Additionally, the law has a big role in the interpretation of the results. As this law is something that I have been taught throughout my studies, I feel like I automatically interpret my results from this viewpoint. When I identified clusters in certain areas, I automatically connected this to the geographical distribution of the variables related to this cluster. I believe Tobler's law is a strong starting point for a geographical study such as this one, especially when the focus is on the community level. It is however, important to be aware of this and to acknowledge that even though Tobler's Law states that near things are more related than distant things, this might also be due to non-geographical processes.

Another key concept in this study, as well as in GIS based research in general, is spatial autocorrelation. Getis (2008) describes in his paper the long road this concept has travelled before being cemented in GIS science. While the research of relationships and influence between geographical locations has been traced back to the 1800's, it were Moran, Iyer and Geary who laid the mathematical foundations of the join count statistic, being coined by Moran as 'spatial correlation' (Moran, 1950). It was a few years later, in 1968, when two – now famous- researchers, Cliff and Ord published their work on spatial autocorrelation. It is important to recognize the economical background of many of the influential researchers on spatial autocorrelation, especially considering the ecological character for this study. Spatial autocorrelation can very well be used in ecological studies, which is proven by the wide occurrence of the subject in ecological research, but it is worth keeping in mind the background and purposes of its creators. The great importance of spatial autocorrelation originates from it being a statistical measure which can confirm or debunk observations. In GIS based research, in which data is often transformed from tables into visual representations of data, a lot is drawn from observations of the data. This is a strength of GIS research, as there are patterns which only appear when they are made visible through spatial visualisation. It can however, be a weakness as it is possible to overemphasize some observations in the data. This is where measures of spatial autocorrelation come into play, as these are methods to statistically review whether observed patterns are indeed patterns or merely caused randomly.

Literature review

A lot of research on dengue virus shares the same conclusion: areas with a higher population density appear to have more registered cases of dengue and people in these areas are more prone to dengue infections (i.e. Yue, 2018; Vincenti-Gonzalez, 2018; Yuan et al. 2010). Schmidt et al. (2011) argued that this is not limited to cities, as a sufficient population density in rural areas might result in even more cases. According to this research, the ratio between vector and host is the most important predictor for dengue cases.

Several dengue related studies have linked dengue infections with socio-economic status (Elsinga, 2018; Yue, 2018; Bisanzio, 2018). These studies concluded that a lower socio-economic status can be related to more dengue infections. It is however yet unclear what the cause of this relationship can be. Causes could be a lower quality of housing, absence of tap water (Schmidt et al. 2011) or a lack of air-conditioning, which has been proven by Liu et al. (2019) to be a risk increasing factor.

It has conclusively been reported that there is a relationship between climate variations and dengue infections. Studies by Hernandez-Gaitan (2017), Sirisena et al. (2017) and Limper et al. (2014) all established relationships between temperature, rainfall and dengue cases. Most studies report a correlation between higher temperatures and dengue cases. Contrastingly, Limper et al. discovered an effect between dengue incidence and a decrease in temperature on Curaçao. This may be due to the higher overall temperatures on Curaçao as the average temperature is just above 27C* (Meteo Curaçao, 2020), while 27C* is suggested as the ideal breeding temperature for mosquitos carrying dengue in the Caribbean (Limper, 2014). It is however important to place a critical note, as it might as well be that an increase in temperature is paired with a decrease in rainfall. A decrease in rainfall decreases the dengue incidence as well, so there may be an interaction effect happening.

There is a growing body of literature on the effects of El Niño Southern Oscillation (ENSO) on dengue infections. Vincenti-Gonzalez et al. (2018) discovered a relationship between dengue and El Niño, the warming period of the ENSO, in Venezuela using Wavelet analysis. Xiao et al. (2018), discovered similar results using Wavelet analysis in Guangdong Province, China.

Even though a lot of research on dengue exists, there are multiple reasons for the importance of this study. First, the main body of literature on dengue consists of studies which are conducted all over the world, albeit mostly in tropical areas. During my literature study, I read articles on dengue from Australia, Indonesia, Vietnam, Venezuela, Brazil and Malaysia. Even though many of these studies find similar results in i.e. the effect of temperature and rainfall, this does not mean that these findings are easily generalizable. There could be massive local differences in climate, knowledge, testing and vector control and it is possible that different variations result in similar conclusions. This strengthens the necessity of this study, as much is yet unknown about the dengue virus on Curaçao and making policy based on research in other countries could yield wrong conclusions. Second, the combination of data being used and the multitude of analysis methods proposed, make this research a thorough investigation, addressing specific subjects that other studies have noted as limitations. For example, the research by Limper et al. (2014) on the effect of climatic data on dengue cases is conducted on the basis of data ranging from 1999 to 2009 and did not include a spatial analysis. The research by Carvalho (2017) on dengue infections in Brazil is conducted using Kernel density analysis and Majid et al. (2019) conducted spatial pattern analysis on dengue cases in Malaysia by comparing Spatial Means of 2008 and 2009. Both these studies provided insights in the distribution of dengue cases, but I feel they could have been improved by using a larger variety of analysis methods. Researching spatial as well as temporal trends will fill the current knowledge gap on the trends of dengue virus infections on Curaçao.

