

**Spatial Temporal and Statistical
Learning of long-term and short-
term trend of hand, foot and
mouth disease, Da Nang city,
Vietnam**

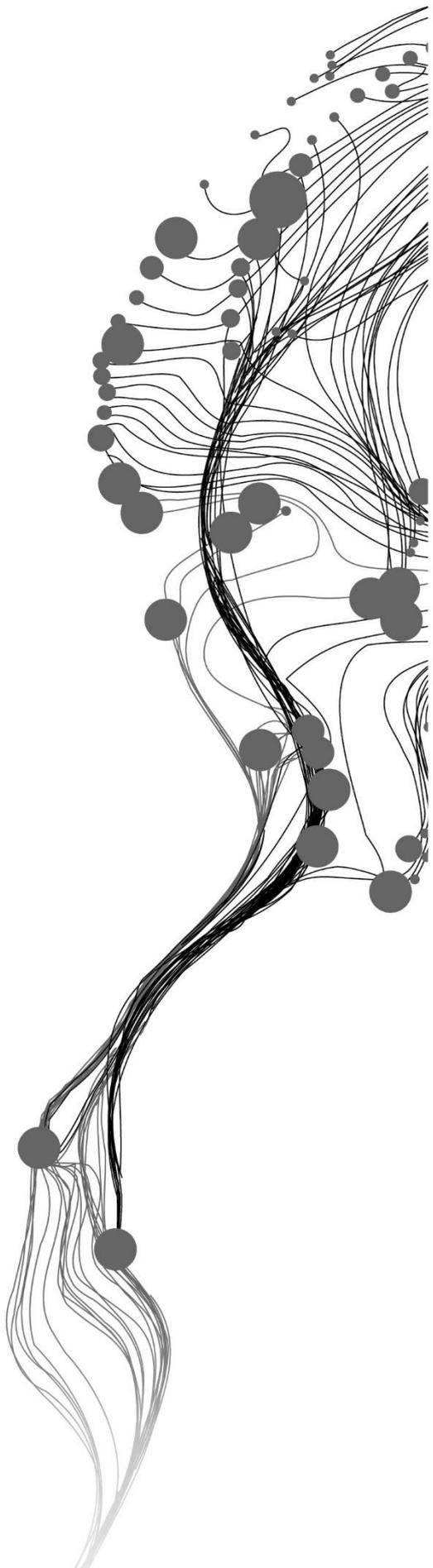
ANDY J.E. BAPTISTE

March, 2019

SUPERVISORS:

dr. Phuong N. Truong

dr. Peng Jia



Spatial Temporal and Statistical Learning of long-term and short- term trend of hand, foot and mouth disease, Da Nang city, Vietnam

ANDY J.E. BAPTISTE

Enschede, The Netherlands, March, 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Geoinformatics (GFM)

SUPERVISORS:

dr. Phuong N. Truong

dr. Peng Jia

THESIS ASSESSMENT BOARD:

Prof. dr. Ir. A. Stein

Prof. dr. Ir. G.B.M. Heuvelink, External Examiner, Wageningen University & Research, Laboratory of Geo-Information Science and Remote Sensing

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Background: Over the years, hand, foot and mouth disease (HFMD) has been considered a public health challenge worldwide, especially in the South East Asian region including Vietnam. This infectious disease mostly occurs in infants and children, and to date, no effective vaccines or drugs to combat this disease have been developed. Existing studies have shown that HFMD cases are related to sociodemographic and environmental factors. This research analysed the relationship between sociodemographic and meteorological factors (temperature, rainfall, sunlight and relative humidity) on HFMD cases in Da Nang City, Vietnam.

Methods: The monthly counts of HFMD cases, population number, and meteorological factors from January 2012 to December 2016 were obtained from a reliable source. Two models, Generalized Linear Mixed Model (GLMM) and Model Based Random Forest (MBRF) were developed to evaluate the relationship between sociodemographic and meteorological factors on HFMD cases in Da Nang City, Vietnam during 2012-2016. Both models included seasonal changing climatic variables to model the seasonal (short-term) component and a function of time, i.e. month used to estimate the linear long-term trend. Similarly, a Bayesian Space Time Conditional Autoregressive (BSTCAR) model was used for the spatial effect of the residuals nested in time that arises from modelling both models.

Results: Temperature was the only factor found to be statistically significant at lag 0. Hai Chau was the district with the highest relative risk of HFMD while Lien Chieu had the lowest. Both models also identified two different seasonal patterns amount the seven districts. Pattern one showed two peaks every year, one in June (highest peak) and a small one in August, i.e. from late spring to summer or in the case of Da Nang, at the end of the dry season. It also showed two troughs, July and January, with the deepest trough in January. The other pattern showed the reverse with peaks in January (highest peak) and a small peak in July, the troughs in June and in August with the deepest trough occurring in June every year. The models also showed that there was variation in the long-term trend. Under 5 Population was found to have a negative correlation with the relative risk while its population density was positively correlated. i.e. the districts with the larger population density had the higher risk while smaller population density had the lower risk.

Conclusion: The models shows that temperature has a significant effect on HFMD in Da Nang City. This has resulted in different seasonal pattern of the risk associated with the disease within the districts. Therefore, this research provides scientific evidence that meteorological monitoring should be considered to help to fight HFMD against susceptible populations.

Keywords: Hand, foot and mouth disease, risk factor, Model Based Random Forest, Generalized Linear Mixed model, Conditional Autoregressive model.

ACKNOWLEDGEMENTS

In carrying out the task of this work, good support, guidance, motivation, health, inspirations were required, and it was with these in mind I would like to acknowledge and say thanks to some special individuals.

1. I would like to thank Almighty God for giving me life, strength, health, supporting family, good supervisors and friends to carry out such an important project as without him all this would not have been possible.
2. Secondly, many thanks go out to ITC scholarship providers for given me this opportunity to improve my academic career.
3. I must also say a special thank you to my two direct supervisors, Dr. Phuong N. Truong, Dr. Peng Jia, and their colleagues Dr. Frank Osei and Dr. Mariana Belgiu, for their guidance, patience, and encouragement through each step of the way.
4. I express my deepest gratitude to my stepfather, Michael Stoddard, for assisting me with the additional funding needed for my studies.
5. Special thanks to my other family members included my son, Andres, mother, Gweneth, sisters Gwenael and Chioniso, aunt Cher, nephew Javar, for being there for me day and night and for their continuing words of encouragement and to whom I dedicate this project too.
6. Finally, a big thank you to some very good friends especially Hannah Quammie, Sylbert Frederick, Ryan Nanton, Ruthan Monrose, Latanya Grant, Melicia Burton, Zoe Thomas, and my church family, for their continuous love, prayers, and support during this challenging period.
7. To those I may have forgotten to mention and help in some form on the other, a heartfelt thank you goes out to you all.

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
1.1.	Background and Motivation.....	1
1.2.	Research Problem.....	2
1.3.	Research Objectives.....	3
1.4.	Research Questions.....	4
1.5.	Thesis Structure.....	4
2.	Literature Review.....	7
2.1.	Spatial Epidemiology.....	7
2.2.	Disease Mapping.....	7
2.3.	Time series Data.....	8
2.4.	HFMD research.....	9
2.5.	Model-Based Random Forest research.....	11
3.	Study area, Data and Methods.....	14
3.1.	Study Area.....	14
3.2.	Research Materials.....	16
3.3.	Data Preparation.....	17
3.4.	Data Analysis.....	17
3.5.	Methodology.....	17
3.6.	Generalized Linear Mixed Model.....	19
3.7.	Model-Based Random Forest.....	23
3.8.	Model Inference.....	25
4.	Results.....	27
4.1.	Descriptive Statistics.....	27
4.2.	Estimated Spatiotemporal SMR/ Relative risk.....	35
4.3.	Model Comparison.....	53
5.	Discussions.....	57
5.1.	Incidence of HFMD in Da Nang city, Veitnam.....	57
5.2.	Predicted relative risk.....	57
5.3.	The influence of meteorological and socio-demographic factors.....	58
5.4.	Spatio-temporal residuals.....	60
5.5.	Performances of the models.....	60
6.	Limitations and Recommendations.....	62
7.	Conclusions.....	63

LIST OF FIGURES

Figure 1: Thesis Structure	5
Figure 2: Random Forest illustration, Source: (Lin et al., 2018)	11
Figure 3: Model-Based Random Forest illustration, Source: (Lin et al., 2018).....	13
Figure 4: Area of Study, Da Nang City, Vietnam.....	15
Figure 5: Study area location in Vietnam.....	15
Figure 6: Study area in Vietnam along with neighbouring countries	16
Figure 7: Methodology	19
Figure 8: Time series plot of monthly HFMD observed for each district.....	28
Figure 9: Rainfall pattern 2012-2016.....	29
Figure 10: Humidity pattern 2012-2016.....	30
Figure 11: Temperature pattern 2012-2016	30
Figure 12: Sunlight pattern 2012-2016.....	30
Figure 13: Spatiotemporal pattern of HFMD crude relative risk for each district, 2012-2016	32
Figure 14: Crude relative risk for Da Nang city, Vietnam, 2012	32
Figure 15: Crude relative risk for Da Nang city, Vietnam, 2013	33
Figure 16: Crude relative risk for Da Nang city, Vietnam, 2014	33
Figure 17: Crude relative risk for Da Nang city, Vietnam, 2015	34
Figure 18: Crude relative risk for Da Nang city, Vietnam, 2016	34
Figure 19: GLMM smooth relative risk, Da Nang City, Vietnam 2012-2016	37
Figure 20: MBRF smooth relative risk, Da Nang City, Vietnam 2012-2016.....	38
Figure 21: Crude relative risk, January 2013	39
Figure 22: MBRF smooth relative risk, January 2013.....	39
Figure 23: GLMM smooth relative risk, January 2013.....	39
Figure 24: GLMM smooth relative risk for three months in 2012	40
Figure 25: MBRF smooth relative risk for three months in 2012	41
Figure 26: Linear trend of relative risk for districts in Da Nang City 2012-2016.....	43
Figure 27: Seasonality of MBRF RR vs Seasonality of GLMM RR for Hai Chau 2012-2016.....	45
Figure 28: Seasonality of MBRF RR vs Seasonality of GLMM RR for Lien Chieu 2012-2016.....	45
Figure 29: Seasonality of MBRF RR vs Seasonality of GLMM RR for Ngu Hanh Son 2012-2016.....	46
Figure 30: Seasonality of MBRF RR vs Seasonality of GLMM RR for Son Tra 2012-2016	46
Figure 31: Seasonality of MBRF RR vs Seasonality of GLMM RR for Thanh Khe 2012-2016.....	46
Figure 32: Seasonality of MBRF RR vs Seasonality of GLMM RR for Hoa Vang 2012-2016	47
Figure 33: Seasonality of MBRF RR vs Seasonality of GLMM RR for Cam Le 2012-2016.....	47
Figure 34: Space-Time residuals for each district (MBRF).....	48
Figure 35: Space-Time residuals for each district (GLMM).....	49
Figure 36: Crude RR vs MBRF RR for Da Nang City.....	54
Figure 37: Crude RR vs GLMM RR for Da Nang City.....	55
Figure 38: Variable Importance plot for the seven districts in MBRF model.....	56

LIST OF TABLES

Table 1: Description of Model Parameters (GLMM).....	21
Table 2: Description of Model Parameters (Model-Based)	25
Table 3: Descriptive statistics of the observed cases of Hand, Foot and Mouth Disease.....	27
Table 4: Descriptive statistics of the population for each district	28
Table 5: Pearson correlations between HFMD cases and meteorological variables along with p-values (temperature)	35
Table 6: Retrospective MSE and RMSE for both models.....	36
Table 7: Maximum and Minimum Relative risk of HFMD along with the month and year observed for both models.....	36
Table 8: Estimated posterior parameters (GLMM), regression coefficients (MBRF), along with the exponent of these coefficients and its corresponding 95% credible interval	44
Table 9: Percentage of the residuals explained spatially (interaction with its neighbours) for each month for GLMM and MBRF.....	51
Table 10: Comparison of the RMSE and MSE of both models.....	53

1. INTRODUCTION

1.1. Background and Motivation

Neglected Tropical Diseases (NTDs) is a leading cause of mortality and morbidity globally. Approximately fifty percent (50%) of deaths in tropical countries and twenty percent (20%) in the Americas (Abad, Bedoya, & Bermejo, 2013) have been attributed to NTDs. NTDs have been linked to environmental (Naumova et al., 2007), poor sanitation conditions and limited access to health care (Cohen, Dibner, & Wilson, 2010). According to Jannin & Gabrielli (2013), NTDs “thrive among the poor populations in tropical countries” and account for 35,000 deaths per day in developing countries (Cohen, Dibner, & Wilson, 2010). Existing literature identified major associated consequences of NTDs such as overcrowded medical facilities, increased healthcare cost and limited productivity among public servants within the Caribbean countries (Waweru, 2018; Francis et al., 2015), thus causing tremendous strains on the health sector of these countries (Straker, 2018; Ryan et al., 2017).

The main NTD focused on in this research is Hand, Foot, and Mouth disease (HFMD), an infectious disease caused by a group of enteroviruses, including Coxsackievirus A16 (CA16) and Enterovirus 71 (EV71) (WHO, 2011); which affects children less than five years old. Its control proves to be a major challenge to societies, no vaccine or drug have been developed to date (Sarma, 2013). Thus, increased focus aims at management. Considering its severity and the absence of effective mechanisms to alleviate this disease, further research is paramount for its mitigation.

Currently, a large volume of published statistical studies regarding modelling methods to investigate the relationship of HFMD and its associated risk factors are available (Liu et al., 2015; Song et al., 2018; Zhu, Yuan, Wang, Li, Wang, Liu, Xue, Liu, et al., 2015; Liao, Qin, Zuo, Yu, & Zhang, 2016). These includes both mixed (Duan et al., 2019; Li, Qiu, Xu, & Wang, 2018) and Generalized Linear models (GLM) (Phung et al., 2018; Gou et al., 2018), both within the Bayesian and Maximum Likelihood framework. However, to my knowledge, within Vietnam and Da Nang city specifically, the combine effect of a mixed and GLM have not been carried out. Generalized Linear Mixed models (GLMM) are a “natural outgrowth of both linear mixed models and generalized linear models” (Mcculloch, 1997). They also allow for non-normally distributed responses with both fixed and random effect parameters. These random effects are done to incorporate overdispersion and correlation within the model (Mcculloch, 1997).

These are all statistical models and have varying accuracy as it results to prediction. However, ensemble methods have been mentioned among the most accurate regression tools currently available for data scientists (Breiman, 2001). Boosted regression trees (W. Zhang, Du, Zhang, Yu, & Hao, 2016), and classification and regression trees (Du, Zhang, Zhang, Yu, & Hao, 2016) were two of those methods that

were used to assess the impact of meteorological factors on HFMD. However, no research was found using Model Based Random Forest (MBRF), to assess these impacts. This machine learning algorithm provides high predictive accuracy when compared other modelling methods (Lin et al., 2018). MBRF has seen limited use in research with two pieces of literature found. Garge, Bobashev, & Eggleston, (2013) used MBRF to evaluate the efficacy of pharmaceuticals and behavioral therapies for the treatment of alcohol dependences. Lin et al., (2018) made a comparison between MBRF and stepwise regression, RF, Gaussian process, neural network and support vector machine regression as it relates to predicting Vt mean and variance based on parallel Id measurement. However, to my knowledge, MBRF has not been used in a spatial aspect. Using the residuals arising from this model, spatial autocorrelation was assessed, thus introducing a spatial structure within the model.

Therefore, the research contributes to the development of these methods, to help in estimating the spatial patterns associated with HFMD risk across areal units; help in isolating high-risk areas, with the potential to contribute to public health practice in the formulation of policy interventions for HFMD (Elliot & Wartenberg, 2004; C. Anderson et al., 2014).

1.2. Research Problem

HFMD continues to be a public health challenge worldwide in particular, the Asian Pacific region which has seen the largest outbreaks (Solomon et al., 2010). This disease was considered as a notifiable communicable disease and was recently reported as an epidemic disease in Vietnam (Phung et al., 2018). The current problem with this disease is no effective vaccine or drugs to combat HFMD has been developed (Sarma, 2013; Z. & B., 2018) and as such remains a challenge for various countries. Literatures have show that HFMD has seasonal pattern (Kim, Ki, Park, Cho, & Chun, 2016; Cheng et al., 2014; Yu et al., 2019), and are affected by Socioeconomic, sociodemographic, meteorological and geo environmental factors (Urashima, Shindo, & Okabe, 2003; Chen et al., 2014; Zhu, Yuan, Wang, Li, Wang, Liu, Xue, & Liu, 2015).

Through the use of Geographical Information Systems (GIS) methods, modelling of infectious disease data in space and time gradually increase due to the availability of such data (Bauer & Wakefield, 2018). This includes the ability to visualise and analyse epidemiological data (Liu et al., 2015; Song et al., 2018; Zhu, Yuan, Wang, Li, Wang, Liu, Xue, Liu, et al., 2015). This approach provides ample possibilities for mapping the spread of diseases and reveals the relationships with various risk factors (Sham, Krishnarajah, Ibrahim, & Lye, 2014).

To help in the fight to combat this disease and to help achieving goal 3.3 of the Sustainable Development Goals (SDG'S) (WHO, 2017), two models were developed to predict the relative risk associated with HFMD and to assess the effect of meteorological and sociodemographic factor on this relative risk in Da Nang city,

Vietnam. These two models being, GLMM and MBRF. Both models residuals or random noise were also modelled to introduce spatial structure within the models. These residuals that were not encapsulated as a result of modelling the trend associated to the covariates were modelled as spatial and non-spatial correlation (C. Anderson et al., 2014; Besag, York, & Molli, 1991). Besag, York, & Molli, (1991) proposed a conditional autoregressive (CAR) prior to model the spatial correlation and as such was adopted for this research. The models use seasonal changing climatic variables to model the seasonal (short-term trend) component, a function of time, i.e. month to estimate the linear long-term trend and the CAR model for the residuals. The log of the additive of three components of the time series representing the relative risk for the models. RMSE and MSE on calibration and validation was used to compare the accuracy and selecting the better of the two models. Through the development of these two models, this research provides scientific pieces of evidence to help to fight HFMD.

1.3. Research Objectives

This research aims at identifying and modelling the seasonality (i.e., the long-term and short-term trends) of HFMD, considering the dynamics of the population at risks and the spatial-temporal structures of the residuals. Ultimately, this research is able to better explain the seasonal patterns of HFMD and the developed models are able to forecast the seasonality in space and time given the information about the dynamics of the susceptible population and the related risk factors

1.3.1. Specific Objectives

1. To compare the accuracy of the Generalized Linear Mixed Model and Model-Based Random Forest method for mapping the relative risk of HFMD;
2. To examine the spatio-temporal distribution of HFMD on the basis of the modelled relative risk of HFMD;
3. To assess the effects of meteorological and socio-demographic factors on the relative risk of HFMD for Da Nang City, Vietnam;
4. To develop a Bayesian Space-Time Conditional Autoregressive (BSTCAR) model for the spatial-temporal random effects of the residuals.

1.4. Research Questions

The research questions outlined below are to be answered for the specific objectives to be met.

1. Specific objective 1
 - i. Which of the two models gives a lower Root Mean Squared Error (RMSE) when estimating the relative risk in space and time (calibration study)?
 - ii. Which of two models gives a lower RMSE on forecasting the relative risk in space and time (validation study)?
2. Specific objective 2
 - i. Which district within the Da Nang City has the highest and lowest relative risk of HFMD?
 - ii. Which period experienced the highest and lowest relative risk of HFMD?
 - iii. Is there a stationary seasonality or non- stationary seasonality associated with HFMD in Da Nang City?
3. Specific objective 3
 - i. Is there a positive or negative linear relationship between the meteorological, socio-demographic factors and the relative risk of HFMD?
4. Specific objective 4
 - i. What proportions of the residuals are being explained spatially and non-spatially?
 - ii. Do the residual variation follows the same pattern as that of the relative risk and what could be the reasons for such behaviour?

1.5. Thesis Structure

The thesis consists of six chapters (**Figure 1**). Chapter one is an introduction to give a background and motivation of the research. It also gives an overview of the problem statement, the objectives and questions that are needed for this research. Chapter two presents the literature review consisting of the history of disease modelling, time series data and concludes with previous studies on HFMD. Chapter 3 presents Study area, Data, Data preparation, Data analysis, Methodology and the development of the two models. Chapter 4 consists of Results; Chapter 5 consists of Discussions; Limitations and Recommendations make up Chapter 6. The final chapter, Chapter 7 is the conclusions.

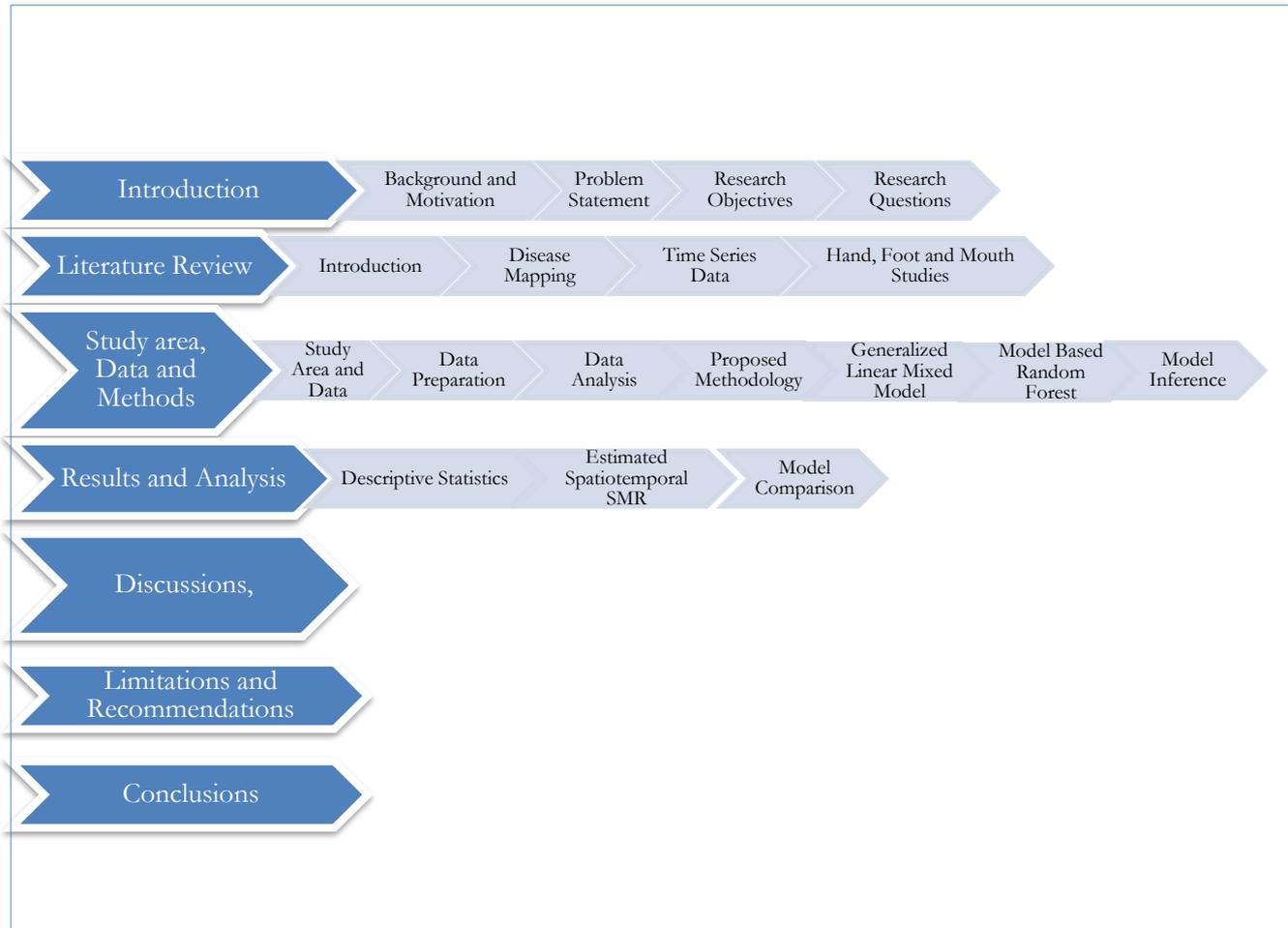


Figure 1: Thesis Structure

2. LITERATURE REVIEW

2.1. Spatial Epidemiology

2.1.1. Geographic Information systems (GIS)

Geographic Information Systems is “a tool for collecting, storing, retrieving, transforming and displaying spatial data from the real world” (Lai, Son, & Chan, 2009). Spatial data tells us about their spatial location in the real world, dimension (Burrough & Mcdonnell, 1998) and is stored as geometry (Lai et al., 2009). Non-spatial data are also important in GIS as they provide descriptive information related to the spatial features mapped (Lai et al., 2009). Through a unique identifier, Spatial and Non- spatial data are related (Lai et al., 2009). Storing the spatial data as a Vector or Raster gives one the ability to visualize spatial information, make maps, analyse spatial information and to ask critical questions of the map and its data (Lai et al., 2009).

Elliot & Wartenberg, (2004) defined Spatial Epidemiology as “the description and analysis of geographically indexed health data with respect to demographic, environmental, behavioural, socioeconomic, genetic, and infectious risk factors.” These authors also mentioned at a small area scale, Spatial epidemiology can be divided into three main areas: 1. disease mapping, 2. geographic correlation studies, 3. clustering, disease clusters, and surveillance. As this thesis focuses on disease mapping, one section would be dedicated to review the history. With increase technology, reports of disease outbreaks are easily collected along with corresponding risks factors and the population at risk. Therefore, spatial analysis can be performed on health data (Lai et al., 2009).

2.2. Disease Mapping

Research into disease mapping has a long history dating back to the 19th century with further developments taking place in the 21st century (Walter, 2001). Through advancement in data availability and analytical methods, reporting of disease at both national or regional scale has vastly improved (Elliot & Wartenberg, 2004). This field of spatial epidemiology focuses on estimating the spatial patterns associated with disease risk across areal units. The aim is to identify high risks areas. Based on that public health interventions are made (C. Anderson et al., 2014). Disease maps are also used for descriptive purposes, etiology hypotheses, as surveillance to accentuate high-risk areas and help policy formation (Elliot & Wartenberg, 2004). A considerable amount of literature has been published on disease mapping. Generally, these studies showed standardized mortality or incidence ratios (SMRs) of different geographic areas, i.e. regions, districts, counties or countries (Elliot & Wartenberg, 2004).