Conceptual model

The conceptual model behind this research is presented in figure 1. The starting point is the main research question: *What are the spatial and temporal trends of dengue virus infections on Curaçao in the period from 1995-2016?* This research question is divided into sub questions which are represented by the three coloured domains and should be interpreted as a research process going from left to right. The first (yellow) domain concerns the geographical aspect of the research question and includes sub-question a and b. The second (blue) domain represents the temporal aspect of this study which is operationalized in sub-question c. The final (green) domain represents the third part of this study in which explanations behind the patterns are being identified, as is questioned in sub-question d. This study design, in which the final sub-question is based upon the findings of the previous sub-questions, is symbolized even further with the combination of colours, as green is a combination of yellow and blue. When zooming in on the individual domains, it becomes apparent that the variables used in this research are stated at the top of the figure,

symbolizing that these are used as a starting point in each domain. The further you look down in the domain, the more advanced methods these input variables are being subjected to. The outcomes of these sub-questions, which are represented by the domains, are used to answer the research question which will ultimately result in an increased understanding of dengue on Curaçao between 1995 and 2016.

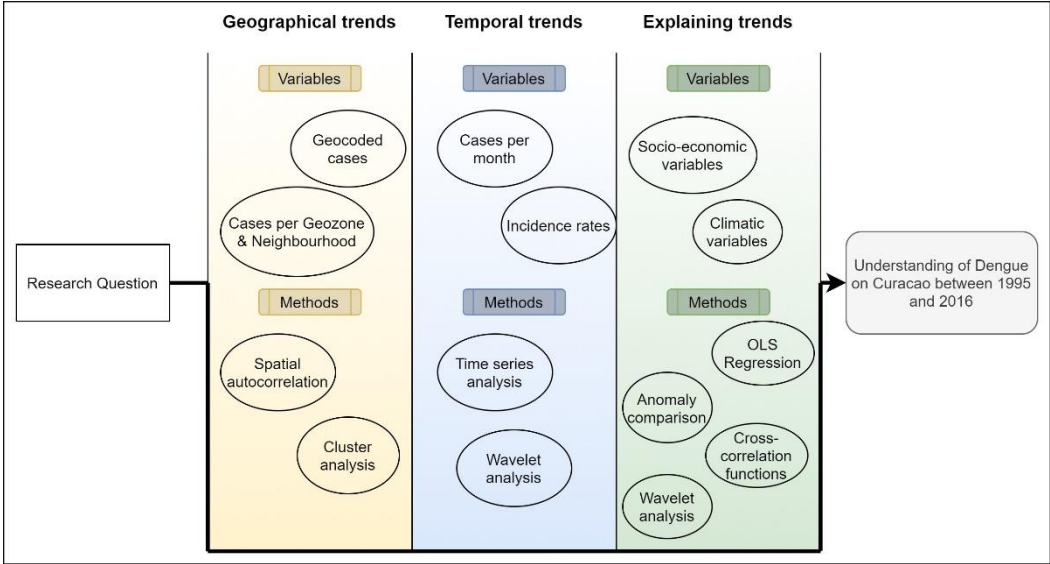


Figure 1: Conceptual Model

Data

The data consists of dengue virus infection databases from the Ministry of Health on Curaçao. These databases contain all registered information from a person that is tested for dengue. Variables included in these databases are specified in the table 1 below. This data is collected as part of the patient registration system. Every person that visits a medical facility and gets tested for dengue is registered in such a database.

Variable	Notes
Patient ID	
Date of birth	
Address (home)	
Address (work)	Not always included
Geozone	Geographical administrative units on Curaçao
Date of testing	
Date of symptoms start	
Month of symptoms start	
Week of symptoms start	
IGM 1/2 values	Number of dengue antibodies found
Interpretation	Probably dengue, confidently dengue, no dengue
Haemorrhagic fever	Severe dengue
Symptomatic variables such as 'Fever, Rash, Vomit, Headache'	Not available for all cases

Table 1: Variables in the dengue dataset

The data used in the OLS regression originates from the Curaçao 2011 Census, which is displayed in table 2. The data on temperature, humidity and precipitation originate from the Hato Airport Meteorological weather station. The SST time-series were obtained from the Climate Prediction Centre of the National Oceanic and Atmospheric Administration (CPC, 2016). All climatic data is combined into a dataset of which the structure is displayed in table 3.

Variable
Geozone
Economic status: Employed
Economic status: Unemployed
Economic status: Not active
Economic status: Not reported
Ratio: Employed
Ratio: Unemployed
Ratio: Inactivity
Total Average gross monthly income per household
Total population
Population density (population per km ²)

Table 2: Variables in the 2011 Census dataset

Variable
Year
Month
Temperature (Average)
Temperature (Maximum)
Temperature (Minimum)
Relative humidity
Precipitation
Sea Surface Temperature (SST)
SST anomalies

Table 3: Variables in the climatic dataset

Data quality

The quality of the data in research like this is of great importance. The main analysis methods are based on a location that should be a correct representation of the actual location as specified by the patient. It is however, because of multiple factors, unlikely that the data is a 100% correct representation of reality, as the data has gone through many stages which all may have contributed to a lower quality.

First, the data is submitted into the computer by the medical practitioner who is talking to the patient. In this stage, mistakes can happen during the communication between medical practitioner and patient; as the people on Curaçao speak multiple languages, translation errors can happen. Patients can provide their address in i.e. English, although the actual address is in Papiamentu, which can have profound implications during the geocoding of this data.

Second, another point of concern is the address system on Curaçao, which uses only street names and numbers. A lack of postal codes results in house addresses being only identifiable with the street

name, which is prone to typing or translation errors. Additionally, lots of streets are named after a nearby landmark or former plantation, which are not official addresses. Multiple records only included the name of this landmark or former plantation as address, which decreased the accuracy of the location.

Third, the geographical representation of the dengue case is the home address of the patient. It is however, unclear where the patient got infected with the dengue virus. This could have happened at the workplace, in the supermarket or during a visit to the beach. It is nevertheless likely that the infection happened at home, due to the fact that the *Aedes Aegypti* mosquito is a mosquito which is most likely to bite during sunrise and sunset (Vincenti-Gonzalez, 2018) which are moments that many people reside in their homes. Additionally, Elsinga (2018) stated that there are many breeding places around the house, making the chance of getting infected at home larger.

Finally, the data consists only of registered cases. People who are infected with dengue fever but only develop mild symptoms will not go to the doctor and get tested, thus making the data an underrepresentation of the actual phenomenon. This does not result in immediate problems with the analysis, as we can expect a distribution in unregistered cases similar to the registered cases, but it is important to keep this in mind.

This research could have been improved by data changes in multiple ways. In an optimal situation, the patient data on dengue cases would have been accompanied by socio-economic variables such as household income and education level. In this study, the socio-economic variables are drawn from the 2011 Curaçao census, but the number of variables in this census is limited. Additionally, there is a discrepancy in the resolution of this data, as the 2011 census is based upon Geozones, which are administrative units that are larger than the neighbourhoods and ultimately less precise as compared to individual data. Resolution is an aspect which could be improved in the climatic data as well. The climatic data (humidity, precipitation, average temperature and sea surface temperature) originates from a single measuring station on Curaçao. Ideally, the climatic variables would have been measured at different locations of the island, as local variation in weather might have an effect on dengue cases as well.

Data preparation

The first part, the geocoding, was an important step as it allowed for data exploration and the GIS methods to be conducted. The process of geocoding has been challenging due to several factors.

First, the data was obtained in separate Excel tables for every year. Most of these Excel tables had different formats and different ways in which the data was structured. To combine all these tables into one complete database, a lot of computations in structure had to be made.

Second, the geocoding of the data did not go as easy as expected. The main geocoder I planned to use was the Esri Geocoder, which is part of the ArcGIS Pro software. This geocoder did however, not cover addresses on Curaçao. After consideration of many other geocoders, the LocalFocus geocoder was chosen, which makes use of OpenStreetMap data and the Pelias geocoder. The OpenStreetMap data was deemed most suitable, as it appeared to be the most complete dataset.

Third, there were, despite the fact that the best geocoder was used, still limitations in the process of geocoding. There is a lack of a decent functioning address system on Curaçao. There are no postal codes which are tied to specific houses, making the geocoding dependent on address data. Many addresses on Curaçao are in either one of three languages: Papiamentu, Dutch or English. There can be a discrepancy between the language in which an address is written in the database and the language of the addresses in the online geocoder. This resulted in many addresses hard to geocode,

which resulted in adjustments by hand based on other street network databases such as the ones by Google or Esri.

Methodology

The aim of this study is to shine new light on the spread and distribution of dengue cases throughout Curaçao from 1995 until 2016. This may provide insights in the way dengue has evolved throughout the 21-year study period, which in turn can increase the understanding of future dengue behaviour. The combination of the investigation of spatial heterogeneity and geographical clusters, as well as the changes over time and weather will provide a concise picture of the situation of dengue. The research methods are designed to answer the research questions in the most adequate manner. Every sub-question is tied to different analysis methods which will strengthen the outcomes, as the overall goal, which is to provide insights in the evolution of the dengue virus, will be approached from different angles. The starting point is the research question, which is:

What are the spatial and temporal trends of dengue virus infections on Curaçao in the period from 1995-2016?

This question contains two clearly different parts, a spatial and a temporal one, which both are best answered using different methodological approaches. The sub-questions are numbered and stated below, during the methodology section references to the numbers of the questions will be made.

- a. What is the extent of spatial heterogeneity for the dengue virus on Curaçao?
- b. What are the geographical patterns of dengue infections on Curaçao?
- c. What are the temporal patterns of dengue infections on Curaçao?
- d. To what other factors can potential geographical and temporal patterns of dengue infections be related?