Due to confidentiality, disease incidence data are generally available as summary counts or rates for a well-defined region such as a district, municipality or county etc. and at a finite period (Waller, Carlin, Xia, & Gelfand, 1997). A typical likelihood model for count data is a realization from conditionally independent Poisson distributions with the means as the product of the relative risk and expected cases (Waller et al., 1997). Elliot & Wartenberg, 2004, stated that the rate in an area is estimated as the standard mortality ratio (SMR_i). This was calculated as O_i / E_i , where O_i is the observed number of deaths or incident cases of a disease in the area; E_i is the expected number of cases (calculated by applying age- and sex-specific death or disease rates to population counts for the area) (Elliot & Wartenberg, 2004). The expected count is given as $N_i * \theta$, where θ is the overall disease risk and N_i as the population for district i (Waller et al., 1997). Through internally standardized (obtained from the given data) or externally standardized (external source), the overall risk is obtained (Waller et al., 1997).

2.3. Time series Data

Time series data can be decomposed into trend, season and random noise (Barnett & Dobson, 2013).

A trend is defined as "the long-term change in disease, representing a gradual improvement or worsening in disease frequency." (Barnett & Dobson, 2013). Previous literature has modelled the trend as function of time (Barnett & Dobson, 2013; Yu et al., 2019). Modelling the seasonality of diseases has been of great concern over the years. It was Hippocrates in 400 BC who stated that "all diseases occur at all seasons of the year, but certain of them are more apt to occur and exacerbate at certain seasons" (Barnett & Dobson, 2013). A seasonal pattern can be stationary or non-stationary. A stationary seasonal pattern does not change from season to season while a non-stationary changes over time (Barnett & Dobson, 2013).

Investigating the seasonal pattern of diseases has been done using different models. Dynamic Harmonic Regression Model (Sofianopoulou, Pless-Mulloli, Rushton, & Diggle, 2017), Auto-Regressive Moving Average (ARMA) model, Multivariate linear regression model, Generalized Additive Models (GAM), Artificial Neural Networks (ANN) (Baquero, Santana, & Chiaravalloti-Neto, 2018), Spatiotemporal Bayesian models and Functional Generalized Least Squares Regression (Basile, Oviedo de la Fuente, Torner, Martínez, & Jané, 2018) are some examples that have been used for disease modelling. Each model has its advantages and disadvantages.

2.4. HFMD research

Existing studies have shown that HFMD cases are related to socioeconomic, sociodemographic, meteorological and geo environmental factors. Historically, research investigating the factors associated with HFMD have showed that there exist a linear association or relationship between HFMD and meteorological factors in different regions (Urashima, Shindo, & Okabe, 2003; Chen et al., 2014). One of the main meteorological factors that has inconsistent relationship according to prior studies is temperature (Urashima et al., 2003; Hii et al., 2011; Huang et al., 2013). In Singapore, a study (Hii, Rocklöv, & Ng, 2011) revealed that an increase of 1°C in temperature above 32 °C was significantly associated with a 36% increase in HFMD incidence. In contrast, a negative association was revealed in a study for Shangong, China, when the average temperature was above 21 °C (Zhu, Yuan, Wang, Li, Wang, Liu, Xue, & Liu, 2015). Other studies have also revealed that sunshine and windspeed have associations with HFMD (W. Zhang, Du, Zhang, Yu, & Hao, 2016b). Likewise, an association between rainfall and HFMD was revealed by Wu et al., (2017) and between humidity and HFMD (Phung et al., 2018). Normalized difference vegetation Index (NDVI), land cover, roadway density and population density were also found to have an influence on HFMD (Stanaway, 2013; Song et al., 2018).

The relationship between HFMD and these meteorological factors have also shown in some articles to be non-linear (Chen et al., 2014; Wu, Wang, Wang, Xin, & Lin, 2014). As it relates to techniques used for predicting nonlinear relationship, Random Forest time series models were used for prediction of avian influenza outbreaks in Egypt (Kane, Price, Scotch, & Rabinowitz, 2014). Also, Support Vector machine, bagging, boosting and Random Forest was used to predicting eight chronic diseases, with Random Forest ensemble learning method outperforming the rest (Khalilia et al., 2011). Distributed lag nonlinear model (Zhu, Yuan, Wang, Li, Wang, Liu, Xue, Liu, et al., 2015; Liao, Qin, Zuo, Yu, & Zhang, 2016), boosted regression trees (W. Zhang et al., 2016a), classification and regression trees (Du et al., 2016) were all used for predicting nonlinear relationship between HFMD and its risk factors.

Other published studies regarding modelling methods used to investigate the relationship of HFMD and its associated risk factors includes: the generalized additive model (Chen et al., 2014), Bayesian network (Liu et al., 2015; Song et al., 2018), geo-additive mixed spatiotemporal model (L. Li et al., 2018), boosted regression trees (W. Zhang et al., 2016a), classification and regression trees (Du et al., 2016), Distributed lag nonlinear model (Zhu, Yuan, Wang, Li, Wang, Liu, Xue, Liu, et al., 2015; Liao, Qin, Zuo, Yu, & Zhang, 2016), among others.

Liao, Qin, Zuo, Yu, & Zhang, (2016) assessed the effect of extreme meteorological factors, air pollution indicators and effects of different lag days on HFMD incidence in Guilin city using Distributed lag nonlinear models (DLNM) with natural cubic spline used to model the nonlinear relationship between meteorological or air pollution variables, and time as the indicator to control term trends, seasonality and differences in the

annual at-risk population. The evidence presented in this research suggests that extreme temperatures, high precipitation and low ozone concentration increases the risk of HFMD.

Phung et al., (2018) examined the province specific association between monthly HFMD and climatic factors, while controlling for spatial lag, seasonality and long-term trend, using a Generalized Linear Model with Poisson family. The results showed that the climate-HFMD relationship varied by regions and provinces across Vietnam. Time series regression was used to examine temporal patterns of HFMD and climate factors by Nguyen et al., (2017). Seasonality and long-term trends were controlled by using a flexible spline function with a Generalized Linear Model with Poisson family used to examine the lag effect of each climatic factor. The results showed that the climate-HFMD relationship varied at different lag days with the Mekong Delta region in Vietnam.

Truong & Stein, (2018) proposed a hierarchically adaptable spatial regression model to link aggregated health data and environmental data in Da Nang City, Vietnam. This model links misaligned health and environmental data especially when health data are available at larger aggregation levels than the environmental data (Truong & Stein, 2018). In Vietnam, cases of HFMD are reported at a district level, ranging between 10 to 10^3 km²; while the environmental risk factors such as daily air temperature and humidity are regularly recorded at only one or two meteorological monitoring stations per province with an average area of about 5×10^3 km² (Truong & Stein, 2018). Most of these studies either focus on space or time but not both simultaneously. Other spatiotemporal models were used to detect clustering of HFMD incidences (Y. Liu et al., 2013; Deng et al., 2013; Wang et al., 2016).

Bernardinelli et al., (1995), proposed a Bayesian model in which both the area specific intercept and trend are modelled as random effects and correlation between them is allowed for, to estimate the cumulative prevalence of insulin-dependent diabetes of 18-year-old military recruits born in Sardinia between 1936 and 1971. Models of such type are Poisson Generalized Linear Mixed models: generalized because its error distribution is other than Normal and mixed because its linear predictors contain both fixed and random effects parameters. These models are under the general class of the Generalized Linear Mixed Model (Breslow & Clayton, 1993). The type of GLMM model used is a Poisson Log normal as this allows us to model the random effect of the relative risk based on the covariates while also accounting for spatial autocorrelation within the residuals (Wakefield, 2007). With regards to spatiotemporal residuals, Waller et al., (1997), proposed a heterogeneity and spatial effected nested within time which allows the examination of the evolution of heterogeneity and spatial patterns over time.

2.5. Model-Based Random Forest research

According to Breiman, (2001), Random Forest (RF) is constructed through an ensemble of decision trees and is used for classification or regression (CART) methods. CART is a commonly used recursive partitioning method, that selects the most important variable from a large number of variables to explain the outcome variable and successively splits the data to identify groups of observations with similar values of the response variable (Breiman, 1994). The following algorithm was defined for a Regression Tree (Garge et al., 2013):

Regression Tree Algorithm

- Several bivariate association models are run using all predictor variable
- The strongest association with the response variable is selected
- The data is split into two or more subgroups on the optimal cut point in the selected predictor (partitioning variable). This point is the one which leads to the greatest possible reduction in the Residuals Sum of Squared (RSS)
- Each subgroup form by such split is called leaves or nodes.
- This continues until nodes contain observations only of one class. i.e. cannot be split further, no predictor variable shows strong association within a given node and number of observations within the node are less than what was specified.

One major disadvantage of CART is that it is a single tree method, thus making it unstable to small changes in the learning data. To overcome this disadvantage, methods such as Random Forest and Bagging were introduced (Strobl, Malley, & Tutz, 2009). Through an ensemble of decisions trees, these methods can greatly improve stability and prediction accuracy.

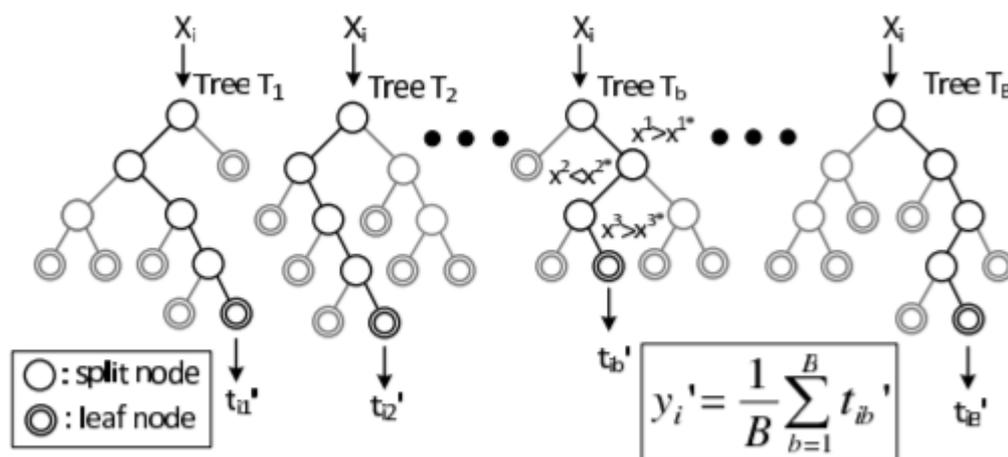


Figure 2: Random Forest illustration, Source: (Lin et al., 2018)

Random Forest makes use of a random selection of the predictor variables during the process of splitting a tree node and constructs trees on random samples of learning data with replacement term bootstrapping (Garge et al., 2013), i.e. each tree is constructed based on a different set of training samples.

Using Random Forest, multiple trees can be constructed at different levels while incorporating predictions that are not as strong as the dominant predictors (Strobl et al., 2009). **Figure 2** illustrates the theory of Random forest for B amount of decisions trees, [T1, T2, ..., TB]. Predictions of the response variable from RF is based on the averaging of the response variable obtained from all trees (Lin et al., 2018). Breiman (2001) stated "Random Forest are A+ predictors but their mechanism for producing a prediction is difficult to understand." This is because random forest models provide a functional measure of the influence that each variable has on accuracy without providing an interpretable measure of how the variable helps to determine the predictions (Kane et al., 2014). Although random forest has shown to provide high prediction accuracy, Lin et al., (2018) have shown that extension of random forest, termed Model-Based Random Forest (MBRF) has even greater prediction accuracy. A comparison between MBRF and stepwise regression, RF, Gaussian process, neural network and support vector machine regression has found that MBRF has the lowest RMS and highest R^2 when a prediction was carried out (Lin et al., 2018). Due to this major advantage and the nature of the data available, MBRF was selected for this research.

Lin et al., (2018) has outlined the two major differences between MBRF and RF. These were 1. MBRF applies the least squares method to build a linear regression model on the same preselected m modelling features for each leaf node of a tree to predict the response feature of a testing sample, instead of using the average of the response features inside a leaf node as in RF and 2. MBRF applies Model-Based Recursive Partitioning (MBRP) to choose a proper partition feature and its split criterion for each split node, instead of using CART as in RF.

This was also illustrated in a diagram as seen in **Figure 3**.

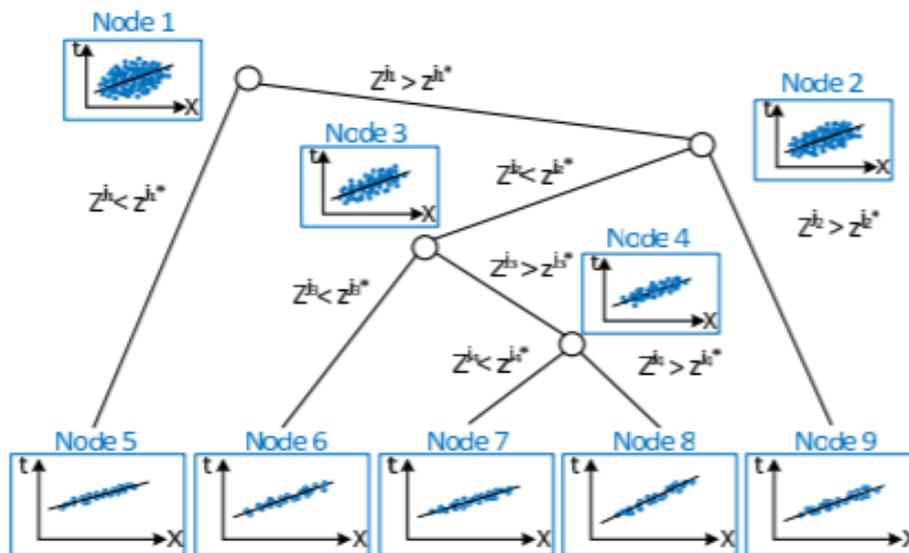


Figure 3: Model-Based Random Forest illustration, Source: (Lin et al., 2018)

As it relates to MBRP, Garge, Bobashev, & Eggleston, (2013) stated it splits group of observations with similar model trends as compare to splitting group of observations that shows similar values of the response variable (CART). For linear regression, MBRP groups the feature space to recognize different subgroups with similar effects. Predictions of the response are based on the different subgroups estimated effects (Garge et al., 2013). This method also makes use of a single tree, thus making it unstable to small changes in the learning data. To stabilize the predictions and improve accuracy, predictions on multiple tree models, MBRF and Bagging were introduced (Garge et al., 2013).

Two pieces of literature have been found using Model-Based Random Forest. Garge, Bobashev, & Eggleston, (2013) used MBRF to evaluate the efficacy of pharmaceuticals and behavioral therapies for the treatment of alcohol dependences. Lin et al., (2018) made a comparison between MBRF and stepwise regression, RF, Gaussian process, neural network and support vector machine regression as it relates to predicting V_t mean and variance based on parallel I_d measurement. In this study, through the MobForest package in R (Garge, Eggleston, & Bobashev, 2018), this model was developed to predict the relative risk of HFMD associated with the meteorological factors and socio-demographic factors.

3. STUDY AREA, DATA AND METHODS

3.1. Study Area

HFMD is currently a worldwide health problem, especially within the South East Asian region. The study area resides in that region. Da Nang City (Figures 4-6) is the fourth largest city related to urbanization and economy and is seen as the largest city in the central coast of Vietnam (General Statistics Office Of Vietnam, 2017). The latitude and longitude for this region are 16.0544° N, 108.2022° E. As of 2016, the average population per thousand was 1,046.2 and has an area of 1,284.7 km² (General Statistics Office Of Vietnam, 2017). This area consisted of 7 districts, one urban and 6 rural (Truong & Stein, 2018) and two seasons, dry (November- April) and rainy (May- October) (General Statistics Office Of Vietnam, 2017).

The first reported cases of HFMD outbreak in Vietnam was for the year 2003 (WPRO, 2011) with the majority of reported and fatal cases being children within five years old (Truong & Stein, 2018). In Vietnam, one of the largest outbreak of hand, foot and mouth disease occurred in 2011, which resulted in 170 deaths among the 113,121 infected children (Nguyen et al., 2014). It was also reported for the first quarter of 2015 in Vietnam, 72.8% of the cases were for the southern part, however, Da Nang city, located in central Vietnam recorded the largest number of cases, 36.8% (WPRO, 2015).

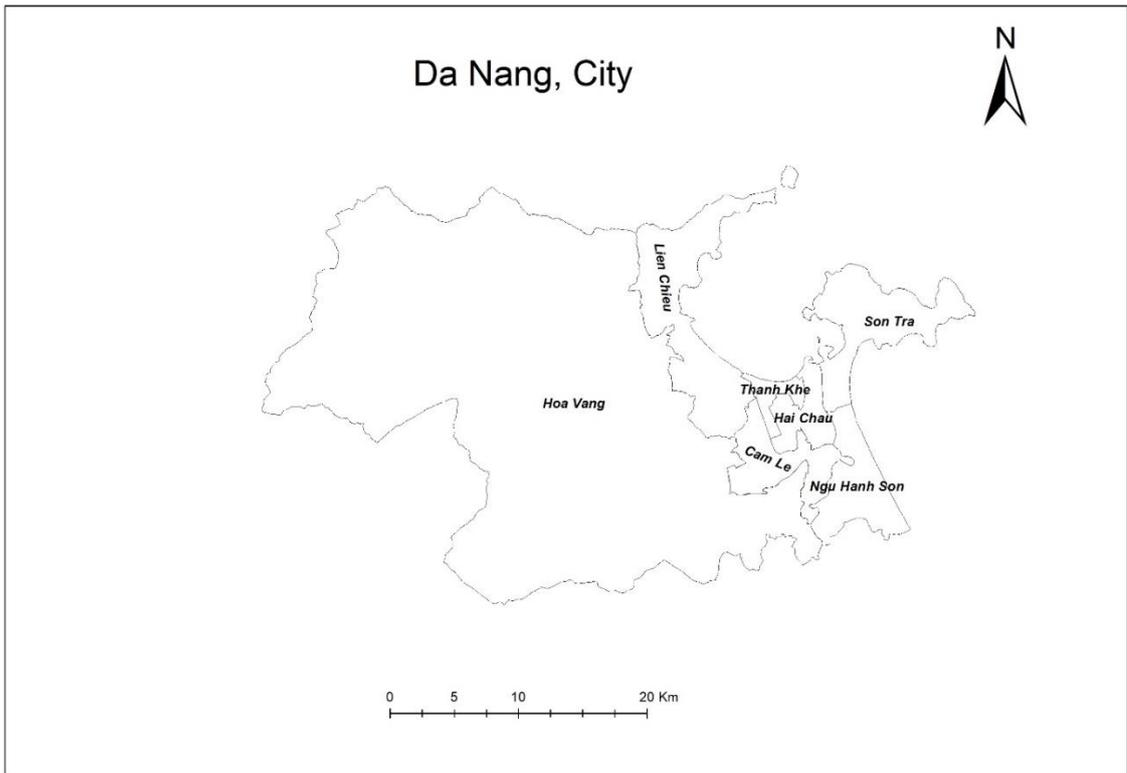


Figure 4: Area of Study, Da Nang City, Vietnam



Figure 5: Study area location in Vietnam



Figure 6: Study area in Vietnam along with neighbouring countries

3.2. Research Materials

Data

Tools required for this research included

- Monthly Hand, Foot, and Mouth summary data between 2012- 2016 for Da Nang City
- Monthly population, average rainfall, average humidity, average sunlight, the average temperature for 2012- 2016 (Calibration data)

- Monthly population, average monthly temperature, weekly observed cases of Hand, Foot and Mouth cases 2017 (Validation data)
- The centroids of the districts were used as the spatial references of HFMD as the exact location of these outbreaks were not recorded.

Softwares

- GIS software, ArcGIS 10.5.1
- R Studio
- WinBUGS
- Microsoft packages

3.3. Data Preparation

To carry out our modelling and analyse in R, the data obtained were reassigned a unique identifier in ArcGIS that was readable for WinBUGS. Also, calculations were done in Excel to calculate the expected and observed cases for each month of each district. These tables were stored as a CSV file and read into R using the `read.csv` command.

3.4. Data Analysis

To identify the effects of the covariates (rainfall, relative humidity, sunlight, temperature) on relative risk, a linear model of covariates against observed cases at both the district level and city level was carried out to show the covariate(s) that is/are statistically significant. Also, the correlation between each covariate and the correlation between the crude relative risk with the covariates at both district level and city level was assessed for the collinearity problem. To estimate the linear long-term trend, a month was used as a dummy variable with values of 1 to 60 representing the months of 5 years. Results of this analysis are presented in the Results section 4.2.

3.5. Methodology

Disease incidence data are generally available as summary counts or rates for a well-defined region such as a district, municipality or county etc. and a finite period (Waller et al., 1997). These are observed from different subgroups of the population within the region. In previous literature, it assumes that summary count data arise from the associated relative risks, i.e. there exists a likelihood model for the count data given the relative risk (Waller et al., 1997). The typical likelihood model that is assumed for count data is a realization that follows the conditionally independent Poisson distribution with the mean as the product of

the relative risk and expected cases (Waller et al., 1997). In the Bayesian Modelling approach, suitable priors for the relative risk are needed to smooth the crude map that arises from the likelihood model as these maps often have large variations for sparsely population regions, i.e. areas with the highest relative risk will have high uncertainties (Waller et al., 1997).

With related to time series data, it can be decomposed into trend, season and random noise (Barnett & Dobson, 2013). In modelling the trend or the fixed effect associated with the relative risk for diseases, it can be modelled linearly or non-linearly as a function of time, i.e. monthly within the given time series. Likewise, the residuals or random effects can be modelled as a spatially structured effect and an unstructured/non-spatial effect.

Therefore, in this study, a Generalized Linear Mixed model and Model-Based Random Forest model, were developed to evaluate the impact of meteorological and socio-demographic factors on cases of HFMD and predict the relative risk of seven districts within Da Nang City, Vietnam for the period 2012 to 2016. These models were compared based on RMSE and MSE for prediction (present data, 2012-2016) and validation (2017) to determine which model gives the better accuracy. The results showed which model predicts better based on past data and can predict for a future outbreak.

Similarly, a Bayesian Space-Time Conditional Autoregressive (BSTCAR) Model was used to model the spatial effect of the residuals nested in time that arises from modelling the GLMM model and MBRF model. Both models included seasonal changing climatic variables to model the seasonal (short-term) component, a function of time, i.e. month used to estimate the linear long-term trend and the (BSTCAR) model for the residuals. The log of the additive of three components of the time series representing the relative risk. Figure 7 below shows the flow chart of the stated methodology.

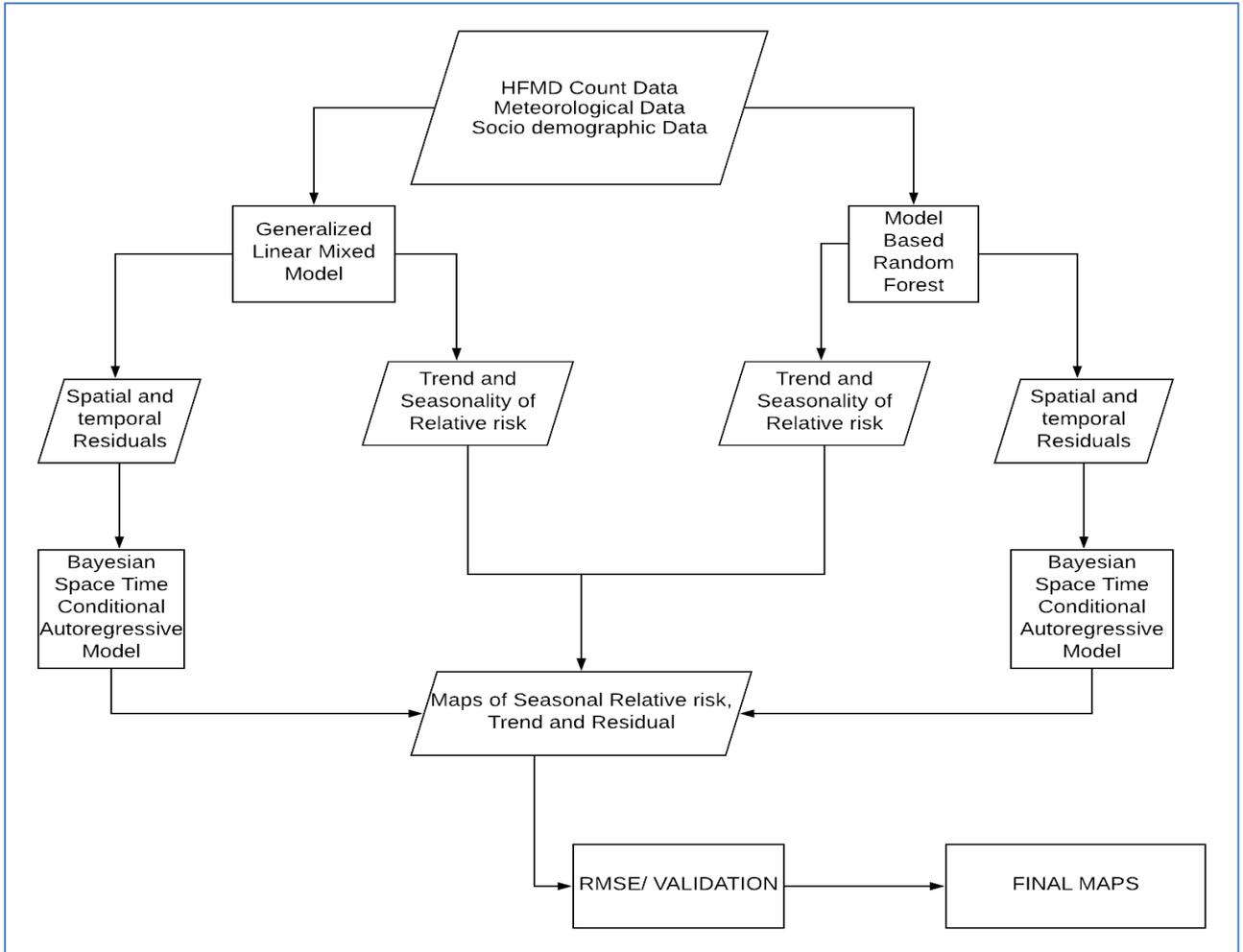


Figure 7: Methodology

3.6. Generalized Linear Mixed Model

3.6.1. Model Specification

As stated above, the disease incidence is count data, i.e. the number of cases in each district per month. This shows the observed cases or the occurrences of HFMD in different districts in Da Nang City. However, due to the different population size among the districts, it cannot reflect the risk of contracting the disease. In literature, it assumes that summary count data arise from the associated relative risks, i.e. there exists a likelihood model for the count data given the relative risk (Waller et al., 1997). This relative risk (RR) or Standard Mortality Rate (SMR) map can be unstable as a result of low event counts or the population at risk, with small changes in the case resulted in a dramatic shift. i.e. they feature large outlying relative risks in districts where the population is sparse (C. Song et al., 2018). In the Bayesian Modelling approach, suitable priors for the relative risk are needed to smooth the crude map that arises from the likelihood model as

these maps often have large variations for sparsely population regions, i.e. areas with the highest relative risk will have high uncertainties.

The typical likelihood model for count data is that a realization follows the conditionally independent Poisson distribution with the mean as the product of the relative risk and expected cases (Waller et al., 1997). For each district in each month, we denoted the observed cases as O_{it} , where $i = 1, \dots, 7$ index the districts and $t = 1, \dots, 60$ index the months. The population for each district at each month was denoted as N_{it} . The risk per district per month is equal to O_{it}/N_{it} , (1), however, our interest is in the relative risk, θ_{it} , which is given as $O_{it}/E_{it}, \dots$ (2), where E_{it} is the expected count in district i at time t . This expected count is given as $N_{it} * \theta_t$, where θ_t is the overall disease risk for a period. The overall risk was internally standardized as it was obtained from the given data $\theta_t = (365.25 * \sum_{i=1}^7 O_{it}) / (12 * D_t * \sum_{i=1}^7 N_{it}) \dots \dots \dots$ (3) where D_t is the number of days per month (Barnett & Dobson, 2013). This adjusted rate was used to account for unequal number of days for each month, as the duration of exposure is likely to change.

Thus, given θ_{it} , the observed cases, $O_{it} \sim \text{Poisson}(E_{it} \theta_{it}) \dots \dots \dots$ (4)

The data can exhibit extra Poisson variation/ over-dispersion because of the within area variance being greater than the expectation of a Poisson distribution. Therefore, to control this overdispersion, one way is to model the relative risk as a random effect. The model chosen to do such was the Poisson Log- Normal Model as it allows for covariate adjustments and accounts for the anticipated similarity of the relative risk in nearby or adjacent regions (Montesinos-López et al., 2017).

The Bayesian modelling approach was used instead of the Maximum likelihood approach because of the sparse data in geographical space (Waller et al., 1997). This approach involves two stages, a likelihood model for the vector of observed counts O_{it} given the vector of the relative risk θ_{it} , and a prior model to be updated by this likelihood model (Waller et al., 1997). The software WinBUGS allows for MCMC computational algorithms and yielding a posterior of the relative risk given the observed cases (Lykou & Ntzoufras, 2011) was used to carry out the Bayesian modelling.

To allow for different error distribution than a Normal with both fixed and random effect parameters, the Generalized Linear Mixed Model was used. This model was developed in the Bayesian Framework through the WinBUGS software. The other analyses and visualizations were carried out in R.

3.6.2. Model Formula

$$\text{Log}(\theta_{it}) = \beta_{0i} + \sum_{j=1}^C \beta_{ij} X_{ijt} + \gamma_i X_t + \phi_{it} + \psi_{it} \dots \dots \dots (5)$$

$$\psi_{it} | \psi_{jt \neq it} \sim N\left(\frac{\sum_{j \neq i} w_{ij} \psi_{jt}}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_{\psi}^2}{\sum_{j \neq i} w_{ij}}\right) \dots \dots \dots (6)$$

Table 1: Description of Model Parameters (GLMM)

Index	Variable	Prior distribution
i, t, j, C	Indexes of Districts, months, regression coefficients, and number of seasonal covariates	N/A
$\log(\theta_{it})$	Log of the relative risk	N/A
β_{0i}	Fixed intercept for each district	Normal distribution with mean zero: $\beta_{0i} \sim N(0, \sigma_{\beta_{0i}}^2)$
β_{ij}	The regression coefficient for seasonal covariates for each district	Normal distribution with mean zero: $\beta_{ij} \sim N(0, \sigma_{\beta_{ij}}^2)$
X_{ijt}	Fixed seasonal covariates for per district per month	N/A
γ_i	The regression coefficient of the trend for each district	Normal distribution with mean zero and variance. $\gamma_i \sim N(0, \sigma_{\gamma_i}^2)$
X_t	Dummy covariate indicating the number of months over a five-year period	N/A
ϕ_{it}	Spatial-temporal unstructured	Normal distribution with mean zero and variance. $\phi_{it} \sim N(0, \sigma_{\phi}^2)$
ψ_{it}	Spatial temporal structured	Conditional Autoregressive prior $\psi_{it} \psi_{jt \neq it} \sim N\left(\frac{\sum_{j \neq i} w_{ij} \psi_{jt}}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_{\psi}^2}{\sum_{j \neq i} w_{ij}}\right)$
$\sigma_{\phi}^2, \sigma_{\psi}^2$	Hyperpriors for the spatial unstructured and structured	Gamma distribution

The two spatial random effects, one followed an independent Gaussian Exchangeable prior to model the unstructured heterogeneity effect nested in time, while an Intrinsic Conditional Autoregressive (CAR) prior was assumed for the spatial structured variability (Waller et al., 1997), i.e. spatial dependence was at its maximum, to update the data and give a posterior distribution of the resulting coefficient (Bernardinelli et al., 1995). When this is used, there is no parameter that controls the strength of the spatial correlation with the assumption being there is a strong spatial correlation (Craig Anderson & Ryan, 2017). This was given by the following:

$$\psi_{it} | \psi_{j \neq i} \sim N\left(\frac{\sum_{j \neq i} w_{ij} \psi_{jt}}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_{\psi}^2}{\sum_{j \neq i} w_{ij}}\right) \dots\dots \text{CAR model for Spatial structured random effects}$$

where i, j indicates that areas i and j are neighbors, w_{ij} indicates whether the districts share boundaries with 1 if region i shares a common edge or border with region j , 0 if $i = j$, 0 otherwise and σ^2 is the variance component (Wall, 2004; Song et al., 2018). Besag, (1974) proposed a CAR prior for spatial dependences with a Gaussian distribution where each spatial effect ψ_i is conditional on its neighbors ψ_j . The variance for such districts depends on the number of neighboring districts and made use of the assumption that disease incidence risk in a spatial area was derived from areas that are geographically near to each other (Besag, 1974). It makes use of Tobler’s law of geography, “everything is related to everything else, but near things are more related than distant things” (Tobler, 2009). As it relates to spatial, this character is called spatial autocorrelation, where it assumes the closer the distances in space are, the similar the disease incidence relative risk will be in those spatial areas (Vieira et al., 2008; Segurado, Araújo, Kunin, & Segurado, 2006).

Both spatial-temporal unstructured and structured are assigned priors. These include the variance parameters with the spatial-temporal unstructured or random effect in time (σ^2_{ϕ}) is on the log of the relative risk and the spatial-temporal structured or spatial effect in time σ_{ψ}^2 is on the log relative risk conditional on the spatial effect of the neighbors (Wakefield, 2007). When assigned the CAR, the prior for the precision of the parameter must also be given. The variance is then given as the one divided by the precision. This hyperprior was assigned from the gamma distribution, $\text{dgamma}(0.5, 0.0005)$, as these are vague and allows the model to get most of its information from the data (Law, 2016).

3.7. Model-Based Random Forest

As stated above, Random Forest (RF) provides a functional measure of the influence that each variable has on accuracy of predictions without providing an interpretable measure of how the variable helps to determine the predictions (Kane et al., 2014), but more focus on the importance of the variable (Kane et al., 2014; Garge, Bobashev, & Eggleston, 2013). Similar characteristics of MBRF were observed, i.e. the MBRF function in the MobForest package is unable to give an interpretable measure or the coefficients of the variables. Therefore, to obtain an interpretable measure or the coefficients of the variables, the linear function or model-based recursive partitioning (MRBP) function can be used. This is because as mentioned previously, MBRF uses the ordinary least squares method (linear regression method or `lm` function) to build the model on selected modelling features while also using model-based recursive partitioning to select a proper partition feature and its splitting criterion for each split node (Lin et al., 2018).

Using the MobForest package within R (Garge et al., 2018), the main function `mobForestAnalysis` for analysis of MBRF functions was used to predict the relative risk of HFMD within Da Nang City. The `mob_rf_tree` function for MBRP from the same package was also used to decompose the predicted related risk into the trend component and the seasonal component.

3.7.1. `mobForestAnalysis ()`: Model-Based Random Forest

Model-based trees to incorporate random forest methodology uses a random subset of the partitioning variables when selecting the process of splitting a tree node. Through this subset, the variable that results in the lowest residual sum of square errors is selected as a splitting variable (Kane et al., 2014; Garge, Bobashev, & Eggleston, 2013). As it relates to Regression, the feature space is a partition to identify similar effects of the covariates and the predicted responses is based on the estimated effects within the different groups. This is the main function for analysis for MBRF. The following takes the necessary arguments to start model-based random forest analysis.

```
mobforest.analysis (formula, partition_vars, data, mobforest_controls = mobforest.control (ntree, mtry,
replace, alpha, bonferroni, minsplit), new_test_data = as.data.frame(matrix (0, 0, 0)), processors = 1, model
= linearModel, family = NULL, prob_cutoff = NULL, seed = sample (1:1e+07, 1)).
```

The `mobForestAnalysis ()` function was used to predict the relative risk based on temperature and the dummy variable. As mentioned previously, this function does not allow the separability of the trend and seasonality, thus the regression coefficients were not obtained. Therefore, to obtain the regression coefficients and to separate the trend and seasonality, MBRP was used. Another function that was used for the same purpose is the linear regression model via the `lm` function.

3.7.2. mob_rf_tree (): Model-Based Recursive Partitioning

Model-based recursive partitioning identifies groups of observations whose parameters of interest are similar and gives predictions from a single tree (Garge et al., 2013). The advantage and the main reason that this function was used was to get the regression coefficients of the covariates in order to define the model. However, a major disadvantage is because it is a single tree model, predictions are very sensitive to small changes. The following takes the necessary arguments for Model-based recursive partitioning, a randomized subset of partition variables considered during each split: `mob.rf.tree (formula, partition_vars, mtry, data, model, control = mob_control (), ...)`.

Carefully analyzing both functions, the major difference is in the `mobforest.control` which contains an argument called `ntree`. This argument represents the number of trees to be constructed in a forest and replaces `which` was set to `true`, thus build the trees on random samples of learning data with replacement (bootstrap).

3.7.3. Model Formula

$$\log(\theta_{it}) = \sum_{j=0}^C \beta_{ij} X_{ijt} + \gamma_i X_t \dots\dots\dots (7).$$

See table 2 below for a description of each component of the model.

3.7.4. Bayesian Conditional Autoregressive model for MBRF Residuals

The model in Equation 7 was used to model the seasonality and the trend combined. The residuals arising from this model were obtained and assessed for spatial-temporal structured and spatial-temporal unstructured residuals. A Conditional Autoregressive prior was assigned for spatial-temporal structured and a Normal distribution prior was assigned for the unstructured residuals. $e_{it} = N (M_{it}, \sigma^2_e)$, where $M_{it} = \phi_{it} + \psi_{it}$.
 (8)

Overall model

$$\log(\theta_{it}) = \sum_{j=0}^C \beta_{ij} X_{ijt} + \gamma_i X_t + \phi_{it} + \psi_{it} \dots\dots\dots (9)$$

The overall relative risk was obtained by taking the exponent of the sum of the predicted relative risk from the Model-Based Random Forest model and the residuals that arises from the spatial structured and unstructured effects.

Table 2: Description of Model Parameters (Model-Based)

Index	Variable	Prior distribution
i, t, j, C	Indexes of Districts, months, regression coefficients, and number of seasonal covariates	N/A
$\log(\theta_{it})$	Log of the relative risk	N/A
β_{ij}	The regression coefficient for seasonal covariates for each district	N/A
X_{ijt}	Fixed seasonal covariates for per district per month	N/A
γ_i	The regression coefficient of the trend for each district	N/A
Xt	Dummy covariate indicating the number of months over a five-year period	N/A
ϕ_{it}	Spatial-temporal unstructured	Normal distribution with mean zero and variance. $\phi_{it} \sim N(0, \sigma^2_\phi)$
ψ_{it}	Spatial temporal structured	Conditional Autoregressive prior $\psi_{it} \psi_{j \neq i} \sim N\left(\frac{\sum_{j \neq i} w_{ij} \psi_{jt}}{\sum_{j \neq i} w_{ij}}, \frac{\sigma^2_\psi}{\sum_{j \neq i} w_{ij}}\right)$
σ^2_ϕ and σ^2_ψ	Hyperpriors for the spatial unstructured and structured	Gamma distribution

3.8. Model Inference

3.8.1. GLMM

The spatiotemporal model was formalized within the Bayesian Framework to include priors to update the likelihood of the data. Using the GLMM, both fixed and random effects were incorporated into the model. The random effects allow the model to incorporate similarities of neighborhoods in space and time while the fixed effects allow the model to incorporate the effects of the covariates. This model includes three main levels; distribution of the data, spatiotemporal process and the parameters. The likelihood model for the data distribution level was Poisson. As it relates to the spatiotemporal process level, spatial-temporal structured and spatial-temporal unstructured were combined to account for the spatial effect and random effects for each month. This model was proposed by Waller, Carlin, Xia, & Gelfand, 1997, where heterogeneity and spatial effects were nested in time. This will allow the examination of these residuals over time. Finally, the parameters which are the variances were assigned an inverse gamma distribution prior

within the Bayesian Framework. The main value of the inverse gamma distribution used was $d\text{gamma}(0.5, 0.0005)$, as these are non-informative and allows the model to get most of its information from the data (Law, 2016). The model was built in Notepad+ and ran in WinBUGS and R using the R2Winbugs package. Three Markov chains with the initial values for the parameters set to zero was used to build the model. A total of 100,000 simulations each with the first 25,000 discarded, though efforts were made to assess the sensitivity based on no. of iterations, no. of chains, and burning length. The selected combination of simulations and discarded amount resulted in the lowest RMSE among those assessed. The retrospective and prospective RMSE of the model was calculated to be used for model comparison. Also, trace plots for the parameters were visualized for convergence and were stored.

3.8.2. MBRF

As stated prior, the MobForest package in R was used to build this model. The `mobForestAnalysis()` function was used to predict the relative risk based on temperature and the dummy variable. Apart from the required formula (equation 7), the function requires the partition variables to split the data into two or more subgroups. The variable which leads to the greatest possible reduction in the Residuals Sum of Squared (RSS) is selected to construct the node (Garge et al., 2013). Other important parameters in the function were `mtry`: the number of input variables that should be randomly sampled at each node; `ntree`: the number of trees to be built in the forest; `minsplit`: the number of observations in a node; `model`: the type of model used for fitting the observations, be it a linear model or generalized linear model and finally `family`: a description of the error distribution and link function that the model will be used. For this given research, the two predictors were used as partition variables, while 300 trees were built in the forest. The number of input variables that should be randomly sampled at each node was set to one with the minimum number of observations set at 30. The model build was a linear model with a Poisson distribution.

The residuals arising from this model were extracted and assessed for spatial-temporal structured and spatial-temporal unstructured residuals. Similar to the GLMM, the model proposed by Waller, Carlin, Xia, & Gelfand, 1997 was used to model for the residuals. This was built in Notepad+ and ran in WinBUGS and R using the R2Winbugs package. Analysis and visualization of the results were carried out in R. Three Markov chains with the initial values for the parameters set to zero was used to build the model. A total of 100,000 simulations each with the first 25,000 discarded. This was done to maintain similarity as it relates to the number of simulations run for both models. The retrospective and prospective MSE and RMSE of the model was calculated to be used for model comparison.

4. RESULTS

4.1. Descriptive Statistics

4.1.1. Observed cases Analysis

A total number of 11,486 cases of HFMD were reported for seven districts within Da Nang city over the 60 months. i.e. from January 2012 to December 2016. The average monthly cases across Da Nang city was 27.35 (Sd:19.36). Among the seven districts, Lien Chieu contributed the largest proportion of the total HFMD cases with 2054, a monthly average of 34.23(Sd: 22.60). The least contributed district was Thanh Khe, which recorded a total of 1324 cases for the 5 years, a monthly average of 22.07 (Sd: 14.33) (Table 3). The highest number of cases recorded for a month was 588 in May 2012 while the lowest recorded, 38 cases, in May 2014. Over the five-year period, 2012 recorded the highest number of cases, a total of 3302 cases, while 2014 recorded the least, 1656 cases (Figure 8). As it relates to the total cases for each month, February recorded the lowest with 456 cases over the 5 years, while August recorded the highest with 1300 cases. Hoa Vang was the district that received the highest cases for a month, 113 in August 2012, while several districts recorded as low as 3 cases for different months. It can be arguably stated that the monthly cases varied with time with most of the observed cases occurred during the first six months of every year for every district with the only exception being 2016 where most of the cases of HFMD occurred in the last six months (Figure 8).

Table 3: Descriptive statistics of the observed cases of Hand, Foot and Mouth Disease

District	Variable	Minimum	25 th	50 th	Mean (Sd)	75 th	Maximum	Total
Hai Chau	OBSERVED CASES OF HFMD	3	13	21	22.23(14.29)	27	75	1,334
Lien Chieu		6	18	27	34.23(22.60)	43	96	2,054
Ngu Hanh Son		3	11	17	22.42(16.92)	29	91	1,345
Son Tra		3	15	20	28.18(19.76)	37	83	1,691
Thanh Khe		4	12	19	22.07(14.33)	29	66	1,324
Hoa Vang		3	19	27	32.67(23.15)	42	113	1,960
Cam Le		3	17	26	29.63(19.12)	37	97	1,778
Total								

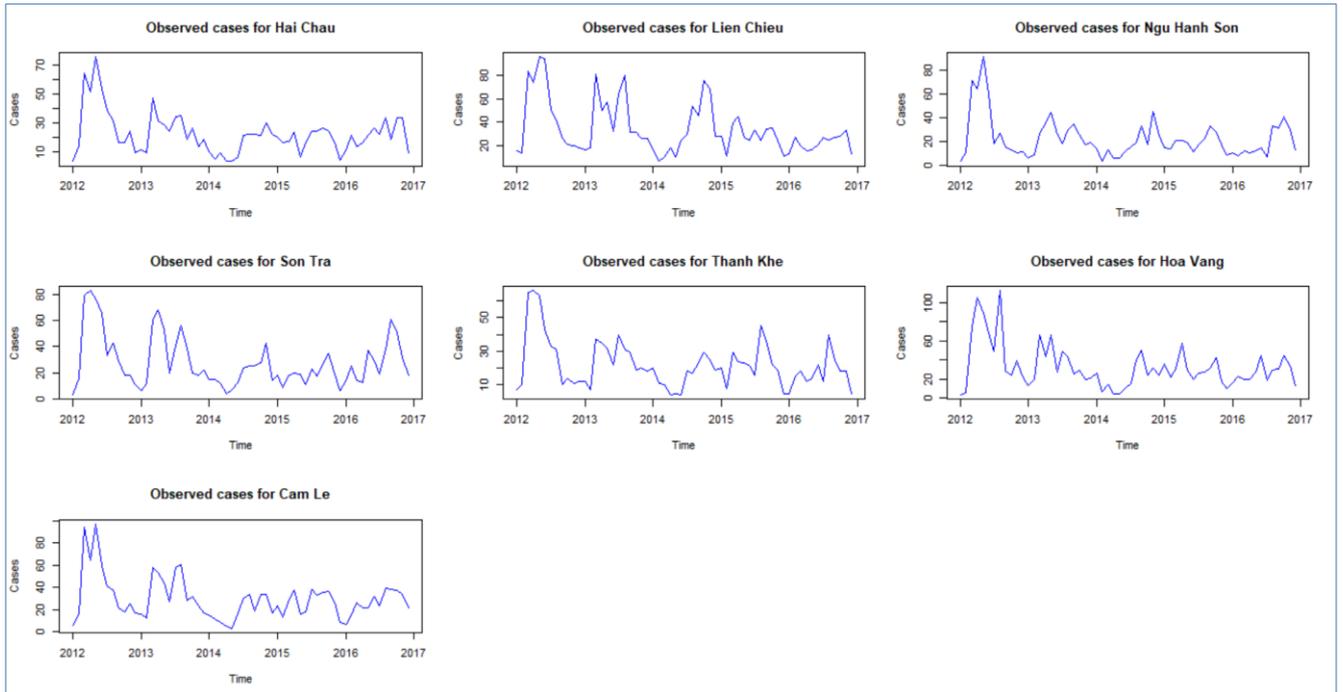


Figure 8: Time series plot of monthly HFMD observed for each district

4.1.2. Population Analysis

The smallest under five population throughout the five years was in Hai Chau district, a population of 210,371, a monthly average of 3,506 (Sd: 271), while the largest was in Lien Chieu, 1,938,655, a monthly average of 32,311 (Sd: 2,456) (Table 4).

Table 4: Descriptive statistics of the population for each district

District	Variable	Min	25 th	50 th	Mean (Sd)	75 th	Max	Total
Hai Chau	Population	3,074	3,278	3,460	3,506 (270.80)	3,724	4,065	210,371
Lien Chieu		28,118	30,215	31,886	32,311 (2,546.16)	34,305	37,520	1,938,655
Ngu Hanh Son		14,764	15,868	16,691	16,873 (1,257.53)	17,847	19,452	1,012,401
Son Tra		7,182	7,644	8,050	8,143(610.94)	8,621	9,403	488,558
Thanh Khe		4,041	4,345	4,584	4,642 (362.58)	4,929	5,386	278,508

Hoa Vang		6,854	7,354	7,760	7,821 (569.07)	8,228	8,958	469,251
Cam Le		7,365	7,992	8,462	8,555 (699.49)	9,109	9,962	513,314
Total								4,911,058

4.1.3. Seasonal Covariates Analyses

The average monthly temperatures of the study period ranged from 20.3 °C to 30.8 °C (mean:26.44; Sd: 2.99). The average monthly rainfall ranged from 0.0 to 819.4 mm (mean:182.22; Sd:224.58). The average monthly relative humidity ranged from 69% to 89% (mean:80.7; Sd:4.89) while the monthly sunny hours ranged from 28.2 to 288.3 (mean: 180.7; Sd:61.23). The last four months for each year recorded the most rainfall, an average of 420.71 mm per month, with the least amount in the first four months of each year, averaging 26.20mm. October 2014 was the month where the most rainfall was recorded while no rainfall was recorded for March 2012, see Figure 9. As it relates to relative humidity, similar results to the rainfall were observed, the last four months having the highest percentage of humidity, 84%, while the middle third of the year, or months May to August, having the lowest, an average of 75%, see Figure 10. Temperature followed a completely different pattern, with an average of 29.6 °C between May to August being the highest period throughout the 5 years, while the lowest was at the beginning of the year, i.e. between January to April, an average of 23.9 °C, see Figure 11. Sunlight was also maximum during the period of May to August every year, 236.7 hours while the lowest was the last period of the year, September to December, averaging 174.3 hours, see Figure 12.