Sub-question a and b are clearly linked to the geographical part of the research question, while sub-question c concerns the temporal aspect. Sub-question d regards the explanation of observed patterns during the previous sub-questions. Logically, the geographically oriented sub question a and b are answered using geographical based analysis methods by means of GIS, while sub question c will be answered using time series analysis. Sub-question d will be approached using a combination of the two, depending on the results of the previous sub-questions. This study design is thus based upon a combination of analysis methods which are focused around pattern detection in space as well as over time. It is the combination of methods regarding geographical as well as temporal patterns which differentiates this study from other studies regarding dengue.

Identifying spatial trends is, as mentioned in the theoretical framework, a process which is often started by observation. The transformation of data tables into visual representations such as maps allows for preliminary analysis by observation, which is often a great starting point for further analysis. It is however, important to validate these observations by means of statistical tests. Sub-question a, regarding spatial heterogeneity, is tied to the tests for spatial autocorrelation. This analysis is incorporated as I think it is necessary to provide such statistical proof whenever possible.

The second type of geographical analysis is linked to sub-question b. As this study aims to detect trends in space, I have sought analysis methods which are suited to this goal. I started out with data exploration and created maps of population, case distribution and incidence. I chose not to include case distribution in the final paper, as it would have been possible to identify individuals based on the location of the case. These preliminary explorations would give an insight in the distribution of

cases through space, but since all cases are included nothing can be concluded about trends over time. Additionally, as no statistical measures are included in these analysis, other methods of trend detection needed to be conducted to provide statistical proof. An example of a geographical analysis method which can identify locations with a statistically significant high intensity of cases is cluster analysis. Conducting cluster analysis would allow me to gain an overview of different clusters through space and time which could be compared afterwards to allow for trend discovery. The two types of cluster analyses that are conducted are Optimized Hot Spot analysis using the Getis-Ord G_i^* statistic using Arcgis Pro and the Kulldorff's Statscan analysis using Clusterseer 2.5. Those two cluster analysis methods have different ways to answer a similar question: Getis-Ord G_i^* assesses whether there are features with a statistically significant positive or negative z-scores, which represent areas with an intense clustering of high values (hot spot) or an intense clustering of low values (cold spot). Kulldorff's scan statistic can detect clusters based on space and time (Kulldorff, 1997). The scan statistic provides a measure of whether the observed number of cases is unlikely for a window of that size, using reference values from the entire study area (BioMedWare, 2012). The reason that two types of cluster analysis methods are conducted is due to the different approaches they take in identifying clusters. Optimized Hot Spot analysis indicates a more specific location of a cluster due to the fact that it is based on the original point data. The Kulldorff's scan uses aggregated data, but due to the inclusion of population data in this method, it presents an ordering in the intensity of the different clusters. By having access to the clusters identified in both analysis methods it provides insights in where the clusters are located but also which ones are of a higher intensity.

To investigate the temporal part of this research, as outlined in sub-question c: *What are the temporal patterns of dengue infections on Curaçao?*, initial data explorations as well as time-series analysis were conducted. Similar to the data exploration in the geographical analysis part, transforming the data from tables into visual representations allowed for instant data exploration. The counts per month were extracted from the data and the incidence was calculated, after which both were plotted in a table which is visible in appendix 1. The main analysis method for this sub-question is Time Series analysis, as this allows for investigation of the data for patterns that appear over time. In this analysis, the time series of dengue cases are reviewed and decomposed, after which the trends and seasonality can be explored. Doing this is important as it can prove the seasonality in dengue cases, which was observed in the graph of cases per month. Additionally, identifying a trend may provide insights in whether dengue is increasing or decreasing on Curaçao. The final temporal analysis is Wavelet analysis, which is a different way of analysing time series. By representing the power of the time series as a function of the time and the duration period, the data gets decomposed which allows for insights in patterns over short as well as long periods (Schulte, 2016). By doing this, insights will be given in not just seasonal effects, but larger patterns over multiple years as well. This has a direct relationship with answering sub-question c as it is important to focus not just on the observed 1-year patterns, but also on potential longer-term patterns which are not visible by observation.

The final sub-question d: *"To what other factors can potential geographical and temporal patterns of dengue infections be related?"* is constructed to provide insights in the reasons behind the potential patterns that are discovered in the previous sub-questions. This question is important as it goes into depth with the identified patterns. Instead of ending this research with stating the various patterns and trends discovered, explanations will be sought to increase the understanding of the evolution of the dengue virus on Curaçao. The first way in which these relationships are studied is by conducting the OLS on the geographical distribution of dengue cases and socio-economic variables. The OLS will evaluate the relationship between dengue cases and the socio-economic variables of interest. The inclusion of this method is based upon research by Elsinga (2018), Yue (2018) and Bisanzio (2018), in