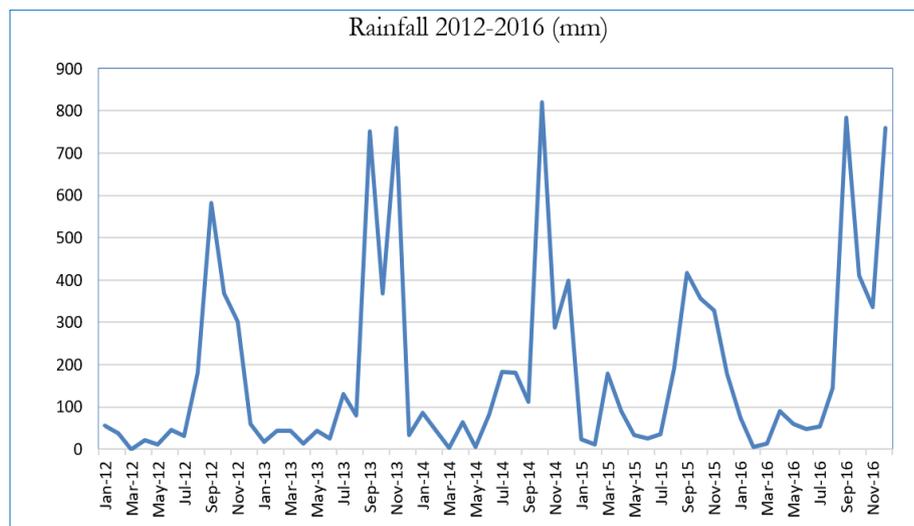


Figure 9: Rainfall pattern 2012-2016

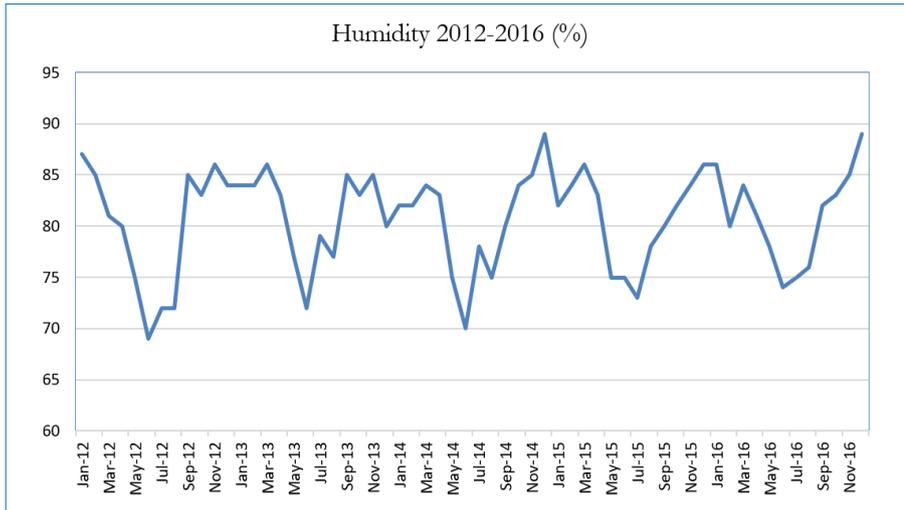


Figure 10: Humidity pattern 2012-2016

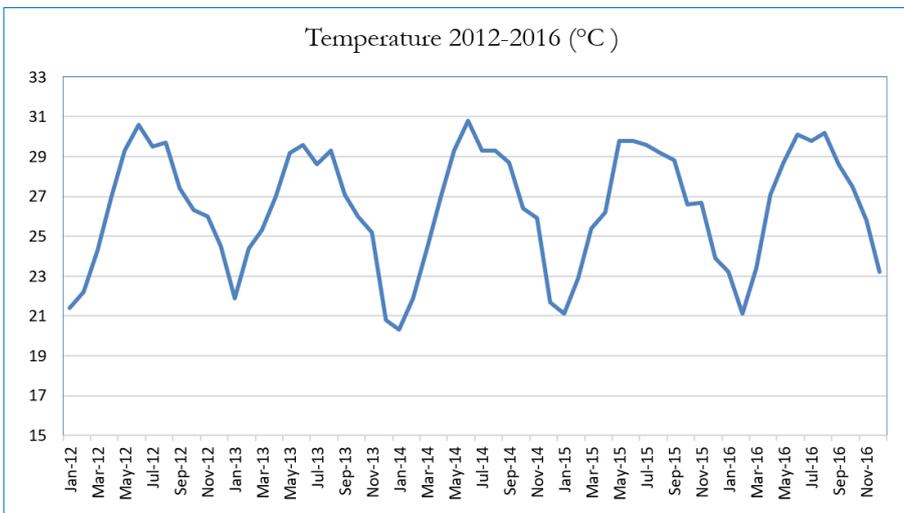


Figure 11: Temperature pattern 2012-2016

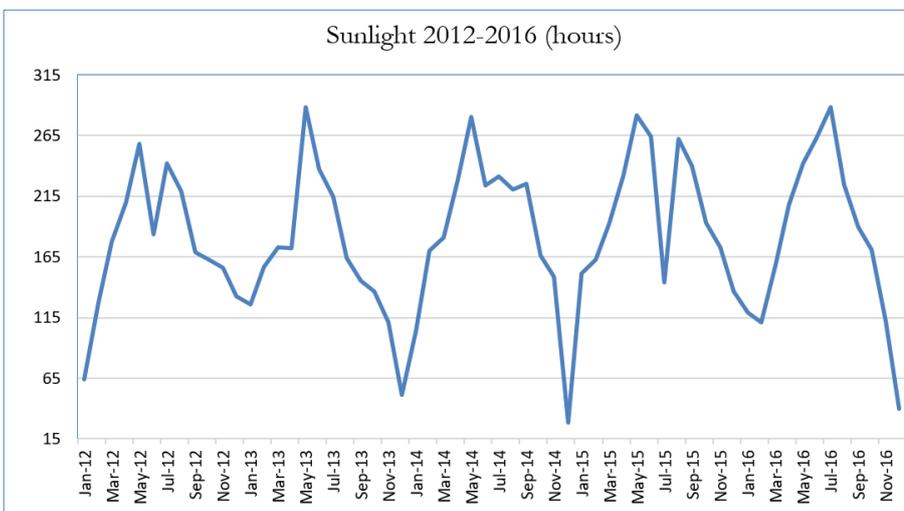


Figure 12: Sunlight pattern 2012-2016

4.1.4. Crude Relative risk Analyses

The spatiotemporal patterns of the crude HFMD relative risk for each district were presented below in Figure 20. This relative risk varied for each district and ranged from 0.2562 (25.62%) to 4.1194 (411.94%) (mean: 1.49; Sd:0.86). Hai Chau was the district over the years with the highest risk of contracting HFMD (mean: 2.697; Sd: 0.66) while, Lien Chieu having the lowest risk (mean: 0.446; Sd: 0.15) (See figures 13-18 for yearly relative risk for Da Nang city). It was observed that the crude relative risks were higher for the district with the smallest population, while the districts with the larger population having a lower risk. This clearly shows that the population and the crude relative risk have an inverse relationship, i.e. as the population increased, the crude relative risk decreased. Lien Chieu and Ngu Hanh Son were the only two districts where the risk of contracting HFMD was below 1. This means the population is negatively correlated with the relative risk for these districts while it was positively correlated for the remaining five districts, thus supporting the claim that population and crude relative risk has an inverse relationship.

A few interesting observations were: for Lien Chieu district, only once did the observed cases exceed the expected cases, in April 2014; the number of observed cases never exceeded the expected for Ngu Hanh Son throughout the five years, ranging between 0.2562 to 0.8324 and finally, throughout the five years, Hai Chau observed cases of HFMD has always exceeded the expectation cases of HFMD for population under five, with as low as 1.002 to the highest being 3.995. i.e. the chances of contracting HFMD within this district ranges between 100% to approximately 400%.

These crude relative risk maps are solely based on the observed cases and have large variances in the relative risk in sparsely populated districts. They also failed to account for the similarity in the relative risks in adjacent districts. Thus, the relative risk was treated as a random variable to introduce into an extra source of variability to capture the impact unobserved confounding factors has on it. Using the Bayesian approach, appropriate priors were assigned to the likelihood data to smooth the risks within the Da Nang city.

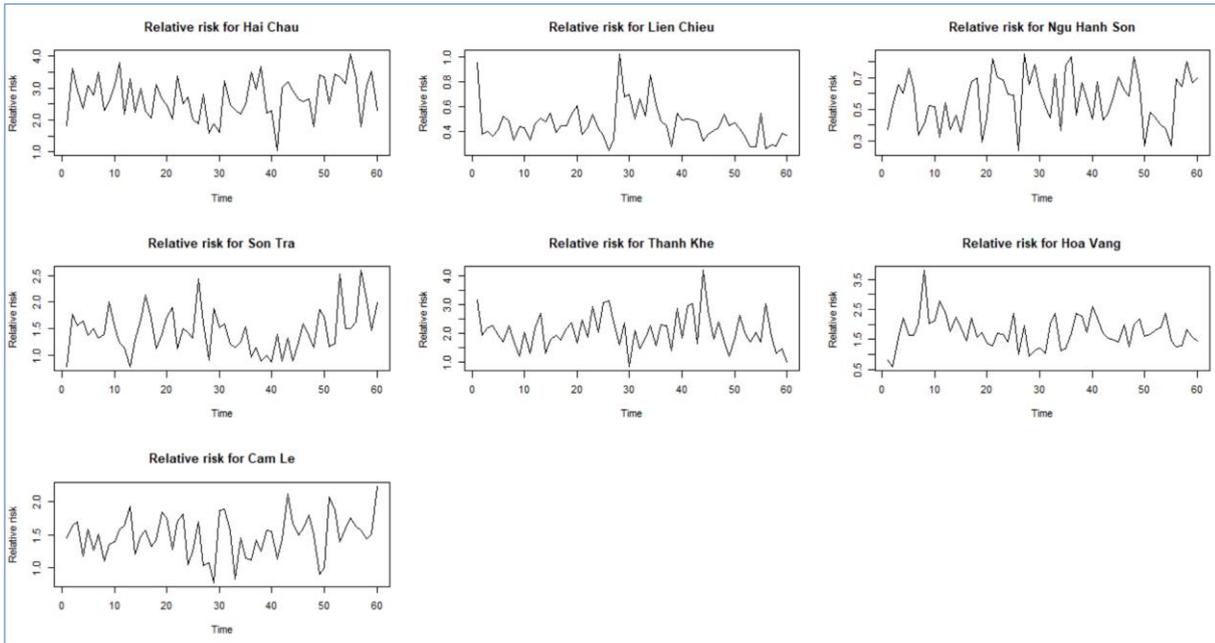


Figure 13: Spatiotemporal pattern of HFMD crude relative risk for each district, 2012-2016

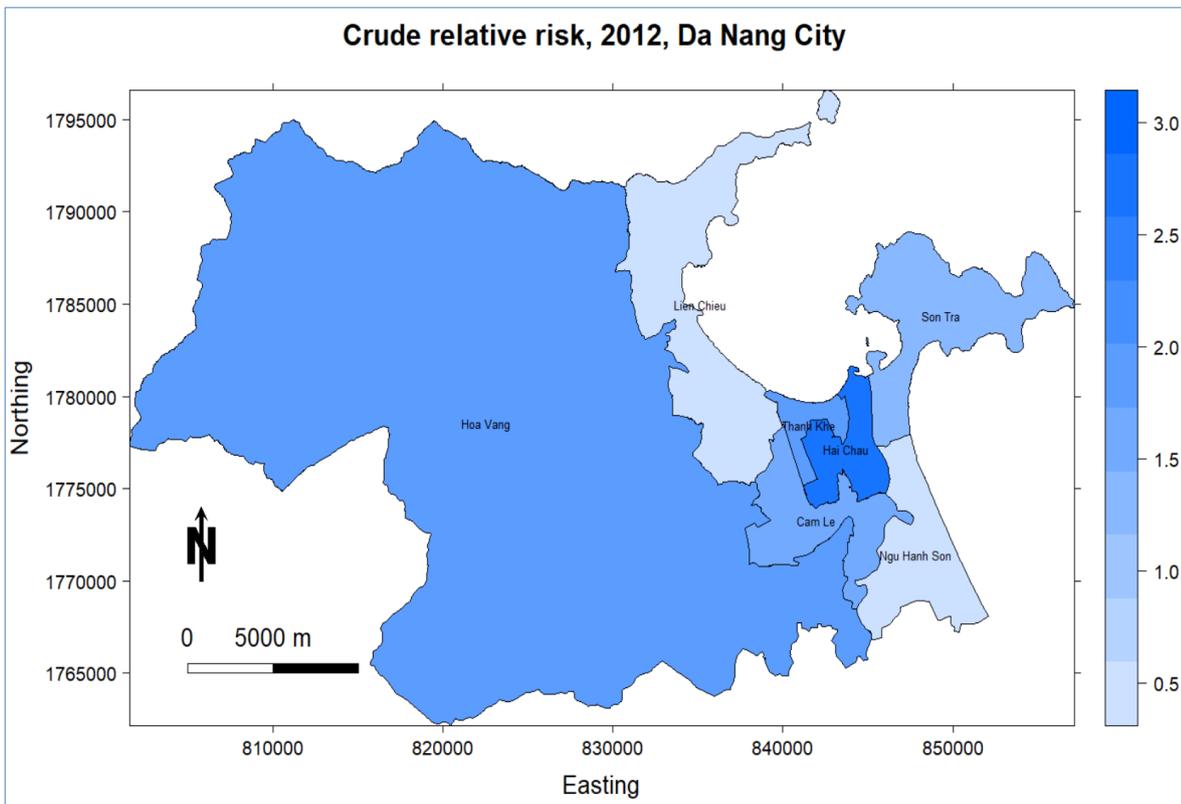


Figure 14: Crude relative risk for Da Nang city, Vietnam, 2012

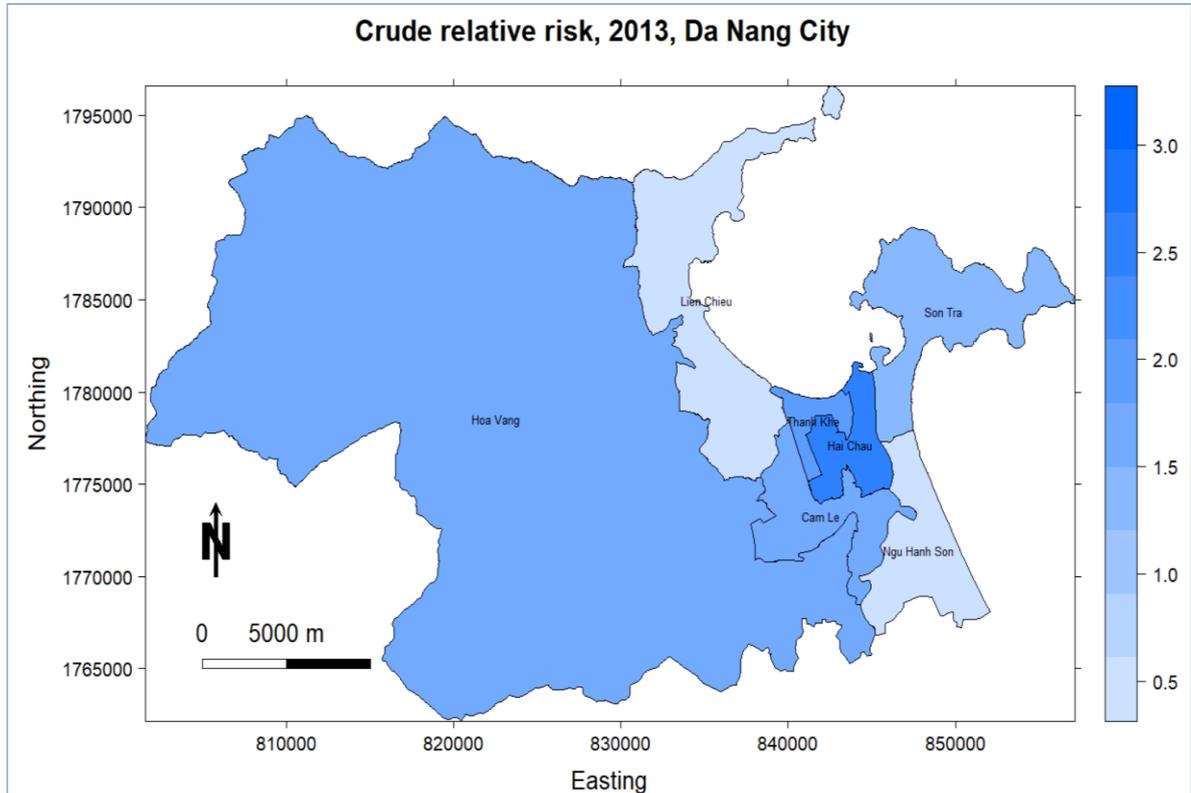


Figure 15: Crude relative risk for Da Nang city, Vietnam, 2013

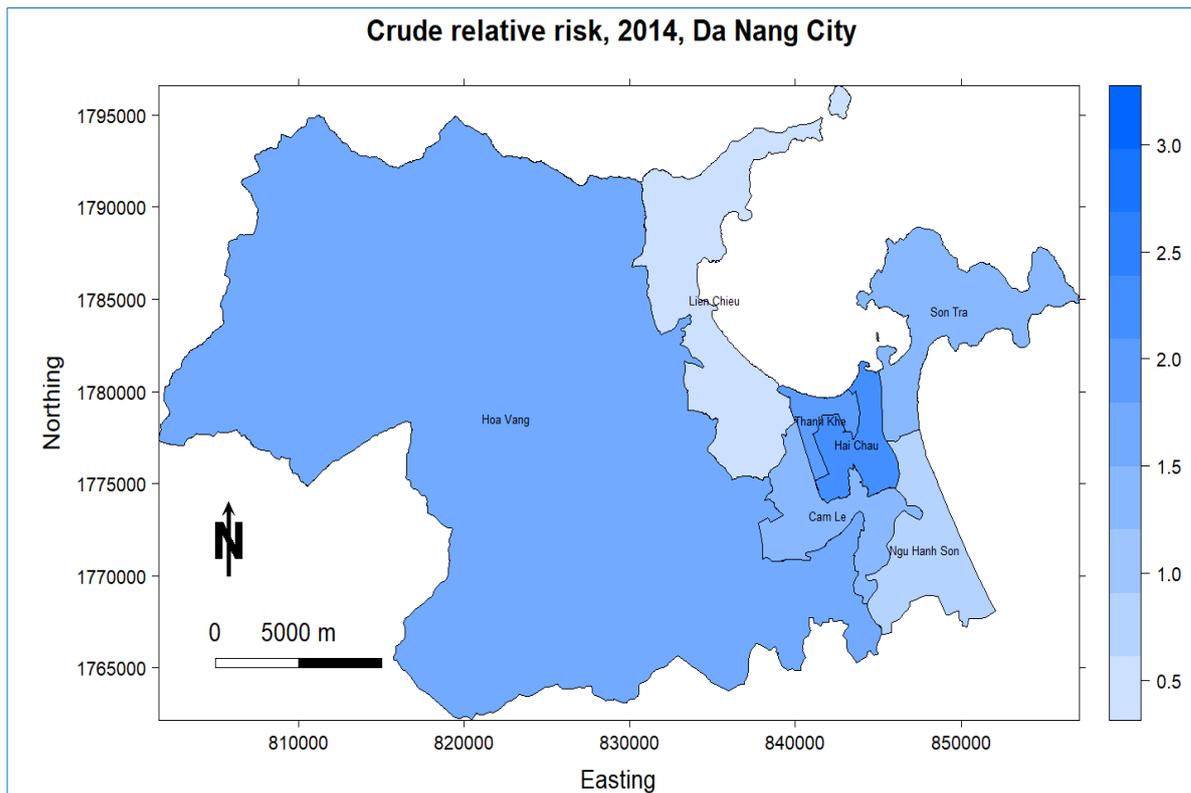


Figure 16: Crude relative risk for Da Nang city, Vietnam, 2014

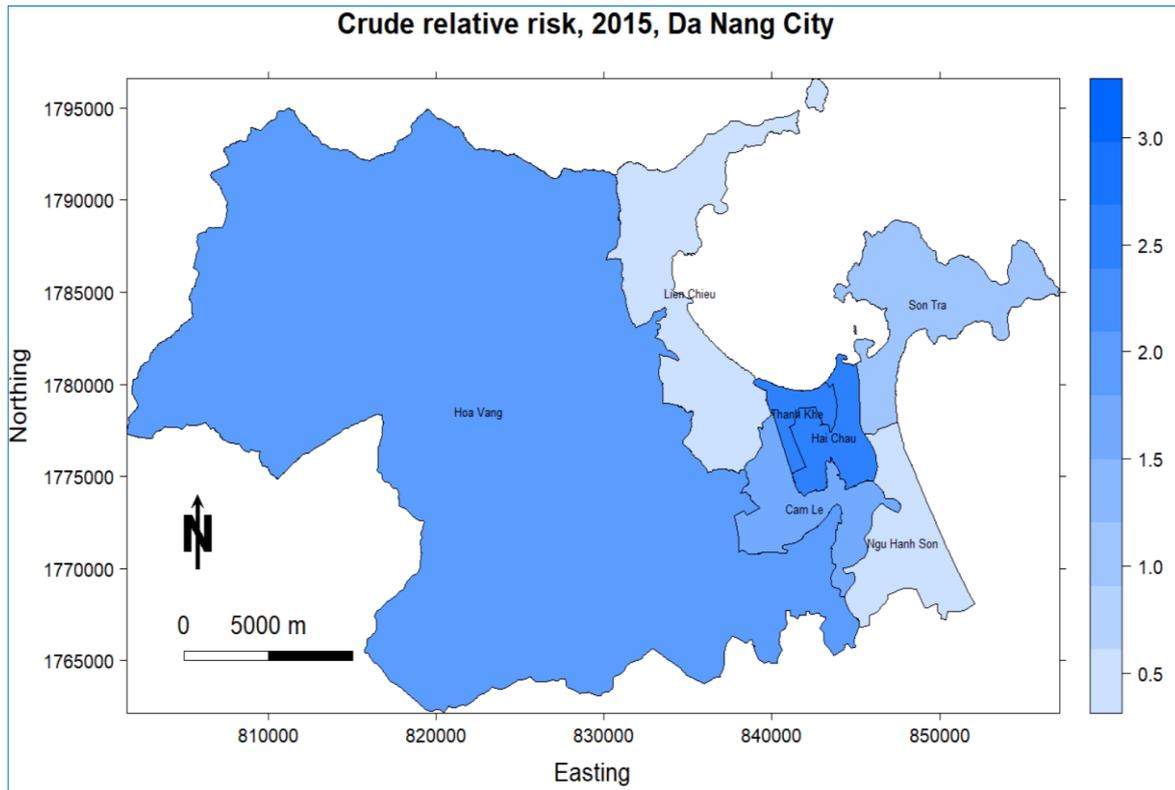


Figure 17: Crude relative risk for Da Nang city, Vietnam, 2015

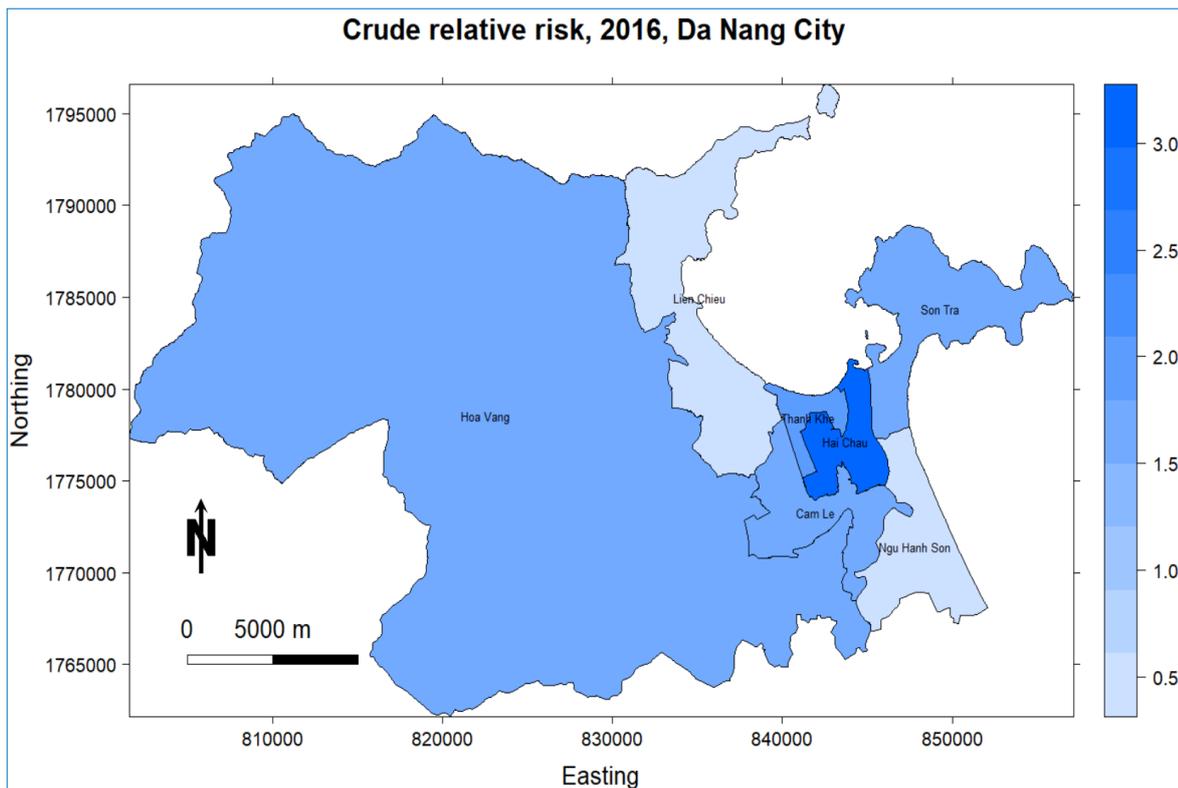


Figure 18: Crude relative risk for Da Nang city, Vietnam, 2016

4.1.5. Seasonal factors affecting relative risk

As stated in Section 3.2, to identify the effects of the covariates (rainfall, relative humidity, sunlight, temperature) on relative risk, a linear model of covariates against observed cases at both the district level and city level was carried out to show the covariate(s) that is/are statistically significant. Also, the correlation between each covariate and the correlation between the crude relative risk with the covariates at both district level and city level was assessed for the collinearity problem. To estimate the linear long-term trend, the month was used as a dummy variable with values of 1 to 60 representing the months of 5 years.

At the city level, the temperature was the only seasonal variable found to be statistically significant with p-values of 5.27×10^{-6} when all the seasonal variables were considered at lag 0 (See table 5). The Pearson correlation coefficient that takes a value between -1 and 1 is used to determine the strength of a linear association. The correlation at lag zero showed that sunlight and relative humidity were significantly correlated with temperature (0.79 and -0.73 respectively). Similarly, the correlation of the observed cases of HFMD at both the city level and district level with the temperature has a higher value as compared to rainfall (see Table 5). Thus, supporting the result of the linear model that temperature was statistically significant and the only covariate considered to build the GLMM and MBRF models.

Table 5: Pearson correlations between HFMD cases and meteorological variables along with p-values (temperature)

District		Rainfall	Temperature	P-value
Hai Chau	HFMD cases	-0.12	0.33	0.0459
Lien Chieu	HFMD cases	-0.03	0.34	0.02141
Ngu Hanh Son	HFMD cases	0.02	0.30	0.0455
Son Tra	HFMD cases	0.03	0.34	0.0324
Thanh Khe	HFMD cases	-0.11	0.32	0.0371
Hoa Vang	HFMD cases	-0.19	0.34	0.0198
Cam Le	HFMD cases	-0.10	0.37	0.0304

4.2. Estimated Spatiotemporal SMR/ Relative risk

As stated in the Methodology section, the log of the relative risk of HFMD can be decomposed into the additive of the three main components of the time series, i.e. season, trend and residuals. The relative risk itself can be viewed as the exponential of the additive of each component. These smooth relative risk maps not only maintained the original spatial risk pattern but also captured the local variation from neighbouring areas. In figures 19 and 20, the spatiotemporal patterns of the smooth HFMD relative risk for each district of both models for the period 2012-2016 were presented. This relative risk varied for each district and

ranged from 0.292 (29.2%) to 4.034 (403.4%) for both models. See Table 7. The predicted mean relative risk was highest for Hai Chau (GLMM: Mean: 2.743; Sd: 0.17, MBRF: Mean: 2.664; Sd: 0.44) and Lien Chieu was the lowest (GLMM: Mean: 0.458; Sd: 0.07, MBRF: Mean: 0.454; Sd: 0.09), for both models.

Hoa Vang for GLMM and Hai Chau for MBRF were the districts that recorded the highest risk for one month throughout the five years while the lowest was Lien Chieu for both models. As it relates to the months where the highest relative risk was predicted, similar months of high prediction were observed for 5 of the 7 districts while both models predict similar months for the lowest in one district, see Table 8 below.

The predicted spatiotemporal smooth relative risk of the MBRF follows a similar pattern of the crude relative risk, Figure 13 and 20. Hence, MBRF does not significantly smooth the relative risk. However, the GLMM predicted spatiotemporal smooth relative risk follows a different pattern for some months. This was also shown in the Mean Square Error (MSE) and Root Mean Square Error (RMSE) for both models, see Table 7.

It was also observed visually that there was no clear indication of the seasonality of the relative risk for each district as throughout the five years, different peaks occurred for different months. However, when modelled using only the seasonal variable, a clear seasonal pattern was observed. Similarly, it was observed that there were different trends associated with the relative risk for each district. Two examples are Cam Le and Hoa Vang of figures 19 and 20, there was a clear increasing trend for Cam Le while the trend decreased over time for Hoa Vang. This was also shown when the relative risk was modelled solely on the dummy variable.

Table 6: Retrospective MSE and RMSE for both models

Model	Retrospective MSE	Retrospective RMSE
GLMM	0.7869	0.8871
MBRF	0.1907	0.4372

Table 7: Maximum and Minimum Relative risk of HFMD along with the month and year observed for both models

District	Maximum Relative Risk, GLMM (MBRF)	Month and Year, GLMM (MBRF)	Minimum Relative Risk, GLMM (MBRF)	Month and year, GLMM (MBRF)
Hai Chau	3.141 (3.736)	November 2012 (February 2012)	2.316 (1.482)	October 2014 (May 2015)
Lien Chieu	0.839 (0.725)	October 2014 (January 2012)	0.347 (0.292)	October 2016 (October 2016)
Ngu Hanh Son	0.694 (0.761)	October 2016 (October 2016)	0.438 (0.351)	October 2014 (July 2013)

Son Tra	2.441 (2.249)	September 2016 (September 2016)	1.076 (0.845)	August 2015 (January 2013)
Thanh Khe	3.657 (3.586)	August 2015 (August 2015)	1.669 (1.201)	October 2016 (June 2014)
Hoa Vang	4.034 (2.663)	August 2012 (August 2012)	1.347 (0.712)	October 2014 (February 2012)
Cam Le	1.655 (1.899)	December 2016 (December 2016)	1.194 (0.997)	August 2012 (May 2014)

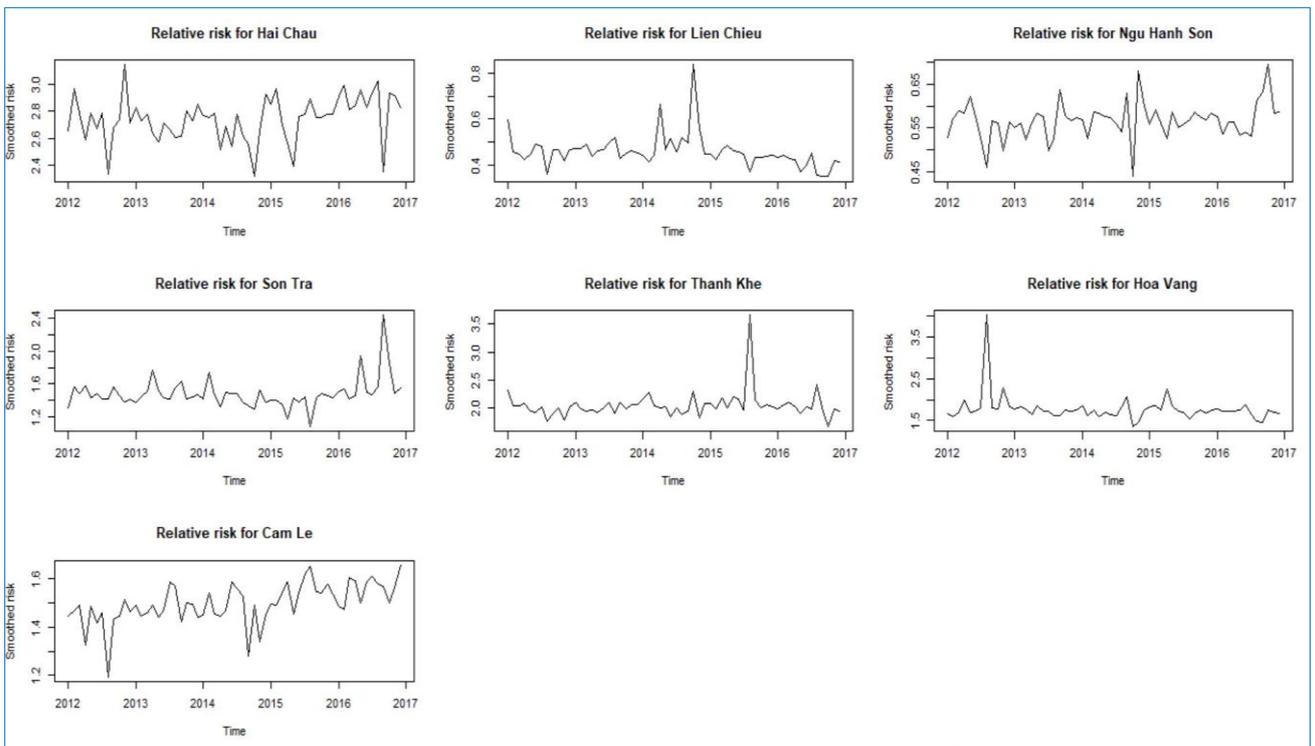


Figure 19: GLMM smooth relative risk, Da Nang City, Vietnam 2012-2016

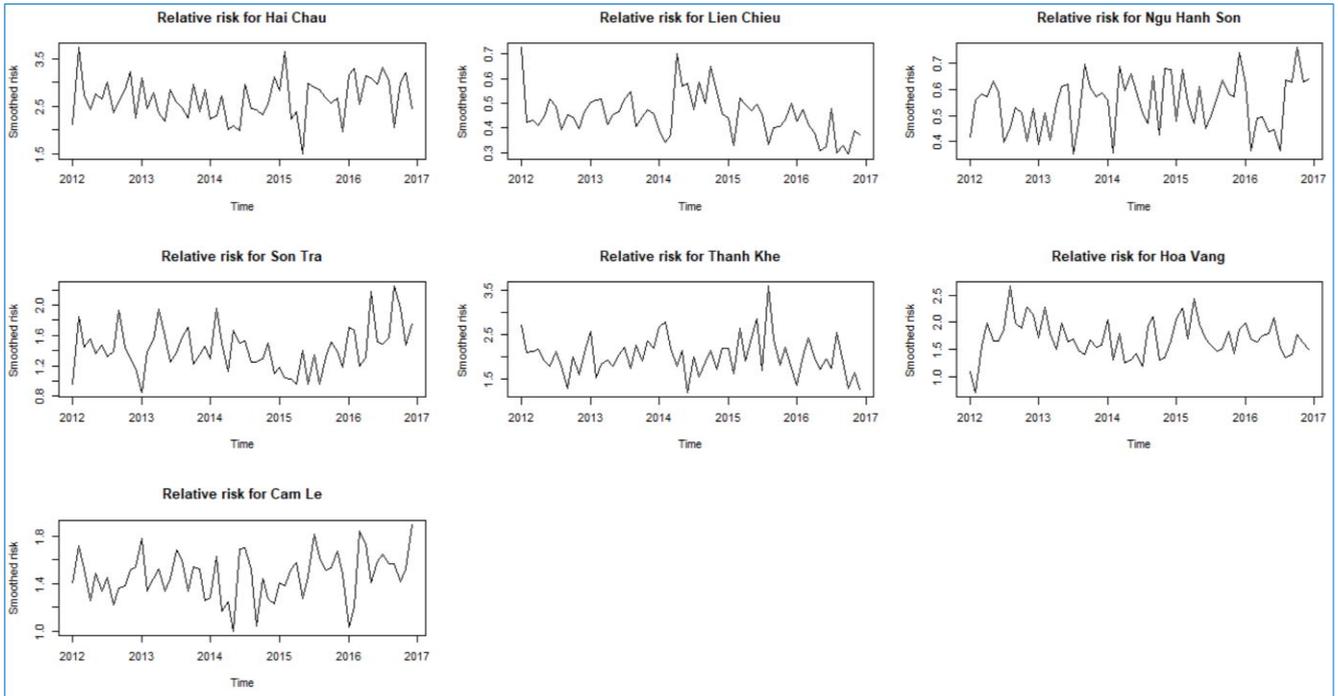


Figure 20: MBRF smooth relative risk, Da Nang City, Vietnam 2012-2016

In figures 21-23, the estimated SMR as explained by the relative risk of the GLMM and MBRF models for January 2013 are compared with the crude relative risk. Cam Le, highlighted with the square below, shows smoother rates as a result of the meteorological factor and spatial dependences from neighbouring districts. The rates for the remaining districts fall within the same range except for Son Tra, which showed a higher rate in the GLMM model, highlighted by the black triangle in figure 23.

Also shown in figures 24 and 25, was the relative risk of three different months for both models. Figure 24.ii showed that Hoa Vang had the highest risk in August 2012. Hai Chau was the district with the highest risk for February 2012 and November 2012 as it relates to the GLMM model, Figures 24.i and iii. Similar results were also shown for MBRF in Figure 25.a, b, c. These smooth SMR maps could give more intuitive information and can be used for disease prevention and control.

These time series predicted smooth relative risk was decomposed into temporal trend, seasonality and spatial-temporal residuals, with the following subsections giving an analysis of these components.

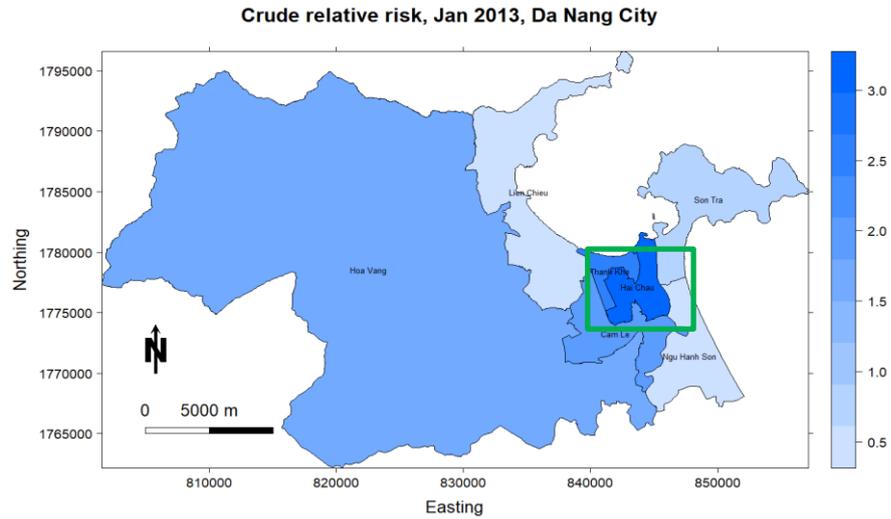


Figure 21: Crude relative risk, January 2013

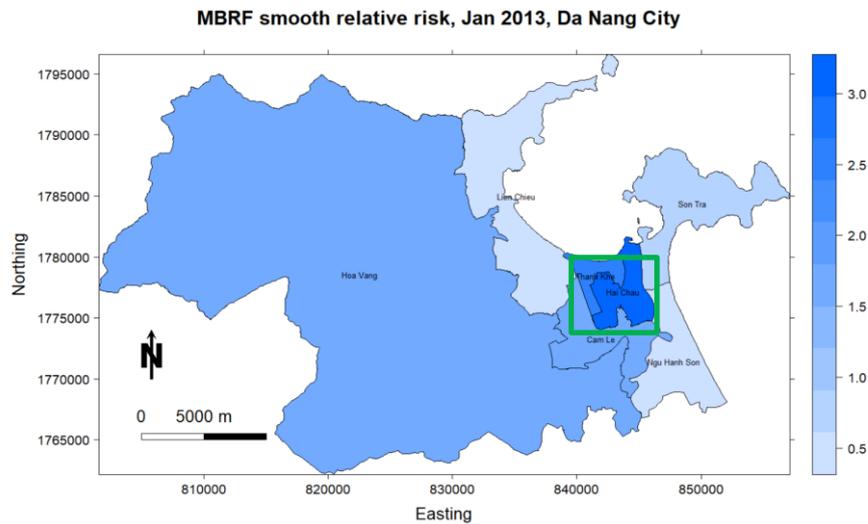


Figure 22: MBRF smooth relative risk, January 2013

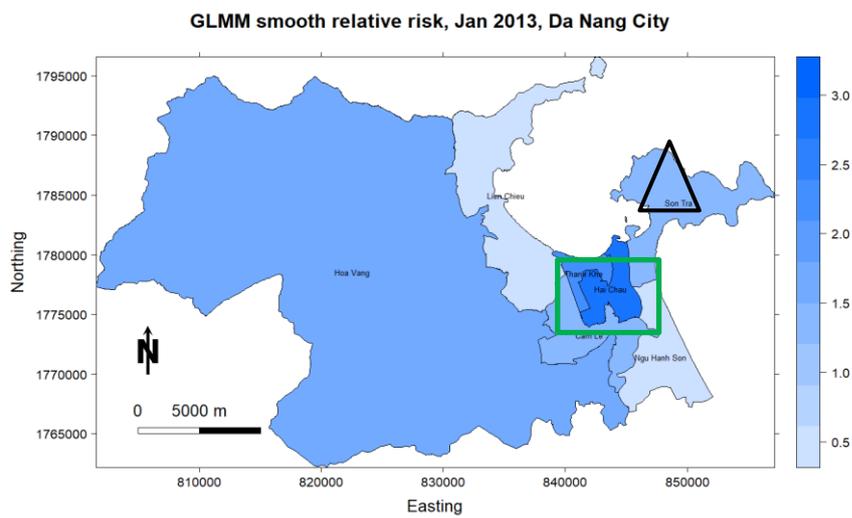
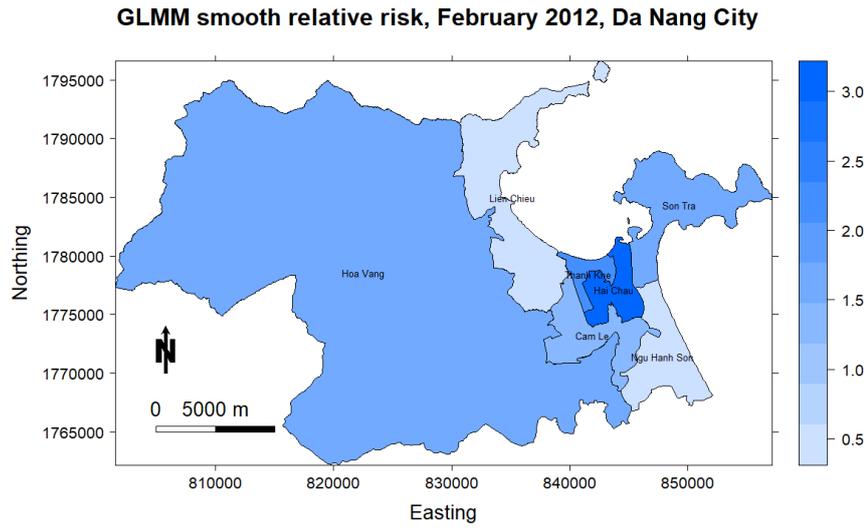
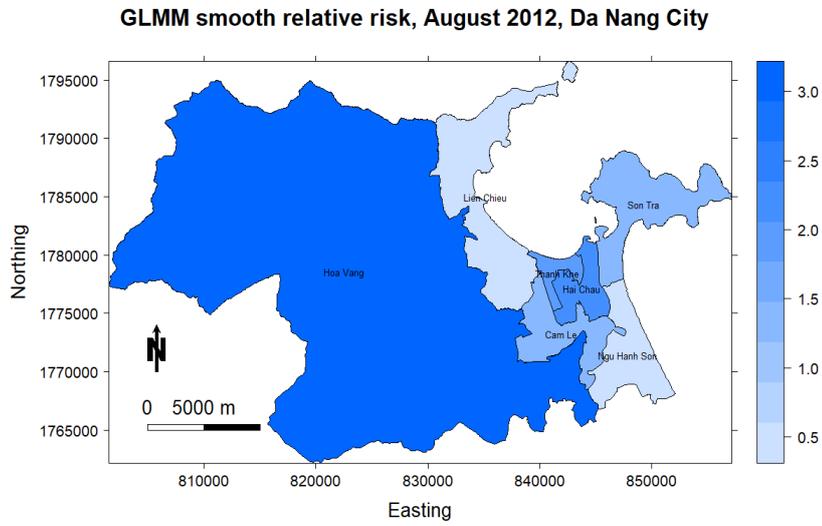


Figure 23: GLMM smooth relative risk, January 2013

i.



ii.



iii.

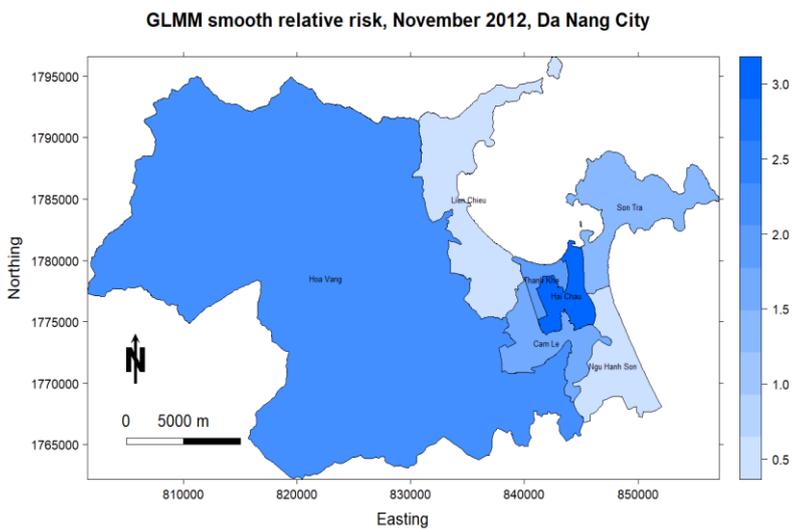
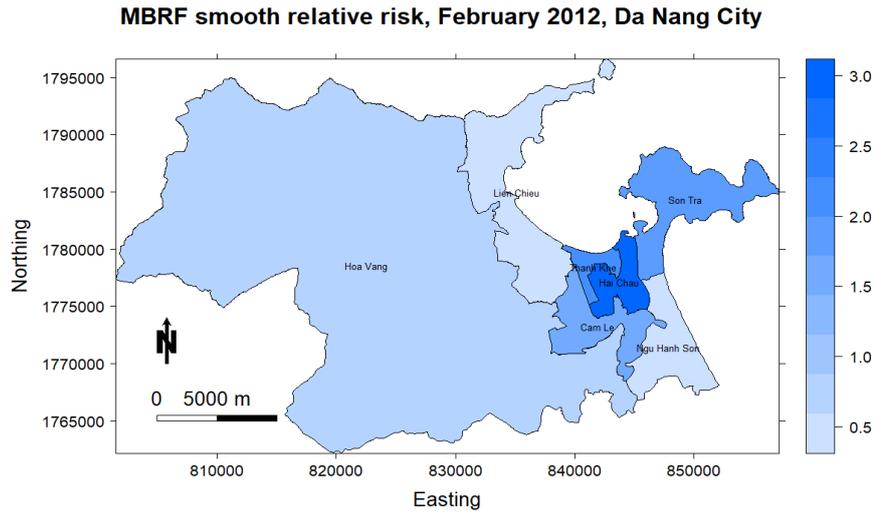
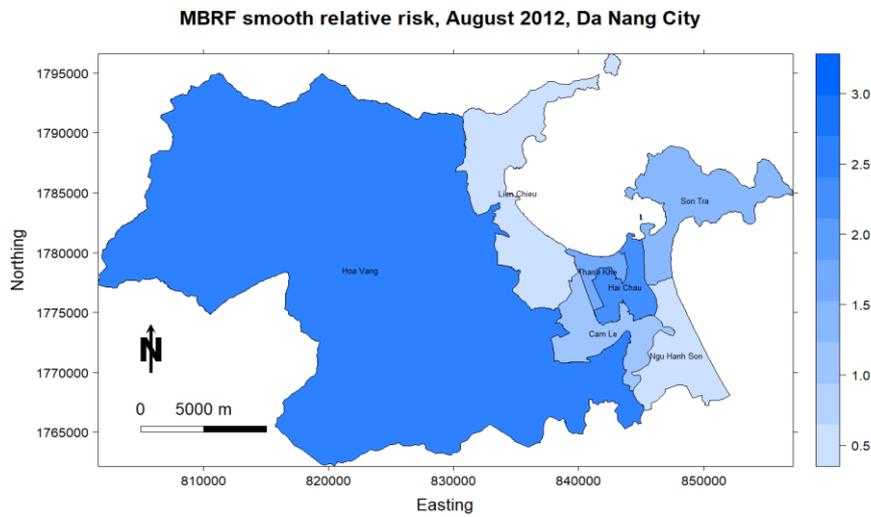


Figure 24: GLMM smooth relative risk for three months in 2012

a.



b.



c.

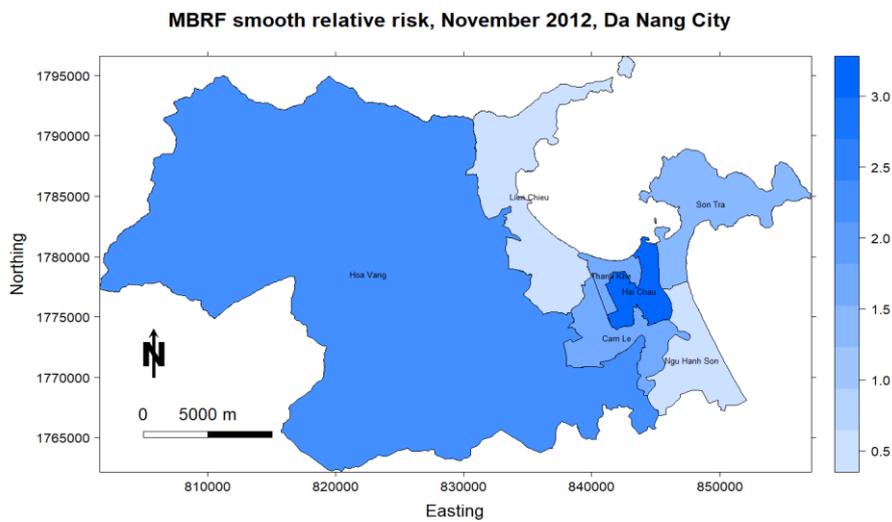


Figure 25: MBRF smooth relative risk for three months in 2012

4.2.1. Temporal Trend

As mentioned in section 3.3, a dummy variable representing each month was used to model the long-term trend of the relative risk. This was calculated by taking the exponent of the product of the dummy variable and its corresponding regression coefficient for each district. This dummy variable was assigned a Normal Distribution prior with the variances assigned an inverse gamma distribution prior within the Bayesian Framework. In figure 26, plots of the long-term trend were shown. The long-term trend ranged from 0.882 to 1.112 (GLMM) and 0.776 to 1.108 (MBRF) and varied linearly over time. Similar trends were observed in both models except for one district, i.e. Thanh Khe. This district showed opposite results from the two models. The GLMM model showed a positive linear trend while the MBRF showed this trend to be negatively linear.

A positive trend was observed in Hai Chau, Ngu Hanh Son, Son Tra, Thanh Khe and Cam Le for the GLMM while Thanh Khe had a negative linear trend for the MBRF. The GLMM model showed two districts, i.e. Ngu Hanh Son and Son Tra, to be linear only for the first two years, 2012 and 2013. At the start of 2014, this trend was shown to be non-linear as there was an increasing gradual curve. Lien Chieu and Hoa Vang were all shown to have a negative linear trend by both models. These variations in the long-term trend can be explained by the size and the location of the districts. The two largest districts in terms of area and with their locations having fewer influences from neighbors showed a negative linear trend. Those districts that showed to have a positive linear trend were centrally located in the city except for Son Tra. The influence from each other and the size of their neighboring districts can be inferred as a major contributing factor as to why similar trends were observed.

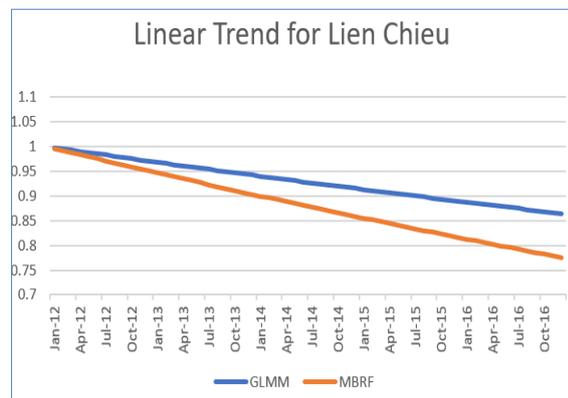
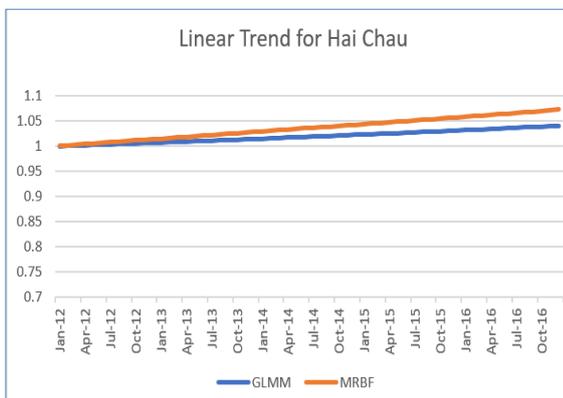




Figure 26: Linear trend of relative risk for districts in Da Nang City 2012-2016

4.2.2. Seasonality

Table 9 shows the overall relative risk (Intercept) for each district and its corresponding Credible Interval. This is the risk that is present when the risk factors are zero. Also, the values of the seasonal coefficient (Temperature). The exponent of the regression coefficients of temperature arising from both models was shown to be within each other. This was also showed at the 95% credible interval, i.e. the range of the values are identical.