which they identified relationships between dengue cases and socio-economic variables. In this study, population numbers, population density, the inactivity ratio and the average gross income per Geozone are tested. The second analysis method related to this sub question regards the effect of climatic variables on dengue cases, a question which is influenced by research of Vincenti-Gonzales (2018), Xiao (2018) and specifically for Curaçao, Limper (2014). This potential effect of climatic variables on dengue cases is studied using a multitude of methods, the first of which being time series analysis. Time series of humidity, precipitation, average temperature and sea surface temperature (SST) are created which are tested for correlations with dengue cases using cross correlation functions. This will shed insights in a potential relationship between the two variables and more specifically the lag in the effect. In this study, the hypothesis is that the amount of dengue cases is influenced by the status of the climate in a certain period before the infections. The third analysis is an anomaly comparison of the dengue cases and the sea surface temperature (SST). This is not a statistical test, but an empirical method to view whether there is a potential relationship between the two. As SST is thought to be a proxy for El Niño Southern Oscillation (ENSO) (Vincenti-Gonzalez et al. 2018), a sudden drop in SST is expected to be caused by La Niña. Observing the anomalies of these variables will provide an insight in whether they might be related. The final analysis consists of a Wavelet analysis of the climatic variables in combination with the dengue cases. The coherence between dengue and climatic variables will be tested which will result in a coherence spectrum. This is a method in which the relationships which are observed in earlier analyses can be tested.

Methods not used

During the writing of this thesis, I have come across many more analysis methods which could have been beneficial in studying patterns of dengue. Commonly used methods in the study of dengue or other arboviral diseases are interpolation methods such as Kriging and IDW. Based on studies in which these methods are used, I initially included these research methods as part of this study. The research question of this research study is however, aimed at exploring the trends found in dengue virus infections. As the data used in this study consists of all the registered cases, it makes most sense to use analysis methods which make use off this data. Kriging and IDW are interpolation methods in which you can estimate the probability of an effect happening, which is especially helpful if there are areas on which you have little to no data. As we have data on all the cases and are interested in why they are in specific locations, it is for this study not that relevant to look for the probabilities that IDW and Kriging provide. These methods are conducted in other studies mainly because those studies have a largely incomplete dataset and aim to estimate the occurrence of dengue in locations in which no data is available. The only possible use of interpolation in this research would be to estimate the locations of non-registered cases. This would however, most likely result in the identification of locations which are already identified as clusters, because all the locations of dengue cases are already known. This would thus not help in answering the research questions which ultimately made me decide not to include these methods. I did compute some of these analyses, mostly because they are the go-to method in all dengue or arboviral disease research and I was trying to base my methodology upon previous studies. This has been a very helpful experience, as it was a learning opportunity to be critical: even if all research on this topic is conducted using a specific method, it is important to question why this method is used and whether your research question and data is suited for this type of analysis.

Another analysis that I conducted but not included was the Space Time Cube analysis and the Emerging Hot Spot analysis. The Space Time Cube is a useful tool for exploring the trends in the data over time, but the output in itself it not something that is useful in a research paper. Additionally, the Emerging Hot Spot analysis is based on the Space Time Cube, but it is a tool for predicting future

possibilities based on the trends in the most recent years of the Space Time Cube. The Emerging Hot Spot analysis returned not that much, as there are not that many dengue cases in the final study years as compared to the large bulk of cases around 2008 until 2011.

Ethical Issues

In the research paper I have reported on the ethical issues of working with personal data which could be identifiable. There are however, some additional ethical considerations I had during the writing of this thesis.

I think the presentation of the research findings is something that is very important and I feel it is a big responsibility to ensure an adequate spreading of findings that can have an effect on the lives of people. I am by no means an expert on knowledge transfer, but I feel that after a research is finished, researchers sometimes stop after publishing an article. I personally believe that it is important to actively promote your findings towards the people that are involved in your research, as expecting them to find and read your article is not sufficient. Originally, I thought that, in the case of the discovery of dengue Hot Spots, I could make flyers or posters to send to these areas to make people aware of their risks for dengue. I do however, think this is something that should be done by a health or -governmental organization on Curaçao. I do think that I have a responsibility in transferring the findings towards the ministry of health or government of Curaçao. This will be tried by transforming this paper in a journal article in the near future, which I will try to share with important actors on the dengue virus on Curaçao. If this does not work out, I will try to send this thesis paper or a summary to important actors in this field on Curaçao. My external supervisor in the UMCG is in contact with such an actor on Curaçao, which makes me confident that this goal will be reached.

Results

Writing the results section is often a straightforward task as it is the presentation of the outcomes of the methods. The large amount of methods that I have used in this research made this a bit more difficult, as a lot of reflection went into considering which results were worth presenting in the paper. As explained in the methods section of the paper, there are some analyses of which the results are not presented. This is due to the fact that these methods were not suited to the research questions as much as expected, for example the interpolation methods, since these were more aimed at predicting instead of analysing. As the results presented in the paper should be aimed at answering the research questions, I ultimately decided to only include those methods which contributed to this goal.

Spatial analysis:

The case exploration at the beginning of the results section is included as it gives an overview of the distribution of cases and people, making it easier to interpret the results in later parts of this section.

The Optimized Hot Spot analysis is conducted on the years in which $N > 100$, due to the fact that conducting this analysis with fewer cases would decrease the reliability drastically. The results are presented in an image in which all valid years are present. In presenting this type of data it is important to choose a correct visualisation and as the changes per year are important, this could not have been possible by displaying the data in a single image. In the case of an online article, these results could be animated or a time-slider could be incorporated to increase the understandability of these results.

The Kulldorff's scan analysis presented interesting results, as it gives an ordering in the power of the clusters. Some clusters of cases in the villages in the northern part of the island appear to be very large. This corresponds to the research by Schmidt (2011), in which is stated that it is not that much related to cities, as villages can be large hot spots as well, as long as it is populated densely enough. As with the Optimized Hot Spot results, it is difficult to present the data on so many years in a single image. Online presentation would be best suited for these kinds of figures.

I think these analysis methods contributed a lot to this research, but I do have some considerations on whether they are optimally suited. The research question is specifically aimed at identifying patterns. Even though these methods detect clusters in specific areas, the patterns over time only become visible by analysing multiple years. In order to do an actual pattern analysis, the patterns of hot spots for the different years could have been tested to review whether there are any significant patterns.

Temporal analysis:

The temporal analysis was originally an umbrella term for all the non-spatial research and it included all analyses on the relationship between dengue and climatic variables. In the later stages of the research I decided to split this into two separate parts: one clear temporal part on the temporal trends of dengue cases and an explanatory part on the explanation of these temporal trends by means of analysing dengue cases in combination with climatic as well as socio-economic variables.

The temporal analysis of dengue cases consists of an analysis on the case distribution per month and week, of which only the former is included in the paper, a time series analysis of dengue cases and a wavelet power spectrum. The decomposition of the time series displayed a clear seasonality in dengue cases, which corresponds with the graph of cases per month and specification of the rainy season by the Curaçao meteorological institute (Meteo Curaçao, 2020). The Wavelet analysis is a great method to gain insights in trends on longer periods of time. The Wavelet analysis of dengue cases indicated a significant trend on a three-year cycle, which I would otherwise not have been able to identify.

Explanatory analysis

The explanatory analysis consists of methods which try to answer the 'why' question behind the patterns explored in the previous methods. This started with conducting the OLS regression on socio-economic variables and the dengue cases. The OLS regression found a significant relationship between population, population density and dengue cases. Relationships between variables such as gross average household income were not found. This may be due to the fact that the data was only available at the Geozone level, which is a relatively low resolution as compared to the individual dengue case data. Additionally, only the data for 2011 was available, a larger dataset on multiple years could have increased the quality of this analysis.

In the time series decomposition of the dengue cases there is clear seasonality, as the graph shows distinct seasonal peaks. The climatic variables all show seasonality as well, which is logical as the climate is a seasonal phenomenon.

The cross-correlation functions provide insight in the effect of a variable in a past moment on dengue infections now. The cross-correlation function between sea surface temperature (SST) and dengue cases revealed a negative relationship up to four months; indication that a colder sea surface temperature results in an increase in dengue cases over the next four months. The cross-correlation function between precipitation and dengue cases shows a positive relationship up to three months.

This indicates that an increase in rainfall has an effect on dengue cases for up to three months. The combination of these results; a lower SST and an increase in rainfall is an interesting observation. This is in line with the observations on anomalies related to La Niña. La Niña causes a drop in SST and an increase in precipitation, which in turn results in an increase in dengue cases over the next months.

The anomaly comparison is used as another method to review the relationship between dengue cases and SST and thus La Niña. Interestingly enough, in large parts of south America, El Niño and its related rise in SST and temperature is related to an increase in dengue infections, while the opposite appears to be the case on Curaçao. La Niña is characterized by a drop in SST and an increase in rainfall. Both these effects are correlated with the amount of dengue infections. It is more difficult to see these types of relationships with humidity or temperature as compared to the SST, as these are much more influenced by the daily and seasonal climate. There appears to be a clear pattern, as an increase in SST happens alongside in a decrease in dengue cases, while a decrease in SST happens simultaneously with an increase in dengue cases.

Conclusion & Discussion

The paper has presented an overview of the spatio-temporal trends of dengue infections on Curaçao from the period 1995 until 2016. The combination of spatial and temporal methods allowed for a thorough overview of the situation.

The spatial analysis confirmed the link between dengue cases and population density, as stated in other dengue research. This was confirmed using two types of cluster analysis, which both identified multiple clusters over large parts of the study period. A relationship with socio-economic variables was not found, but this may be due to only one year of data (2011) being available for this analysis and the interesting division in geographical units on Curaçao.

The temporal analysis identified a relationship between dengue infections and climatic variables with a high likelihood of this relationship being tied to La Niña. This is an interesting finding, as for many other countries, it is El Niño, the opposite climatic trend of La Niña, which causes an increase in dengue infections. A lag effect of up to four months was discovered for climatic variables on dengue infections. I think it was difficult to translate the results into a clear discussion which entailed everything that I wanted to discuss. The discovery of the relationship between La Niña and dengue infections on Curaçao is something that I think is really interesting and I feel that it should be the main take away message of the discussion. As I did not find any sources who reported on the relationship between dengue and La Niña, only on the effect between El Niño and dengue infections, I think it is a particularly important finding.