Table 8: Estimated posterior parameters (GLMM), regression coefficients (MBRF), along with the exponent of these coefficients and its corresponding 95% credible interval

District	Variable	Coefficient GLMM (MBRF)	Relative risk of HFMD (exponential of coefficients)	95% Credible Interval
Hai Chau	Intercept	1.1446 (1.1556)	3.141 (3.176)	2.825, 3.457 (2.830 ,3.522)
	Temperature	-0.0058 (-0.0087)	0.994 (0.991)	0.740 ,1.260 (0.740, 1.260)
Lien Chieu	Intercept	-0.7775 (-0.7405)	0.459 (0.477)	0.160, 0.758 (0.124, 0.830)
	Temperature	0.0015 (0.0025)	1.002 (1.002)	0.740 ,1.260 (0.740, 1.260)
Ngu Hanh Son	Intercept	-0.5692 (-0.7580)	0.566 (0.469)	0.266, 0.866 (0.103, 0.835)
	Temperature	-0.0022 (0.0032)	1.000 (1.003)	0.740 ,1.260 (0.740, 1.260)
Son Tra	Intercept	0.3389 (0.0146)	1.403 (1.015)	1.104, 1.702 (0.661, 1.369)
	Temperature	0.0013 (0.0106)	1.001 (1.011)	0.740 ,1.260 (0.740, 1.260)
Thanh Khe	Intercept	0.7723 (1.0439)	2.165 (2.840)	1.862, 2.468 (2.479, 3.201)
	Temperature	-0.0032 (-0.0122)	0.997 (0.988)	0.740 ,1.260 (0.740, 1.260)
Hoa Vang	Intercept	0.6347 (0.4267)	1.886 (1.532)	1.586, 2.186 (1.165, 1.899)
	Temperature	-0.0024 (0.0035)	0.998 (1.003)	0.740 ,1.260 (0.740, 1.260)
Cam Le	Intercept	0.3015 (0.2191)	1.352 (1.245)	1.054, 1.650 (0.917, 1.573)
	Temperature	0.0018 (0.0041)	1.002 (1.004)	0.740 ,1.260 (0.740, 1.260)

The relative risk of HFMD shows two different seasonal patterns among the seven districts for both models. Pattern one showed two peaks every year, one in June (highest peak) and a small one in August, i.e. during the summer time or in the case of Da Nang, at the end of the dry season. It also showed two troughs, July and January, with the deepest trough in January. The other pattern showed the reverse with peaks in January (highest peak) and a small peak in July, the troughs in June and in August with the deepest trough occurring in June every year. The GLMM and MBRF seasonal components showed a similar pattern for 5 of the 7 districts. The two districts, Ngu Hanh Son and Hoa Vang were the only two districts where opposite results were obtained from the two models.

Lien Chieu, Son Tra and Cam Le for both models followed pattern one as described above (Figures 28, 30 and 33). As it relates to the seasonality of temperature, two peaks occurred in June and August with the deepest trough being in January. Based on these observations, the seasonality of Relative Risk for Lien Chieu, Son Tra, and Cam Le are all positively correlated with the seasonality of temperature as the peak and trough of both occurred in the same months. Hence, the seasonality of relative risk increases and are said to be in phase.

Hai Chau and Thanh Khe districts (Figures 27 and 31) for both models followed pattern two with the highest peaks of contracting HFMD in January for each year. This shows that there was a negative correlation, as when the temperature was at its maximum, the seasonality was at its minimum. Hence, the seasonality of relative risk decreases and are said to be out of phase. However, it can be inferred that the maximum temperature resulted in an increased in the relative risk in the months to follow. The final two districts, Ngu Hanh Son and Hoa Vang (Figures 29 and 32) showed opposite results for both models. The GLMM results showed that these districts are negatively correlated with temperature while the MBRF showed that they are positively correlated with temperature.

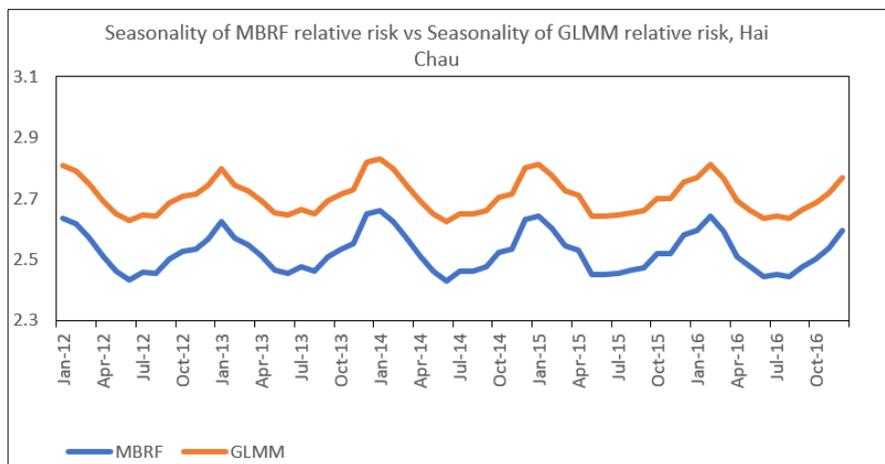


Figure 27: Seasonality of MBRF RR vs Seasonality of GLMM RR for Hai Chau 2012-2016

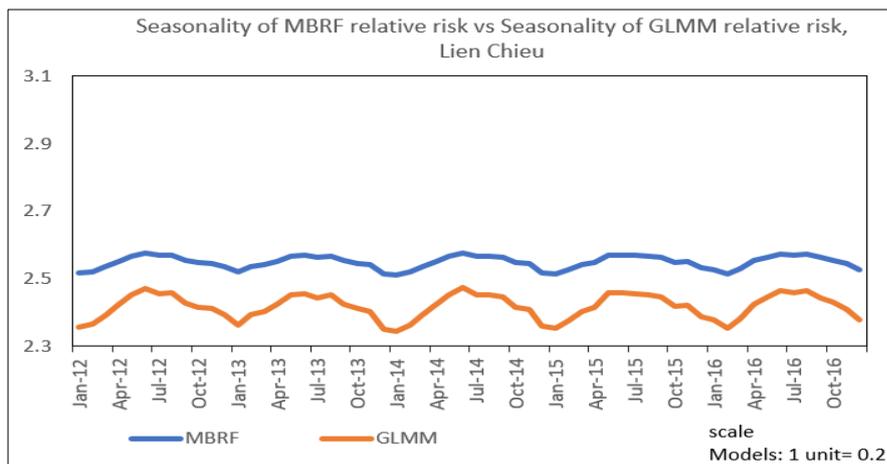


Figure 28: Seasonality of MBRF RR vs Seasonality of GLMM RR for Lien Chieu 2012-2016

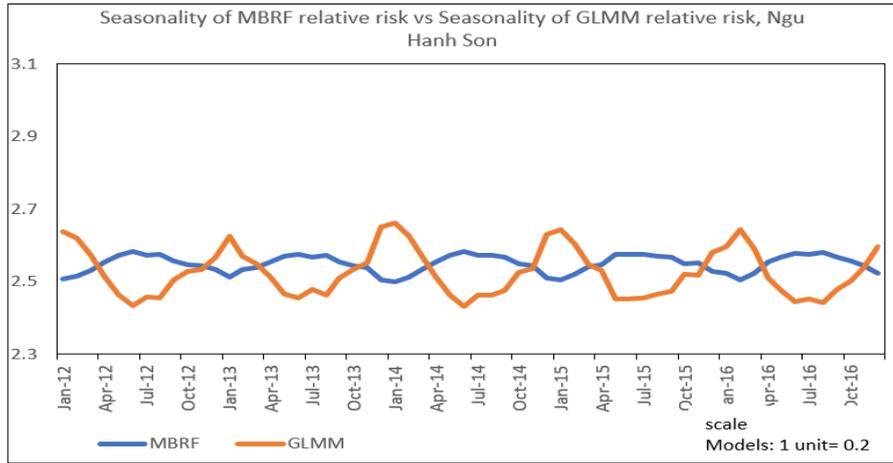


Figure 29: Seasonality of MBRF RR vs Seasonality of GLMM RR for Ngu Hanh Son 2012-2016

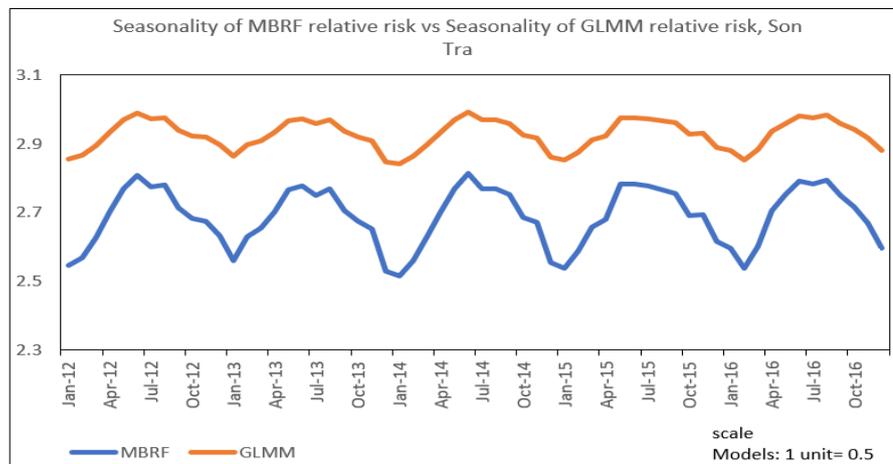


Figure 30: Seasonality of MBRF RR vs Seasonality of GLMM RR for Son Tra 2012-2016

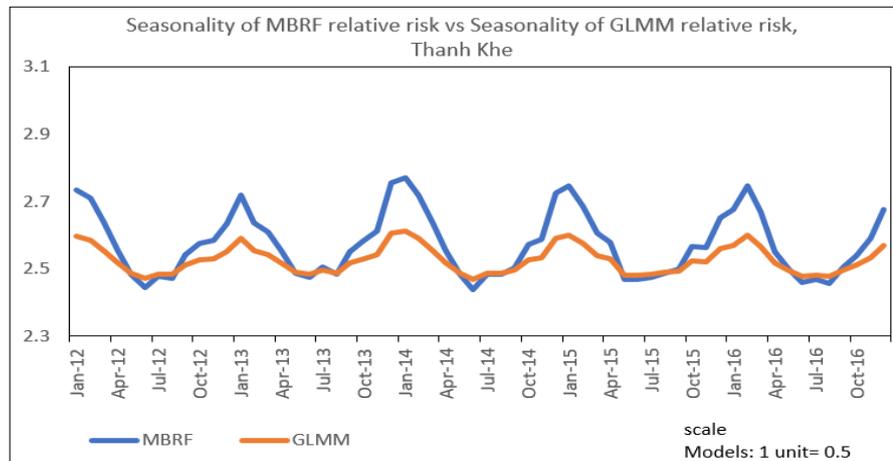


Figure 31: Seasonality of MBRF RR vs Seasonality of GLMM RR for Thanh Khe 2012-2016

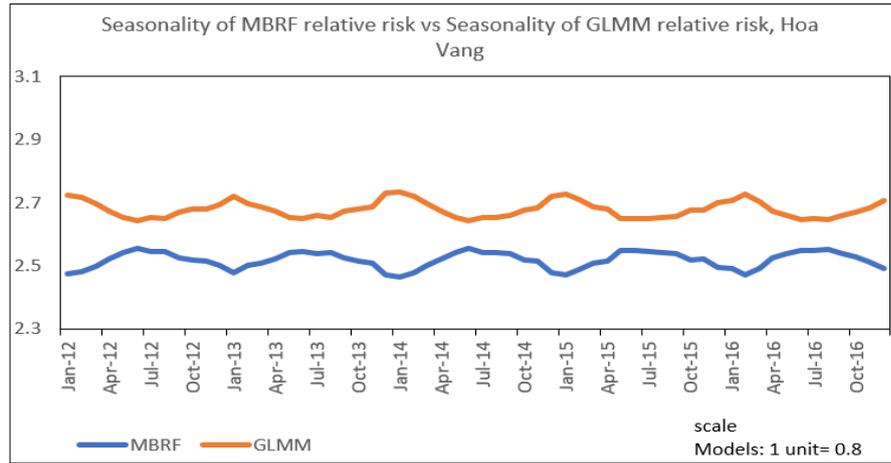


Figure 32: Seasonality of MBRF RR vs Seasonality of GLMM RR for Hoa Vang 2012-2016

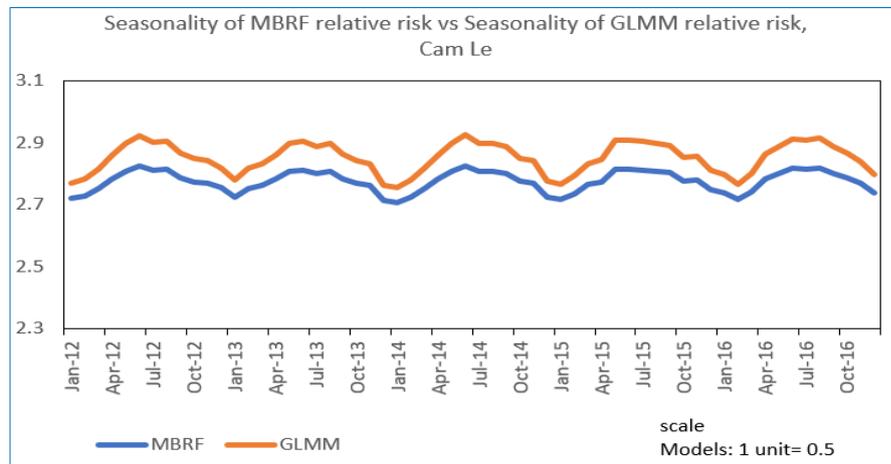


Figure 33: Seasonality of MBRF RR vs Seasonality of GLMM RR for Cam Le 2012-2016

4.2.3. Spatial temporal residuals

The final component of the models was the residuals. These residuals represent variation not yet explained by the temporal trend and the seasonal risk factor temperature.

In figures 34 and 35, the combined spatial and heterogeneity effects nested in time for each district from both models are plotted. Table 11 shows the proportion of these residuals that are being explained by the spatial structure for each time period from both models.

From the residual’s analyses, a few major observations were found:

- The residual risks of both models have smaller values than the relative risk. However, its spatiotemporal pattern follows the same pattern of the smooth relative risks of each predicted

model. This is because the trend and the seasonality due to temperature do not have a spatial structure.

- In some months, the predicted relative risk was higher than the observed relative risk for the MBRF as it resulted in negative residuals while for the GLMM all the predicted relative risk was lower than the observed.
- The districts with the highest relative risk also had the highest residuals arising from the models.
- For the GLMM model, for every month, most of the residuals were spatially structured, i.e. residuals arising due on their neighboring values with the lowest percentage being 53.26 for June 2014, while for the MBRF, in some months, most of the residuals change between spatial or random with the lowest percentage being 20.44 for April 2014 of the spatial effect.
- It was also observed that in 2015 and 2016, most of the residuals were explained spatially for the MBRF, i.e. they arise due to interaction with its neighbors. This was also seen as predictions were closer to the observed relative risk in those years as the Non-spatial or random errors decreased.

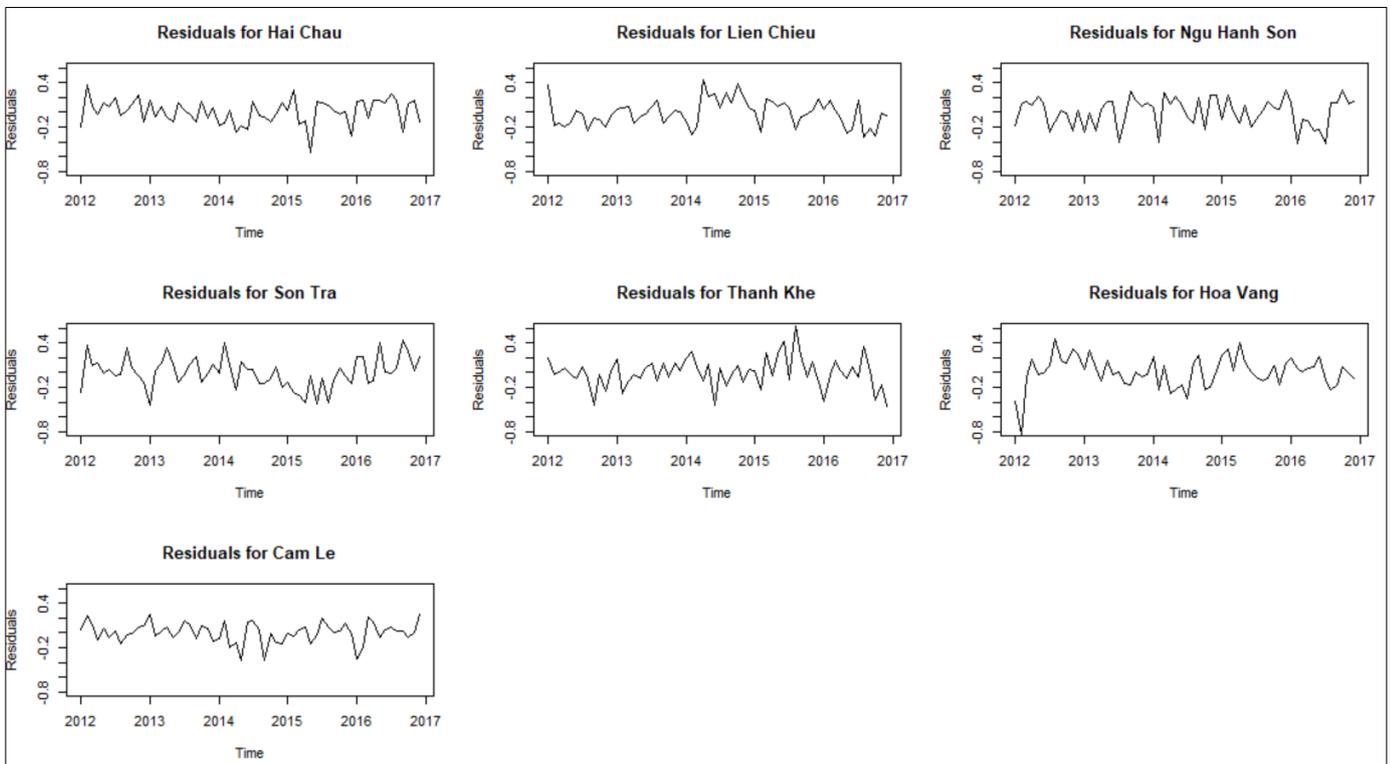


Figure 34: Space-Time residuals for each district (MBRF)

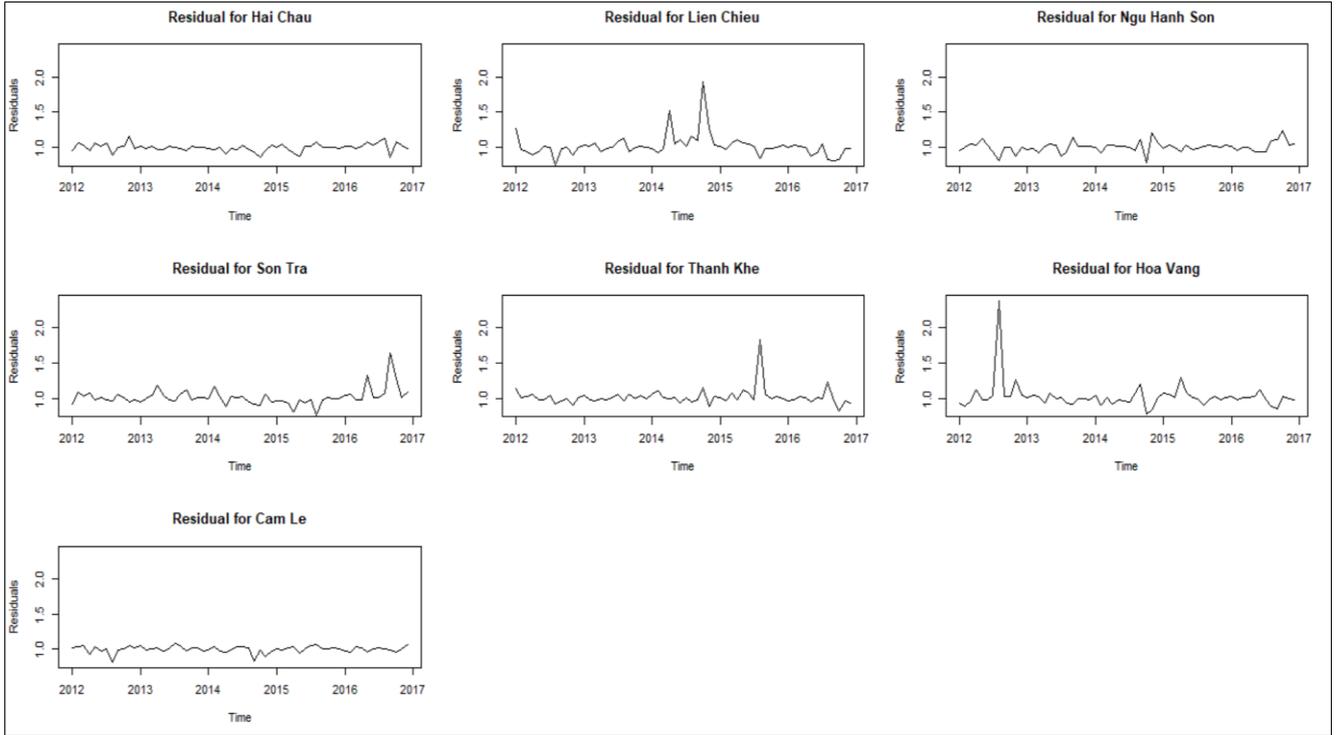


Figure 35: Space-Time residuals for each district (GLMM)

Table 9: Percentage of the residuals explained spatially (interaction with its neighbours) for each month for GLMM and MBRF

Month	Percentage of the residuals explained spatially GLMM (MBRF)	Month	Percentage of the residuals explained spatially GLMM (MBRF)	Month	Percentage of the residuals explained spatially GLMM (MBRF)	Month	Percentage of the residuals explained spatially GLMM (MBRF)	Month	Percentage of the residuals explained spatially GLMM (MBRF)
Jan-12	76.28 % (33.01%)	Jan-13	70.59% (95.66%)	Jan-14	59.40% (66.08%)	Jan-15	60.65% (75.05%)	Jan-16	65.26% (91.86%)
Feb-12	66.83% (98.42%)	Feb-13	63.24% (91.46%)	Feb-14	62.69% (31.08%)	Feb-15	62.05% (92.15%)	Feb-16	62.13% (78.92%)
Mar-12	67.66% (54.12%)	Mar-13	58.87% (68.96%)	Mar-14	59.36% (48.85%)	Mar-15	71.99% (94.16%)	Mar-16	63.82% (80.63%)
Apr-12	61.50% (58.37%)	Apr-13	77.46% (86.26%)	Apr-14	71.98% (20.44%)	Apr-15	89.75% (95.06%)	Apr-16	62.81% (55.71%)
May-12	63.75% (48.82%)	May-13	70.03% (81.45%)	May-14	60.30% (55.13%)	May-15	59.75% (47.11%)	May-16	80.45% (89.06%)
Jun-12	61.33% (47.15%)	Jun-13	59.44% (40.15%)	Jun-14	53.26% (22.14%)	Jun-15	70.19% (96.60%)	Jun-16	60.31% (70.58%)
Jul-12	63.25% (66.15%)	Jul-13	66.71% (78.86%)	Jul-14	61.90% (79.17%)	Jul-15	55.65% (63.58%)	Jul-16	55.67% (60.41%)
Aug-12	86.41% (32.78%)	Aug-13	59.76% (71.53%)	Aug-14	71.47% (81.74%)	Aug-15	87.71% (95.95%)	Aug-16	81.51% (79.10%)
Sep-12	67.79% (96.49%)	Sep-13	64.34% (70.17%)	Sep-14	65.83% (74.11%)	Sep-15	60.45% (45.65%)	Sep-16	90.05% (87.52%)
Oct-12	64.30% (44.57%)	Oct-13	58.95% (43.67%)	Oct-14	92.29% (63.64%)	Oct-15	62.82% (64.99%)	Oct-16	91.28% (97.85%)
Nov-12	57.73% (54.61%)	Nov-13	61.12% (49.91%)	Nov-14	54.76% (73.60%)	Nov-15	62.08% (53.26%)	Nov-16	61.60% (52.58%)
Dec-12	60.34% (85.59%)	Dec-13	64.42% (41.51%)	Dec-14	58.80% (37.97%)	Dec-15	63.35% (90.19%)	Dec-16	63.36% (67.29%)

4.3. Model Comparison

The accuracy of Generalized Linear Mixed Model and Model-Based Random Forest for mapping the relative risk associated with HFMD for Da Nang City, Vietnam were compared based on RMSE and MSE for retrospective analysis or calibration (past data, 2012-2016), and prospective analysis or validation (2012-2017). Table 12 shows both MSE and the RMSE each for retrospective and prospective analysis.

The results from both analyses show that MBRF model is more accurate in terms of predicting the relative risk on past data and for future outbreaks of the disease, RMSE equal to 0.4372 (MBRF) and 0.8871 (GLMM), respectively. It was also observed that the MBRF prospective RMSE is lower than the GLMM retrospective (calibration) and prospective (validation) RMSE.

Figures 36 and 37 show the predicted vs observed values for both models along with their R^2 values. These were calculated by taking the square of the correlation between the actual and predicted outcomes. The higher values from the MBRF shows that it has higher predictive accuracy as the values close to 1 represents predictions close to the actual.

Figure 38 shows the important variables within the model for each district. Our models show that temperature was the most important variable for three of the districts, Son Tra, Thanh Khe and Hoa Vang in gaining a more accurate prediction while the remaining four districts, the month which represents the temporal trend was the important variable. These results were discussed more in details in the discussion section.