At the start of this reflection I listed four hypotheses which I think have contributed to keeping my focus in this research. The hypotheses are:

- *There is a relationship between population density and dengue cases*
- *There is a relationship between socio-economic variables and dengue cases*
- *There is a relationship between time, specifically seasonality, and dengue cases*
- *Rainfall or temperatures changes affect dengue cases after a certain time period has passed.*

The first of these hypotheses is one I am confident to accept. The various spatial analysis methods and the OLS regression all concluded this relationship. The second hypothesis however, I am not able to accept. Even though other researchers have established such relationships, I did not find any. This might be due to the fact that my data was not suitable as I only had access to data for 2011. When

analysing this over a longer period of time it might result in the discovery of patterns on a larger scale. The third hypothesis is accepted, a clear seasonal link between dengue cases was discovered. What I did not expect to find, was the link I was able to make with La Niña based on the relationships with sea surface temperature and precipitation. The fourth hypothesis is accepted as well, the cross-correlation functions gave a clear insight in the cause and effect relationships between temperature, rainfall and dengue cases.

Process reflection

Writing the Master's thesis has been a valuable experience in many different ways. It was the final assignment in which I could (and was expected to) build upon all the knowledge I have gathered during the Research Master. I have once again learned a lot about the way in which I approach a research, what my work habits are and how I can create structure in my working process. I think it is important for me to keep reflecting on what I am doing and why, during the whole research process. I have more than once found myself diving head on in an interesting research method, to only question myself halfway through whether this is actually beneficial for my research. This is something I will try to take with me to my further research projects. Additionally, I think I could make more use of my supervisors, although I have made a lot of progress throughout the process of writing this thesis. I often find it difficult to ask questions or consult my supervisors, as I did not want to bother them as they are often quite busy as well. As I was very aware of this fact, I deliberately discussed this with my supervisors, after which they assured me to ask whatever was needed, as this was something they had already agreed upon by supervising me. This has been very helpful and I have asked more than ever, which resulted in lower stress levels and more direction on what I wanted to do on a day to day basis.

A great influence in the coming in to being of this research is the interplay between the research aim, the research questions and the conceptual model. I think having a research aim which is relatively broad has been a blessing as well as a curse. The opportunities that it provided is that it allowed for exploration of many methods, which was very suitable for the type of data that was at the heart of this research. The flipside of this is that it was often difficult to find a direction in the early stages of this research. The creation of the conceptual model provided structure in finding this direction, as it resulted in the first visualisation of splitting up the research aim and research question into multiple parts. The decision to ultimately split the research question into the sub questions, which questioned the aspects of geographical patterns, temporal patterns and the explanation of the patterns, resulted in a clear structure which has been very helpful.

As a reflection on the research subject itself, I feel that working as a researcher on something that has happened 7000 kilometres away, will most likely affect the process in some ways. As I have never been to Curaçao, I do not have an idea of what my study area looks like, what sizes the urban areas are and so on. Although this might not look that important, I think that there will always be some issues that I did not notice because I have not been there.

Finally, I think it is important to acknowledge the fact that this research is a data oriented as well as computer-based research. This was a purposeful decision, as I wanted to explore this data using a variety of computer-based research methods. It is however, worth reflecting on different ways in which this research could have been conducted. I could have tried to collect data myself, either through sending (online) surveys, or collecting interviews, about experiences of people with the dengue virus or their thoughts on the spread of the dengue virus. This would of course have drastically changed the structure and the outcomes of this research and I feel I would be less likely to

have confidence in answering my research questions. I do think that a combination of quantitative and qualitative research could have been very valuable. Talking to people about the dengue virus and for example their experiences in wet and dry seasons could shed insights on the human aspect of dengue virus infections. This would provide insights in their knowledge and best practices regarding dengue, as exposure to mosquitos might have resulted in the development of specific strategies to prevent infections. These strategies could in turn be used in order to strengthen educational programmes or policy.

To conclude this reflection, I would like to end with that I am very happy with how the process of writing this thesis went. I feel like I have worked very hard and that the hard work ultimately paid off, not just in the form of this output, but in the development as my research skills as well. Ultimately, I would like to thank my supervisors, Daniella Vos and Maria Vincenti-Gonzalez for all their help during the writing of this thesis. Not only did they provide valuable feedback, they also contributed to my mental wellbeing by being supportive throughout the process.