Table 10: Comparison of the RMSE and MSE of both models

Model	Retrospective MSE (RMSE)	Prospective MSE (RMSE)
GLMM	0.7869 (0.8871)	0.8668 (0.9310)
MBRF	0.1907 (0.4372)	0.3146 (0.5609)

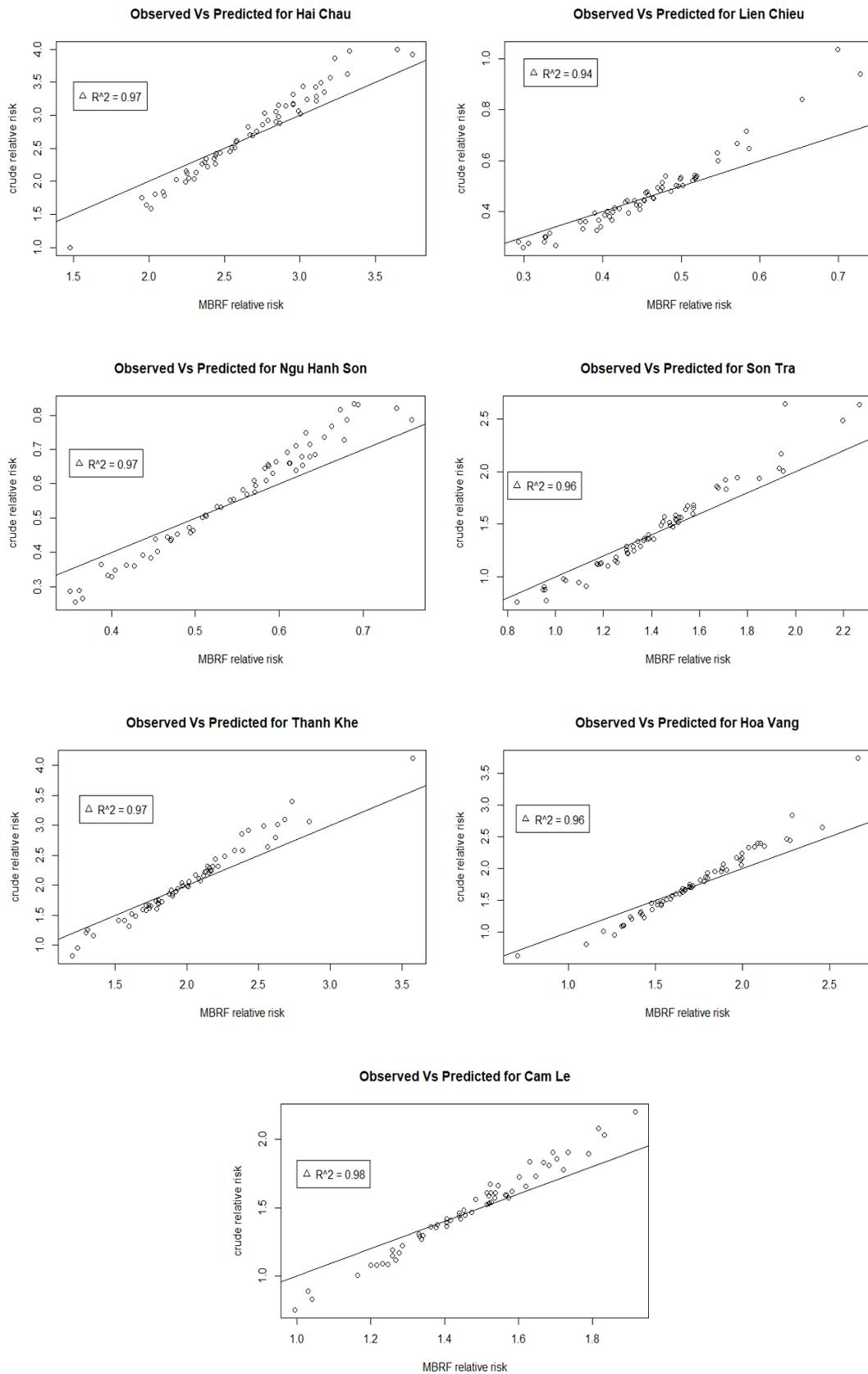


Figure 36: Crude RR vs MBRF RR for Da Nang City

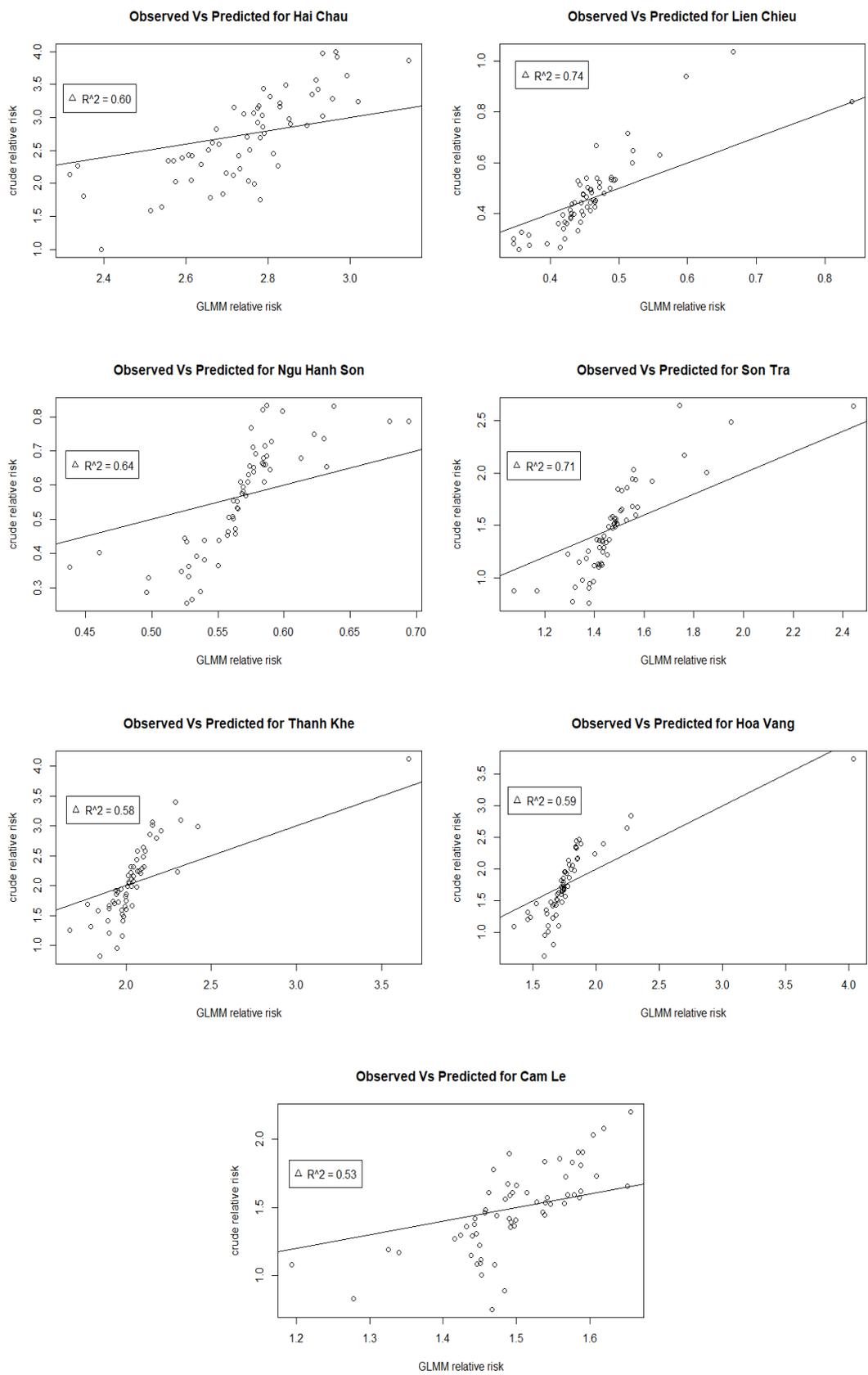


Figure 37: Crude RR vs GLMM RR for Da Nang City

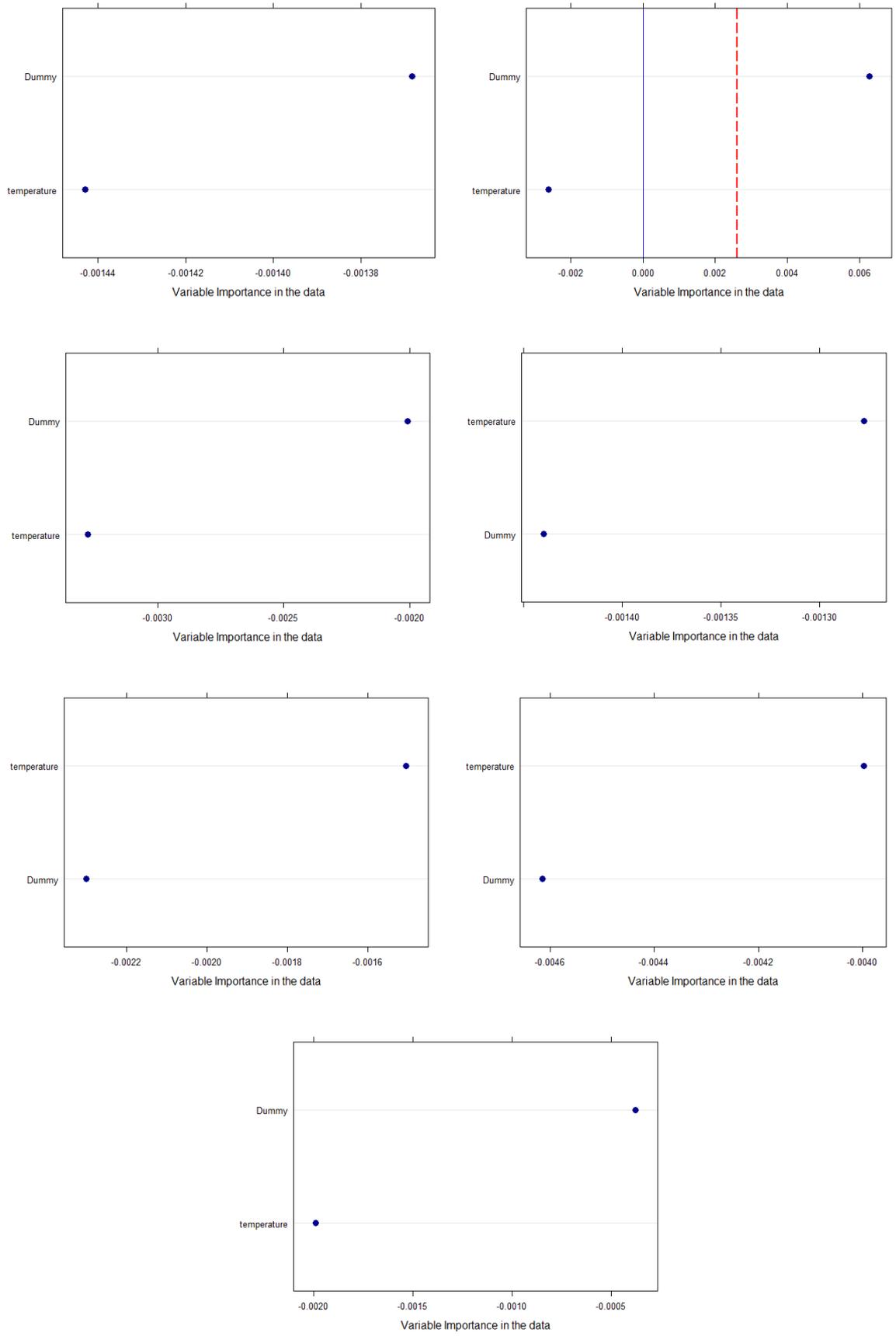


Figure 38: Variable Importance plot for the seven districts in MBRF model

5. DISCUSSIONS

In this section, four main findings were discussed. First, the relative risk of HFMD was discussed. Secondly, the decomposed time series with a specific focus on the seasonal pattern was discussed. Thirdly, the effects of meteorological and socio-demographic factors on the relative risk associated with HFMD in Da Nang City, Vietnam were discussed. Finally, evaluating the performance of both models and identifying the possible reasons why MBRF outperformed GLMM in terms of predictions were discussed. Our findings were compared with similar research to check for consistencies and contradictions/mismatches. The section concludes with limitations and recommendations of this research.

5.1. Incidence of HFMD in Da Nang city, Vietnam

As mentioned prior, a total number of 11,486 cases of Hand, Foot, and Mouth disease were reported for seven districts within Da Nang city over the 60 months. i.e. from January 2012 to December 2016. The average monthly cases across Da Nang city was 27.35 (Sd:19.36). Lien Chieu contributed the largest proportion of the total HFMD cases with 2054, a monthly average of 34.23 (Sd: 22.60) while Thanh Khe contributed the least, a total of 1324 cases, a monthly average of 22.07 (Sd: 14.33). It can be arguably stated that the monthly cases varied with time with most of the observed cases occurred during the first six months of every year for every district with the only exception being 2016 where most of the cases of HFMD occurred in the last six months. The districts with the highest number of cases were all located on the outskirts of the city. Examples were Lien Chieu, Hoa Vang, and Son Tra.

5.2. Predicted relative risk

Our models have identified two districts (Hai Chau and Thanh Khe) of Da Nang City, Vietnam with the highest relative risk of HFMD. This can be attributed to these districts having the largest population density, smallest population, most number of pre-schools (43 and 39 respectively) (General Statistics Office Of Vietnam, 2017), central of the urban areas thus having the most neighboring districts, and highest land surface temperature (L. Liu & Zhang, 2011). The districts with opposite conditions had the lowest relative risk. This finding is consistent with other findings in the literature. For example, Chang et al. (2002) suggested that school children share toys along with other items among each other, that contribute to the virus infections. The authors stated that it could be transmitted through the fecal-oral route and respiratory droplets of the children. Li et al. (2013) stated that temperature affects the survival and transmission of HFMD, hence with the highest land surface temperature in those districts, it can be inferred that greater transmission of HFMD is possible, thus higher the risk. Also, other evidence has suggested the higher temperature may lead to a higher risk of a person getting infected (Song et al., 2015; Urashima et al., 2003). Similarly, the relative risk tends to be higher in urban areas than in rural areas (Yan et al., 2014). Although

not in the same geographic location, Fabre, (2015) and Huang et al., (2014) stated that a densely populated area tends to expedite the spread of HFMD.

5.3. The influence of meteorological and socio-demographic factors

The influence of meteorological and socio-demographic factors on diseases is an ongoing field of study in recent years. These factors have been recognized as important factors influencing HFMD outbreaks as this disease generally displays seasonality (Yu et al., 2019a). Previous literature has shown that meteorological factors influence HFMD occurrences. Temperature, humidity, sunlight, rainfall were shown to be significantly associated with HFMD occurrences (Chen et al., 2014; Hii et al., 2011; Urashima et al., 2003; Huang et al., 2013). These meteorological factors influence HFMD in two main ways: 1. Affecting the external environment resulting in a change to the biological activity and transmission of the pathogen; and 2. Impacting the behaviour of humans (Tian et al., 2018).

In this study, two models were developed to predict the relative risk of HFMD in space and time, considering meteorological and socio-demographic factors of Da Nang City for the period 2012-2016. Firstly, four meteorological variables, monthly temperature, humidity, sunlight, and rainfall were assessed for statistical significance at lag 0 with the occurrence of HFMD. Average monthly temperature was shown to be the only variable to have statistical significance at lag 0 with the occurrence of HFMD in the districts of Da Nang City. Three reasons could be attributed to this. First, assessing the correlation between temperature and humidity, temperature and sunlight have shown that both humidity and sunlight have a strong correlation with temperature. Because temperature has a stronger correlation with HFMD occurrences, it was used as the main meteorological variable. Secondly, in a univariate model, these variables showed statistical significance for some of the districts but not all. However, as a multivariate model, only temperature showed statistical significance. This might be attributed to the fact that different geographic units within the study area are exposed to different climatic or meteorological variables. Finally, as stated, only at lag 0 the association was tested for statistical significance, however, their association may have been at different time scales, i.e. the meteorological variables may have been associated with HFMD occurrence in different lags (J. Liao et al., 2016).

Within the time series predicted relative risk, statistical evidence in our findings show a variation of the seasonality of the relative risk associated with the average monthly temperature within Da Nang City, Vietnam. Two different seasonal patterns amongst the seven districts were identified. Pattern one showed two peaks every year, one in June (highest peak) and a small one in August, i.e. from late spring to summer or in the case of Da Nang, at the end of the dry season. This pattern also showed two troughs, July and January, with the deepest trough in January. The other pattern showed the reverse with peaks in January (highest peak) and a small peak in July, the troughs in June and in August with the deepest trough occurring

in June every year. Our findings of late spring into summer were consistent with other works of literature such as Yu et al. (2019), Cheng et al. (2014) and Zhang et al. (2016).

Peaks in pattern one occurred when the temperature was at its maximum while pattern two showed peaks in January when the temperature was at its minimum. One major observation was that for the two central districts, Hai Chau and Thanh Khe, with the highest relative risk, the seasonal pattern was negatively correlated with seasonal temperature, i.e. when the temperature increased to its maximum, the relative risk decreased to its minimum. However, when the maximum temperature has reached, the relative risk within those districts starts to increase. This mismatch in those districts clearly shows that maximum temperature was the driving factor in the sudden increase of the relative risk. It can be inferred that the effect of temperature was at a different lag for these districts and may have been positively correlated at different lags. Phung et al. (2018) examined the province-specific association between monthly HFMD and climatic factors in Vietnam while controlling for spatial lag, seasonality and long-term trend, using a Generalized Linear Model with Poisson family. The results showed that the climate-HFMD relationship varied by regions, different lags, and provinces across Vietnam.

The districts that showed to have a negative correlation with temperature was found to have a positive correlation ranging from 0.61 to 0.75 with relative humidity, i.e. the peak in January occurs when the relative humidity was the highest. Also, a univariate model of observed cases against the relative humidity showed the p-values of these districts were all less than 0.05 and as such was statistically significant. Therefore, it can be inferred that the relative humidity contributed to the high relative risk within those districts and its mismatch with temperature. This finding is consistent with other pieces of literature that shows high relative humidity increases the risk of HFMD (Yang et al., 2018; Huang et al., 2013; Yang et al., 2017). Due to high relative humidity, the virus tends to be attached to toys, limit sweating and the metabolism of children thus increasing the risk associated with HFMD (Yang et al., 2018).

As it relates to the effect of socio-demographic on the relative risk, our models showed an overall negative correlation. The districts with the highest population or least dense resulted in the lowest risk and vice versa. However, the models have also shown that there were periods where an increase in population resulted in an increase in the relative risk. These periods of increasing risk with increasing population were on average for two months. For the effect of the meteorological factors, a delayed effect was observed as our models showed that there were different seasonal relative risks within the districts of Da Nang city. Among the seven districts within Da Nang city, five districts show a positive relationship, i.e. in phase with temperature, while the remaining 2 districts show a negative relationship, i.e. out of phase with temperature.

5.4. Spatio-temporal residuals

The spatial-temporal residuals which represent variation not yet explained by the temporal trend and the seasonal risk factor temperature spatiotemporal pattern follow the same pattern of the smooth relative risks of each predicted model. Reason can be argued to be because the trend and the seasonality due to temperature do not have a spatial structure. Also, literature has shown other meteorological factors such as wind speed (Y. Liao, Ouyang, Wang, & Xu, 2015) and precipitation (W. Zhang et al., 2016a) to have an influence on HFMD. Similarly, none meteorological factors such as socio-economic levels, health and medical facilities access, surveillance and controlling capacities have all been mentioned as potential factors of HFMD (Tian et al., 2018). All these factors can be inferred as factors influencing the residuals variations.

5.5. Performances of the models

Our final findings focus on the performance of the models. This research was the first to my knowledge that uses MBRF to predict the relative risk with the residuals arising from this model being split into spatial-temporal structured and unstructured. Likewise, a GLMM model was developed to predict the relative risk. The MBRF model was inferred to be the better model as it relates to predicting the relative risk of HFMD associated with meteorological factors on a retrospective (past) and prospective (future) data. The MBRF MSE and RMSE were substantially very low as compared to the GLMM for both on the developed calibration models (past data) and the validation models (future data). MSE represents the squared difference between predicted and observed parameter, with RMSE being the square root of the MSE (Dietrich, 2008). As this research was the first to compare the accuracy of the predictions from both models, reasons for their differences are inferred from similar researches.

Random Forest makes use of a random selection of the predictor variables during the process of splitting a tree node and constructs trees on random samples of learning data with replacement, i.e. bootstrapping (Garge et al., 2013). Thus, each tree is constructed based on a different set of training samples. For every prediction, some of the data is not used in the prediction, termed Out of Bag observations (Cutler, Cutler, & Stevens, 2009). To calculate the importance of a variable m for a single tree, predictions are first carried out on the out of bag observations. Furthermore, the values of the variable m are randomly permuted or reshuffle randomly, keeping the other variables fixed. Predictions are then computed on these new values, giving two sets of out of bag predictions for each observation, i.e. a set for the real data and the other for the permuted variable m . The variable importance is calculated as an error rate from the permuted data and the real data. The overall variable importance for is calculated by averaging over all the observations (Cutler et al., 2009). A large value indicates an important predictor in producing accurate predictions (Cutler, Cutler, & Stevens, 2009; Garge et al., 2013).

MBRF is an extension to Random Forest , a machine learning algorithms (MLA). Newton & Wernisch (2017) stated that the major advantage machine learning methods have over a Bayesian method is that within

the MLA features in the data that are important for accurate predictions are easily identified. However, within the Bayesian modelling approach this is not the case. Bayesian approach requires prior knowledge that should be applied to the model. According to Newton & Wernisch, (2017), these priors can be subjective and incomplete. This variable importance shows which variables significantly affect the prediction and how much its squared error is improved (Cutler et al., 2009). Garge et al., (2013) stated variable importance assessment is ranking variables within the predictor set according to how important they are in producing accurate predictions. The ability of MBRF models to give preference to the most important variable for prediction can be argued to be a major advantage as to why it was more accurate than the GLMM. Although vague priors were specified to allow the model GLMM to get most of its information from the data, the MBRF with no specified prior still gives better predictive accuracy.

Another reason that can be argued is the different seasonal pattern of the relative risk based on temperature for the districts. As seen in the results section, MBRF showed two of the districts having different seasonal patterns for the relative risk as compared to GLMM. These predicted values will affect the overall prediction accuracy of the models. Also, the accuracy of MBRF can be attributed to the inability of the MBRF to smooth the relative risk, with values close to the crude relative risk obtained. These results are consistent with literature as it was stated that ensemble methods like random forest tends to be the most accurate regression tools currently available for data scientists (Breiman, 2001). Thus, with MBRF being an extension of random forest, this research shows that it was one of the most accurate regression tools available. Finally, the R^2 which represents the ratio of the sum of the squares explained by a regression model and “total” sum of squares around the mean” (Itaoka, 2012) were shown in the results section. Values of R^2 lies between 0 and 1, with 1 implying that most of the variability in the dependent variable was explained by the regression model. This was used to assess the goodness of fit of regression models (Bewick, Cheek, & Ball, 2003). Those plots along with the high R^2 values confirm that the MBRF model was more accurate in terms of predicting the relative risk.

There were two disadvantages observed of MBRF. Firstly, its inability to significantly smooth the relative risk. The predicted spatiotemporal smooth relative risk of the MBRF follows a similar pattern of the crude relative risk. However, this is an advantage for GLMM, as with prior specified, it was able to smooth the relative risk. Secondly, it overpredicts the relative risk in some districts at a different time period. Further research is required as a recommendation as to why these disadvantages arises.

6. LIMITATIONS AND RECOMMENDATIONS

This research has a few limitations and recommendations and are stated as follows: the meteorological variables used were obtained from one monitoring center and was used to represent the meteorological conditions of all the districts in Da Nang. Therefore, as a recommendation, spatial variation in these variables can be considered. However, for such a small city, a lot of spatial variabilities is not expected. Secondly, to estimate the monthly population of under 5s, a constant monthly growth rate was applied to the yearly data. Thirdly, limitation in the data, 5 years for 7 districts, hence why Bayesian was applied. It would be ideal to analyze data for a much longer time period and more districts. Fourthly, the study was carried out at the district level and as such cannot be applied for a different level of spatial aggregation (Tian et al., 2018). Finally, this research only takes into consideration lag 0, however, delayed effects or different lags for the remaining meteorological factors were not assessed. As a recommendation, different lags of the meteorological factors can be considered.

7. CONCLUSIONS

The main objective of this research was to identify and model the seasonality of HFMD, considering the dynamics of the population at risks and the spatial-temporal structures. Two models, GLMM and MBRF were developed to predict the relative risk of HFMD in space and time for Da Nang City. This time series predicted relative risk was decomposed into temporal trend, seasonality, and spatial-temporal residuals components. To achieve this main aim, the research had four sub-objectives along with seven research questions.

As it relates to sub-objective one, MBRF was the more accurate model in terms of predicting the relative risk associated with HFMD for Da Nang City, Vietnam for both the validation and calibration model. Sub-objective two was to examine the spatio-temporal distribution of HFMD based on the modelled relative risk of HFMD. For both models, Hai Chau was the district identified to have the highest relative risk while Lien Chieu the lowest relative risk associated with HFMD. Two different seasonal patterns amount the seven districts for both models were identified. Pattern one showed two peaks every year, one in June (highest peak) and the other in August with the deepest trough in January. The other pattern showed the reverse with the highest peak occurred in January and the peaks fell in June and in August with the deepest trough occurring in June every year. This season was stationary as it repeats itself every year.

The next sub-objective was to assess the effects of meteorological and socio-demographic factors on the relative risk associated with HFMD in Da Nang City, Vietnam. For the effect of the meteorological factors, a delayed effect was observed as our models showed that there were different seasonal relative risks associate with temperature within the districts of Da Nang city. It was also observed relative risks were higher for the district with the smallest population, while the districts with the larger population having a lower risk. This clearly shows that the population and the relative risk have an inverse relationship, i.e. as the population increased, the relative risk decreased. Finally, the last sub-objective was to develop a Bayesian Space-Time Conditional Autoregressive (BSTCAR) model for the spatial-temporal structured residuals. The result of this component was used to determine the overall relative risk and were shown in the results section.

In conclusion, the models showed that temperature had a significant effect on relative risk of HFMD in Da Nang City. This has resulted in the different seasonal pattern of the disease risk within the districts. Similarly, Model Based Random Forest was shown to be an accurate model in terms of predicting the relative risk associated with HFMD. Therefore, this research provides a model for future works and scientific evidence that meteorological monitoring should be considered to help to fight HFMD against susceptible populations.