References

- Abd Majid, N., Muhamad Nazi, N. & Mohamed, A. (2019). Distribution and Spatial Pattern Analysis on dengue Cases in Seremban District, Negeri Sembilan, Malaysia. *Sustainability*. 11. 3572. 10.3390/su11133572.
- Babbie, E. R. (2010). *The practice of social research*. Belmont, Calif: Wadsworth Cengage.
- Barnes, T.J. (2004). A Paper Related to Everything but More Related to Local Things, *Annals of the Association of American Geographers*, 94:2, 278-283, DOI: 10.1111/j.1467-8306.2004.09402004.x
- BioMedWare (2012) *ClusterSeer User Manual book 1 version 2.5*. As retrieved on 20-06-2020 from: http://www.biomedware.com/wp-content/uploads/2018/05/ClusterSeer_manual1_2.5_web.pdf
- Bisanzio D, et al. (2018). Spatio-temporal coherence of dengue, chikungunya and Zika outbreaks in Merida, Mexico. *PLOS Negl. Trop. Dis.* 12, e0006298 doi: <https://10.1371/journal.pntd.0006298>
- Carvalho, S. & Magalhães, M. & Medronho, R. (2017). Analysis of the spatial distribution of dengue cases in the city of Rio de Janeiro, 2011 and 2012. *Revista de Saúde Pública*. 51. 10.11606/s1518-8787.2017051006239.
- Climate prediction Center (CPC). *Monthly Atmospheric and SST Indices*. As retrieved on 24-06-2020 from: <https://www.cpc.ncep.noaa.gov/data/indices/>
- Dutilleul, P., & Legendre, P. (1993). Spatial Heterogeneity against Heteroscedasticity: An Ecological Paradigm versus a Statistical Concept. *Oikos*. 66(1), p. 152-171. doi: <https://10.2307/3545210>
- Elsinga, J. (2018). Towards sustainable management of arboviral diseases: A multidisciplinary mixed methods approach in Curaçao and Venezuela. [Groningen]: University of Groningen.
- Getis, Arthur. (2008). A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective. *Geographical Analysis*. 40. 297 - 309. 10.1111/j.1538-4632.2008.00727.x.
- Hernandez-Gaitan et al. (2017) 20 Years Spatial-Temporal Analysis of dengue Fever and Hemorrhagic Fever in Mexico. *Archives of Medical Research*. 48(7), pp. 653-662
- International Journal of Infectious Diseases (IJID) (2020). *Aim and Scope*. As retrieved on 01-07-2020 from: <https://www.ijidonline.com/content/aims>

- Kulldorff, M. (1997). A Spatial Scan Statistic. *Communications in Statistics - Theory and Methods*. 26. 1481-1496. 10.1080/03610929708831995.
- Limper, M., Thai, K. T. D., Gerstenbluth, I., Osterhaus, A. D. M. E., Duits, A.J. & van Gorp, E.C.M. (2014). Climate Factors as Important Determinants of dengue Incidence in Curaçao. *Zoonoses and Public Health*. 63. P. 129-137. Doi: <https://10.1111/zph.12213>
- Liu, J., Tian, X., Deng, Y., Du, Z., Liang, T., Hao, Y. & Zhang, D. (2019). Risk Factors Associated with dengue Virus Infection in Guangdong Province: A Community-Based Case-Control Study. *International Journal of Environmental Research and Public Health*. 16. 617. 10.3390/ijerph16040617.
- Miller, J. H. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers* Vol. 94, No. 2, pp. 284-289
- Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*. 37, 17–23.
- Schmidt, W.P., Suzuki, M., Dinh Thiem, V., White, R.G., Tsuzuki, A., et al. (2011) Population Density, Water Supply, and the Risk of dengue Fever in Vietnam: Cohort Study and Spatial Analysis. *PLOS Medicine* 8(8): e1001082. <https://doi.org/10.1371/journal.pmed.1001082>
- Sirisena, P., Noordeen, F., Kurukulasuriya, H., Romesh, T.A. & Fernando, L. (2017) Effect of Climatic Factors and Population Density on the Distribution of dengue in Sri Lanka: A GIS Based Evaluation for Prediction of Outbreaks. *PLoS ONE* 12(1): e0166806. <https://doi.org/10.1371/journal.pone.0166806>
- Sui, D. Z. (2004). Tobler's First Law of Geography: A Big Idea for a Small World?, *Annals of the Association of American Geographers*, 94:2, 269-277, DOI: 10.1111/j.1467-8306.2004.09402003.x
- Tobler W. (2004). On the First Law of Geography: A Reply. *Annals of the Association of American Geographers*, 94:2, 304-310, DOI: 10.1111/j.1467-8306.2004.09402009.x
- Vincenti-Gonzalez, M.F., Tami, A., Lizarazo, E.F. et al. (2018). ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. *Sci Rep* 8, 5727 <https://doi.org/10.1038/s41598-018-24003-z>
- Vincenti-Gonzalez, M. F. (2018). Spatio-temporal dynamics of dengue and chikungunya: Understanding arboviral transmission patterns to improve surveillance and control. [Groningen]: University of Groningen.
- WHO (2019). Top 10 threats to global health in 2019. As retrieved on 20-01-2020 from: <https://www.who.int/news-room/feature-stories/ten-threats-to-global-health-in-2019>
- Yue Y., Sun J., Liu X., Ren D., Liu Q., Xiao X., Lu L. (2018). Spatial analysis of dengue fever and exploration of its environmental and socio-economic risk factors using ordinary least squares: A case study in five districts of Guangzhou City, China, 2014. *International Journal of Infectious Diseases*, 75, p. 39-48. <https://doi.org/10.1016/j.ijid.2018.07.023>

Appendices

Appendix 1. Incidence counts and incidence rates of dengue cases on Curaçao from 1995-2016