LIST OF REFERENCES

- Abad, M. J., Bedoya, L. M., & Bermejo, P. (2013). Essential Oils from the Asteraceae Family Active against Multidrug-Resistant Bacteria. In *Fighting Multidrug Resistance with Herbal Extracts, Essential Oils and Their Components* (pp. 205–221). Elsevier. <https://doi.org/10.1016/B978-0-12-398539-2.00014-8>
- Anderson, C., Lee, D., & Dean, N. (2014). Identifying clusters in Bayesian disease mapping. *Biostatistics*, *15*(3), 457–469. <https://doi.org/10.1093/biostatistics/kxu005>
- Anderson, C., & Ryan, L. M. (2017). A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, *14*(2). <https://doi.org/10.3390/ijerph14020146>
- Baquero, O. S., Santana, L. M. R., & Chiaravalloti-Neto, F. (2018). Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLOS ONE*, *13*(4), e0195065. <https://doi.org/10.1371/journal.pone.0195065>
- Barnett, A., & Dobson, A. (2013). *Analysing Seasonal Health Data. Journal of Chemical Information and Modeling* (Vol. 53). <https://doi.org/10.1017/CBO9781107415324.004>
- Basile, L., Oviedo de la Fuente, M., Torner, N., Martínez, A., & Jané, M. (2018). Real-time predictive seasonal influenza model in Catalonia, Spain. *PLOS ONE*, *13*(3), 1–15. <https://doi.org/10.1371/journal.pone.0193651>
- Bauer, C., & Wakefield, J. (2018). *Stratified space-time infectious disease modelling, with an application to hand, foot and mouth disease in China. Appl. Statist* (Vol. 67). <https://doi.org/10.1111/rssc.12284>
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in Medicine*, *14*(21–22), 2433–2443. <https://doi.org/10.1002/sim.4780142112>
- Besag, J. (1974). *Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 36). Retrieved from https://www.cise.ufl.edu/~anand/fa11/Besag_spatial_interaction.pdf
- Besag, J., York, J., & Molli, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1), 1–20. <https://doi.org/10.1007/BF00116466>
- Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. <https://doi.org/10.1186/cc2401>
- Breiman, L. (1994). *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Breiman, L. (2001a). Random Forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001b). *Statistical Modeling: The Two Cultures. Statistical Science* (Vol. 16). Retrieved from https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726
- Breslow, N. E., & Clayton, D. G. (1993). *Approximate Inference in Generalized Linear Mixed Models. Source: Journal of the American Statistical Association* (Vol. 88). Retrieved from <https://www.jstor.org/stable/pdf/2290687.pdf?refreqid=excelsior%3Ac7e868b6a9291699c6f14905c00bf32>
- Burrough, P. A., & Mcdonnell, R. A. (1998). *Principles of Geographical Information Systems*. Retrieved from <https://pdfs.semanticscholar.org/d8e0/b6b225b36cac23608b41a51e13ddb2746cbd.pdf>
- Chang, L.-Y., King, C.-C., Hsu, K.-H., Ning, H.-C., Tsao, K.-C., Li, C.-C., ... Lin, T.-Y. (2002). *Risk Factors of Enterovirus 71 Infection and Associated Hand, Foot, and Mouth Disease/Herpangina in Children During an Epidemic in Tainan*. Retrieved from <http://www.pediatrics.org/cgi/content/full/109/6/e88>
- Chen, C., Lin, H., Li, X., Lang, L., Xiao, X., Ding, P., ... Liu, Q. (2014a). Short-term effects of meteorological factors on children hand, foot and mouth disease in Guangzhou, China. *International Journal of Biometeorology*, *58*(7), 1605–1614. <https://doi.org/10.1007/s00484-013-0764-6>
- Chen, C., Lin, H., Li, X., Lang, L., Xiao, X., Ding, P., ... Liu, Q. (2014b). Short-term effects of meteorological factors on children hand, foot and mouth disease in Guangzhou, China. *International Journal of Biometeorology*, *58*(7), 1605–1614. <https://doi.org/10.1007/s00484-013-0764-6>
- Cheng, J., Wu, J., Xu, Z., Zhu, R., Wang, X., Li, K., ... Su, H. (2014). Associations between extreme precipitation and childhood hand, foot and mouth disease in urban and rural areas in Hefei, China. *Science of The Total Environment*, *497–498*, 484–490. <https://doi.org/10.1016/J.SCITOTENV.2014.08.006>

- Cohen, J., Dibner, M. S., & Wilson, A. (2010a). Development of and access to products for neglected diseases. *PLoS One*, 5(5), e10610. <https://doi.org/10.1371/journal.pone.0010610>
- Cohen, J., Dibner, M. S., & Wilson, A. (2010b). Development of and Access to Products for Neglected Diseases. *PLoS ONE*, 5(5), e10610. <https://doi.org/10.1371/journal.pone.0010610>
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2009). High-Dimensional Data Analysis in Cancer Research, (December). <https://doi.org/10.1007/978-0-387-69765-9>
- Deng, T., Huang, Y., Yu, S., Gu, J., Huang, C., Xiao, G., & Hao, Y. (2013). Spatial-Temporal Clusters and Risk Factors of Hand, Foot, and Mouth Disease at the District Level in Guangdong Province, China. *PLoS ONE*, 8(2), e56943. <https://doi.org/10.1371/journal.pone.0056943>
- Dietrich, B. J. (2008). Mean Square Error (MSE) In: Encyclopedia of Survey Research Methods. <https://doi.org/10.4135/9781412963947>
- Du, Z., Zhang, W., Zhang, D., Yu, S., & Hao, Y. (2016). The threshold effects of meteorological factors on Hand, foot, and mouth disease (HFMD) in China, 2011. *Scientific Reports*, 6, 36351. <https://doi.org/10.1038/srep36351>
- Duan, C., Zhang, X., Jin, H., Cheng, X., Wang, D., Bao, C., ... Min, J. (2019). Meteorological factors and its association with hand, foot and mouth disease in Southeast and East Asia areas: A meta-analysis. *Epidemiology and Infection*, 147(November). <https://doi.org/10.1017/S0950268818003035>
- Elliot, P., & Wartenberg, D. (2004). Spatial Epidemiology: Current Approaches and Future Challenges. *Environmental Health Perspectives*, 112(9), 998. <https://doi.org/10.1289/ehp.6735>
- Fabre, G. (2015). The Chinese healthcare challenge Comment on “Shanghai rising: avoidable mortality as measured by avoidable mortality since 2000” *International Journal of Health Policy and Management*, 4(3), 195–197. <https://doi.org/10.15171/ijhpm.2015.36>
- Francis, L., Hunte, S.-A., Marie Valadere, A., Polson-Edwards, K., Asin-Oostburg, V., & James Hospedales, C. (2015). Zika virus outbreak in 19 English-and Dutch-speaking Caribbean countries and territories, 2015-2016 Special report Suggested citation. <https://doi.org/10.26633/RPSP.2018.120>
- Garge, N., Bobashev, G., & Eggleston, B. (2013). Random forest methodology for model-based recursive partitioning: The mobForest package for R. *BMC Bioinformatics*, 14. <https://doi.org/10.1186/1471-2105-14-125>
- Garge, N., Eggleston, B., & Bobashev, G. (2018). *Package “mobForest” Type Package Title Model Based Random Forest Analysis*. Retrieved from <https://cran.r-project.org/web/packages/mobForest/mobForest.pdf>
- General Statistics Office Of Vietnam. (2017). General Statistics Office Of Vietnam. Retrieved August 8, 2018, from https://www.gso.gov.vn/default_en.aspx?tabid=774
- Gou, F., Liu, X., He, J., Liu, D., Cheng, Y., Liu, H., ... Hu, W. (2018). Different responses of weather factors on hand, foot and mouth disease in three different climate areas of Gansu, China. *BMC Infectious Diseases*, 18(1), 15. <https://doi.org/10.1186/s12879-017-2860-4>
- Hii, Y. L., Rocklöv, J., & Ng, N. (2011). Short Term Effects of Weather on Hand, Foot and Mouth Disease. *PLoS ONE*, 6(2), e16796. <https://doi.org/10.1371/journal.pone.0016796>
- Huang, J., Wang, J., Bo, Y., Xu, C., Hu, M., & Huang, D. (2014). Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. *International Journal of Environmental Research and Public Health*, 11(3), 3407–3423. <https://doi.org/10.3390/ijerph110303407>
- Huang, Y., Deng, T., Yu, S., Gu, J., Huang, C., Xiao, G., & Hao, Y. (2013). Effect of meteorological variables on the incidence of hand, foot, and mouth disease in children: a time-series analysis in Guangzhou, China. *BMC Infectious Diseases*, 13(1), 134. <https://doi.org/10.1186/1471-2334-13-134>
- Itaoka, K. (2012). *Regression and interpretation low R-squared! Social Research Network 3 nd Meeting Noosa*. Retrieved from https://ieaghg.org/docs/General_Docs/3rd_SRN/Kenshi_Itaoka_RegressionInterpretationSECURED.pdf
- Jannin, J., & Gabrielli, A. F. (2013). Neurological aspects of neglected tropical diseases: an unrecognized burden. *Handbook of Clinical Neurology*, 114, 3–8. <https://doi.org/10.1016/B978-0-444-53490-3.00001-7>
- Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*, 15(1), 276. <https://doi.org/10.1186/1471-2105-15-276>
- Khalilia, M., Chakraborty, S., Popescu, M., Yu, W., Hebert, P., Fuster, V., ... Palmer, D. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, 11(1), 51. <https://doi.org/10.1186/1472-6947-11-51>

- Kim, B. I., Ki, H., Park, S., Cho, E., & Chun, B. C. (2016). Effect of Climatic Factors on Hand, Foot, and Mouth Disease in South Korea, 2010-2013. *PLOS ONE*, *11*(6), e0157500. <https://doi.org/10.1371/journal.pone.0157500>
- Lai, P.-C., Son, F.-M., & Chan, K.-W. (2009). *Spatial Epidemiological Approaches in Disease Mapping and Analysis*. (C. Press, Ed.).
- Law, J. (2016). Exploring the Specifications of Spatial Adjacencies and Weights in Bayesian Spatial Modeling with Intrinsic Conditional Autoregressive Priors in a Small-area Study of Fall Injuries, *3*(May 2015), 65–82. <https://doi.org/10.3934/publichealth.2016.1.65>
- Li, L., Qiu, W., Xu, C., & Wang, J. (2018). A spatiotemporal mixed model to assess the influence of environmental and socioeconomic factors on the incidence of hand, foot and mouth disease. *BMC Public Health*, *18*(1), 274. <https://doi.org/10.1186/s12889-018-5169-3>
- Li, W., Yi, L., Su, J., Lu, J., Zeng, H., Guan, D., ... Ke, C. (2013). *Seroepidemiology of human enterovirus71 and coxsackievirus A16 among children in Guangdong province, China*. *BMC Infectious Diseases*. <https://doi.org/10.1186/1471-2334-13-322>
- Liao, J., Qin, Z., Zuo, Z., Yu, S., & Zhang, J. (2016). Spatial-temporal mapping of hand foot and mouth disease and the long-term effects associated with climate and socio-economic variables in Sichuan Province, China from 2009 to 2013. *Science of the Total Environment*, *563–564*, 152–159. <https://doi.org/10.1016/j.scitotenv.2016.03.159>
- Liao, Y., Ouyang, R., Wang, J., & Xu, B. (2015). A study of spatiotemporal delay in hand, foot and mouth disease in response to weather variations based on SVD: a case study in Shandong Province, China. *BMC Public Health*, *15*(1), 71. <https://doi.org/10.1186/s12889-015-1446-6>
- Lin, C.-H., Tsai, C.-Y., Lee, K.-C., Yu, S.-C., Liao, W.-R., Hou, A. C.-L., ... Chao, M. C.-T. (2018). A Model-Based-Random-Forest Framework for Predicting σ^2 Mean and Variance Based on Parallel I_d . *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *37*(10), 2139–2151. <https://doi.org/10.1109/TCAD.2017.2783304>
- Liu, L., & Zhang, Y. (2011). Urban Heat Island Analysis Using the Landsat TM Data and ASTER Data: A Case Study in Hong Kong. *Remote Sensing*, *3*(7), 1535–1552. <https://doi.org/10.3390/rs3071535>
- Liu, W., Ji, H., Shan, J., Bao, J., Sun, Y., Li, J., ... Zhu, Y. (2015). Spatiotemporal Dynamics of Hand-Foot-Mouth Disease and Its Relationship with Meteorological Factors in Jiangsu Province, China. *PLOS ONE*, *10*(6), e0131311. <https://doi.org/10.1371/journal.pone.0131311>
- Liu, Y., Wang, X., Liu, Y., Sun, D., Ding, S., Zhang, B., ... Xue, F. (2013). Detecting Spatial-Temporal Clusters of HFMD from 2007 to 2011 in Shandong Province, China. *PLoS ONE*, *8*(5), e63447. <https://doi.org/10.1371/journal.pone.0063447>
- Lykou, A., & Ntzoufras, I. (2011). WinBUGS: a tutorial HISTORICAL BACKGROUND, *3*, 385–396. <https://doi.org/10.1002/wics.176>
- Mcculloch, C. E. (1997). *An Introduction to Generalized Linear Mixed Models*. Retrieved from <https://ecommons.cornell.edu/bitstream/handle/1813/31937/BU-1340-MA.pdf;jsessionid=6ECC5F7F8CD2AAA7AF2DF1E140134968?sequence=1>
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Montesinos-López, J. C., Singh, P., ... Salinas-Ruiz, J. (2017). A Bayesian Poisson-lognormal Model for Count Data for Multiple-Trait Multiple-Environment Genomic-Enabled Prediction. <https://doi.org/10.1534/g3.117.039974>
- Naumova, E. N., Jagai, J. S., Matyas, B., DeMaria, A., MacNeill, I. B., & Griffiths, J. K. (2007). Seasonality in six enterically transmitted diseases and ambient temperature. *Epidemiology and Infection*, *135*(2), 281–292. <https://doi.org/10.1017/S0950268806006698>
- Newton, R., & Wernisch, L. (2017). A comparison of machine learning and Bayesian modelling for molecular serotyping, *18*, 606. <https://doi.org/10.1186/s12864-017-3998-6>
- Nguyen, H. X., Chu, C., Nguyen, H. L. T., Nguyen, H. T., Do, C. M., Rutherford, S., & Phung, D. (2017). Temporal and spatial analysis of hand, foot, and mouth disease in relation to climate factors: A study in the Mekong Delta region, Vietnam. *Science of the Total Environment*, *581–582*, 766–772. <https://doi.org/10.1016/j.scitotenv.2017.01.006>
- Nguyen, N. T. B., Pham, H. V., Hoang, C. Q., Nguyen, T. M., Nguyen, L. T., Phan, H. C., ... Tran Minh, N. N. (2014). Epidemiological and clinical characteristics of children who died from hand, foot and mouth disease in Vietnam, 2011. *BMC Infectious Diseases*, *14*, 341. <https://doi.org/10.1186/1471-2334-14-341>

- Phung, D., Nguyen, H. X., Nguyen, H. L. T., Do, C. M., Tran, Q. D., & Chu, C. (2018). Spatiotemporal variation of hand-foot-mouth disease in relation to socioecological factors: A multiple-province analysis in Vietnam. *Science of the Total Environment*, 610–611(August 2017), 983–991. <https://doi.org/10.1016/j.scitotenv.2017.08.158>
- Ryan, S. J., Carlson, C. J., Stewart-Ibarra, A. M., Borbor-Cordova, M. J., Romero, M. M., Cox, S.-A., ... Ahmed, S. (2017). Outbreak of Zika Virus Infections, Dominica, 2016. *Emerging Infectious Diseases*, 23(11), 1926–1927. <https://doi.org/10.3201/eid2311.171140>
- Sarma, N. (2013). Hand, foot, and mouth disease: current scenario and Indian perspective. *Indian Journal of Dermatology, Venereology and Leprology*, 79(2), 165–75. <https://doi.org/10.4103/0378-6323.107631>
- Segurado, P., Araújo, M. B., Kunin, W. E., & Segurado, P. (2006). Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43, 433–444. <https://doi.org/10.1111/j.1365-2664.2006.01162.x>
- Sham, N. M., Krishnarajah, I., Ibrahim, N. A., & Lye, M. S. (2014). Temporal and spatial mapping of hand, foot and mouth disease in Sarawak, Malaysia. *Geospatial Health*, 8(2), 503–507. <https://doi.org/10.4081/gh.2014.39>
- Sofianopoulou, E., Pless-Mulloli, T., Rushton, S., & Diggle, P. J. (2017). Modeling Seasonal and Spatiotemporal Variation: The Example of Respiratory Prescribing. *American Journal of Epidemiology*, 186(1), 101–108. <https://doi.org/10.1093/aje/kww246>
- Solomon, T., Lewthwaite, P., Perera, D., Cardoso, M. J., McMinn, P., & Ooi, M. H. (2010). Virology, epidemiology, pathogenesis, and control of enterovirus 71. *The Lancet Infectious Diseases*, 10(11), 778–790. [https://doi.org/10.1016/S1473-3099\(10\)70194-8](https://doi.org/10.1016/S1473-3099(10)70194-8)
- Song, C., He, Y., Bo, Y., Wang, J., Ren, Z., & Yang, H. (2018). Risk Assessment and Mapping of Hand, Foot, and Mouth Disease at the County Level in Mainland China Using Spatiotemporal Zero-Inflated Bayesian Hierarchical Models. *International Journal of Environmental Research and Public Health*, 15(7). <https://doi.org/10.3390/ijerph15071476>
- Song, Y., Wang, F., Wang, B., Tao, S., Zhang, H., Liu, S., ... Zeng, Q. (2015). Time Series Analyses of Hand, Foot and Mouth Disease Integrating Weather Variables. <https://doi.org/10.1371/journal.pone.0117296>
- Stanaway, J. D. (2013). *Insights from disease ecology: focus on hand, foot and mouth disease in China*. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/24263/Stanaway_washington_0250E_12222.pdf?sequence=1&isAllowed=y
- Straker, L. (2018). Ministry confirms Dengue and Gastro outbreak | NOW Grenada. Retrieved June 5, 2018, from <https://www.nowgrenada.com/2018/06/ministry-confirms-dengue-and-gastro-outbreak/>
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348. <https://doi.org/10.1037/a0016973>
- Tian, L., Liang, F., Xu, M., Jia, L., Pan, X., & Clements, A. C. A. (2018). Spatio-temporal analysis of the relationship between meteorological factors and hand-foot-mouth disease in Beijing, China. <https://doi.org/10.1186/s12879-018-3071-3>
- Tobler, W. R. (2009). *A COMPUTER MOVIE SIMULATING URBAN GROWTH IN THE DETROIT REGION*. Retrieved from [https://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Tobler Waldo \(1970\) A computer movie simulation urban growth in the Detroit region.pdf](https://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-IV/Tobler%20Waldo%20(1970)%20A%20computer%20movie%20simulation%20urban%20growth%20in%20the%20Detroit%20region.pdf)
- Truong, P. N., & Stein, A. (2018). A hierarchically adaptable spatial regression model to link aggregated health data and environmental data. *Spatial Statistics*, 23, 36–51. <https://doi.org/10.1016/j.spasta.2017.11.002>
- Urashima, M., Shindo, N., & Okabe, N. (2003). Seasonal models of herpangina and hand-foot-mouth disease to simulate annual fluctuations in urban warming in Tokyo. *Japanese Journal of Infectious Diseases*, 56(2), 48–53. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12824684>
- Vieira, C. M., Blamires, Diniz-Filho, Jaf, Bini, L. M., & Rangel. (2008). *Autoregressive modelling of species richness in the Brazilian Cerrado*. *Braz. J. Biol* (Vol. 68). Retrieved from <http://www.scielo.br/pdf/bjb/v68n2/a03v68n2.pdf>
- Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8(2), 158–183. <https://doi.org/10.1093/biostatistics/kxl008>
- Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2), 311–324. [https://doi.org/10.1016/S0378-3758\(03\)00111-3](https://doi.org/10.1016/S0378-3758(03)00111-3)

- Waller, L. A., Carlin, B. P., Xia, H., & Gelfand, A. E. (1997). Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, 92(438), 607–617. <https://doi.org/10.1080/01621459.1997.10474012>
- Walter, S. D. (2001). Disease mapping: a historical perspective. In *Spatial Epidemiology* (pp. 223–239). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198515326.003.0012>
- Wang, C., Li, X., Zhang, Y., Xu, Q., Huang, F., Cao, K., ... Guo, X. (2016). Spatiotemporal Cluster Patterns of Hand, Foot, and Mouth Disease at the County Level in Mainland China, 2008–2012. *PLOS ONE*, 11(1), e0147532. <https://doi.org/10.1371/journal.pone.0147532>
- Waweru, N. (2018). Caribbean countries put on alert over dengue fever outbreak - Face2Face Africa. Retrieved February 8, 2019, from <https://face2faceafrica.com/article/caribbean-countries-put-on-alert-over-dengue-fever-outbreak>
- WHO. (2011). A Guide to Clinical Management and Public Health Response for Hand, Foot and Mouth Disease (HFMD). *Public Health*.
- WHO. (2017). SDG 3: Ensure healthy lives and promote well-being for all at all ages. Retrieved June 5, 2018, from <http://www.who.int/sdg/targets/en/>
- WPRO. (2011). *A Guide to Clinical Management and Public Health Response for Hand, Foot and Mouth Disease (HFMD)*.
- WPRO. (2015). *Hand, Foot, and Mouth Disease Situation Update Number* (Vol. 455). Retrieved from http://www.wpro.who.int/emerging_diseases/hfmd_biweekly_20150210.pdf
- Wu, H., Wang, H., Wang, Q., Xin, Q., & Lin, H. (2014). The effect of meteorological factors on adolescent hand, foot, and mouth disease and associated effect modifiers. *Global Health Action*, 7(1), 24664. <https://doi.org/10.3402/gha.v7.24664>
- Wu, X., Hu, S., Kwaku, A. B., Li, Q., Luo, K., Zhou, Y., & Tan, H. (2017). Spatio-temporal clustering analysis and its determinants of hand, foot and mouth disease in Hunan, 2009–2015. <https://doi.org/10.1186/s12879-017-2742-9>
- Yan, L., Li, X., Yu, Y., de Vlas, S. J., Li, Y., Wang, D., ... Cao, W. (2014). Distribution and risk factors of hand, foot, and mouth disease in Changchun, northeastern China. <https://doi.org/10.1007/s11434-013-0069-5>
- Yang, H., Wu, J., Cheng, J., Wang, X., Wen, L., Li, K., & Su, H. (2017). Is high relative humidity associated with childhood hand, foot, and mouth disease in rural and urban areas? *Public Health*, 142, 201–207. <https://doi.org/10.1016/J.PUHE.2015.03.018>
- Yang, Y., You, E., Wu, J., Zhang, W., Jin, J., Zhou, M., ... Huang, F. (2018). Effects of relative humidity on childhood hand, foot, and mouth disease reinfection in Hefei, China. *Science of the Total Environment*, 630(81), 820–826. <https://doi.org/10.1016/j.scitotenv.2018.02.262>
- Yu, G., Li, Y., Cai, J., Yu, D., Tang, J., Zhai, W., ... Qin, J. (2019a). Short-term effects of meteorological factors and air pollution on childhood hand-foot-mouth disease in Guilin, China. *Science of the Total Environment*, 646. <https://doi.org/10.1016/j.scitotenv.2018.07.329>
- Yu, G., Li, Y., Cai, J., Yu, D., Tang, J., Zhai, W., ... Qin, J. (2019b). Short-term effects of meteorological factors and air pollution on childhood hand-foot-mouth disease in Guilin, China. *Science of the Total Environment*, 646, 460–470. <https://doi.org/10.1016/j.scitotenv.2018.07.329>
- Z., Ś., & B., D.-B. (2018). Hand, foot and mouth disease as an emerging public health problem: Case report of familial child-to-adult transmission. *Dental and Medical Problems*, 55(1), 99–104. <https://doi.org/10.17219/dmp/80995> LK - <http://sfx.library.uu.nl/utrecht?sid=EMBASE&iissn=1644387X&id=doi:10.17219%2Fdmp%2F80995&atitle=Hand%2C+foot+and+mouth+disease+as+an+emerging+public+health+problem%3A+Case+report+of+familial+child-to-adult+transmission&stitle=Dent.+Med.+Probl.&title=Dental+and+Medical+Problems&volume=55&issue=1&spage=99&epage=104&aulast=%C5%9Alebioda&aufirst=Zuzanna&aunit=Z.&aufull=%C5%9Alebioda+Z.&coden=&isbn=&pages=99-104&date=2018&aunit1=Z&aunitm=>
- Zhang, W., Du, Z., Zhang, D., Yu, S., & Hao, Y. (2016a). Boosted regression tree model-based assessment of the impacts of meteorological drivers of hand, foot and mouth disease in Guangdong, China. *Science of The Total Environment*, 553, 366–371. <https://doi.org/10.1016/j.scitotenv.2016.02.023>
- Zhang, W., Du, Z., Zhang, D., Yu, S., & Hao, Y. (2016b). Boosted regression tree model-based assessment of the impacts of meteorological drivers of hand, foot and mouth disease in Guangdong, China. *Science of the Total Environment*, 553, 366–371. <https://doi.org/10.1016/j.scitotenv.2016.02.023>
- Zhang, X., Hou, F., Qiao, Z., Li, X., Zhou, L., Liu, Y., & Zhang, T. (2016). Temporal and long-term trend analysis of class C notifiable diseases in China from 2009 to 2014. *BMJ Open*, 6(10), e011038.

<https://doi.org/10.1136/bmjopen-2016-011038>

Zhu, L., Yuan, Z., Wang, X., Li, J., Wang, L., Liu, Y., ... Liu, Y. (2015). The Impact of Ambient Temperature on Childhood HFMD Incidence in Inland and Coastal Area: A Two-City Study in Shandong Province, China. *International Journal of Environmental Research and Public Health*, 12(8), 8691–8704. <https://doi.org/10.3390/ijerph120808691>

Zhu, L., Yuan, Z., Wang, X., Li, J., Wang, L., Liu, Y., ... Liu, Y. (2015). The Impact of Ambient Temperature on Childhood HFMD Incidence in Inland and Coastal Area: A Two-City Study in Shandong Province, China. *International Journal of Environmental Research and Public Health*, 12(8), 8691–8704. <https://doi.org/10.3390/ijerph120808691>