Potential for using climate forecasts in spatio-temporal prediction of dengue fever incidence in Malaysia

Submitted by

Norziha Che Him

to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, October 2015.

I certify that all material in this thesis which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.

....
Norziha Che Him

Abstract

Dengue fever is a viral infection transmitted by the bite of female *Aedes aegypti* mosquitoes. It is estimated that nearly 40% of the world's population is now at risk from Dengue in over 100 endemic countries including Malaysia. Several studies in various countries in recent years have identified statistically significant links between Dengue incidence and climatic factors. There has been relatively little work on this issue in Malaysia, particularly on a national scale. This study attempts to fill that gap. The primary research question is 'to what extent can climate variables be used to assist predictions of dengue fever incidence in Malaysia?'. The study proposes a potential framework of modelling spatio-temporal variation in dengue risk on a national scale in Malaysia using both climate and non-climate information.

Early chapters set the scene by discussing Malaysia and Climate in Malaysia and reviewing previous work on dengue fever and dengue fever in Malaysia. Subsequent chapters focus on the analysis and modelling of annual dengue incidence rate (DIR) for the twelve states of Peninsular Malaysia for the period 1991 to 2009 and monthly DIR for the same states in the period 2001 to 2009.

Exploratory analyses are presented which suggest possible relationships between annual and monthly DIR and climate and other factors. The variables that were considered included annual trend, in year seasonal effects, population, population density and lagged dengue incidence rate as well as climate factors such as average rainfall and temperature, number of rainy days, ENSO and lagged values of these

climate variables. Findings include evidence of an increasing annual trend in DIR in all states of Malaysia and a strong in-year seasonal cycle in DIR with possible differences in this cycle in different geographical regions of Malaysia. High population density is found to be positively related to monthly DIR as is the DIR in the immediately preceding months. Relationships between monthly DIR and climate variables are generally quite weak, nevertheless some relationships may be able to be usefully incorporated into predictive models. These include average temperature and rainfall, number of rainy days and ENSO. However lagged values of these variables need to be considered for up to 6 months in the case of ENSO and from 1-3 months in the case of other variables.

These exploratory findings are then more formally investigated using a framework where dengue counts are modelled using a negative binomial generalised linear model (GLM) with a population offset. This is subsequently extended to a negative binomial generalised additive model (GAM) which is able to deal more flexibly with non-linear relationships between the response and certain of the explanatory variables. The model successfully accounts for the large amount of overdispersion found in the observed dengue counts. Results indicated that there are statistically significant relationships with both climate and non-climate covariates using this modelling framework. More specifically, smooth functions of year and month differentiated by geographical areas of the country are significant in the model to allow for seasonality and annual trend. Other significant covariates included were mean rainfall at lag zero month and lag 3 months, mean temperature at lag zero month and lag 1 month, number of rainy days at lag zero month and lag 3 months, sea surface temperature at lag 6 months, interaction between mean temperature at lag 1 month and sea surface temperature at lag 6 months, dengue incidence rate at lag 3 months and population density.

Three final competing models were selected as potential candidates upon which an early warning system for dengue in Malaysia might be able to be developed. The model fits for the whole data set were compared using simulation experiments to allow for both parameter and negative binomial model uncertainty and a single model preferred from the three models was identified. The 'out of sample' predictive performance of this model was then compared and contrasted for different lead times by fitting the model to the first 7 years of the 9 years monthly data set covering 2001-2009 and then analysing predictions for the subsequent 2 years for lead time of 3, 6 12 and 24 months. Again simulation experiments were conducted to allow for both parameter and model uncertainty. Results were mixed. There does seem to be predictive potential for lead times of up to six months from the model in areas outside of the highly urbanised South Western states of Kuala Lumpur and Selangor and such a model may therefore possibly be useful as a basis for developing early warning systems for those areas. However, none of the models developed work well for Kuala Lumpur and Selangor where there are clearly more complex localised influences involved which need further study.

This study is one of the first to look at potential climatic influences on dengue incidence on a nationwide scale in Malaysia. It is also one of the few studies worldwide to explore the use of generalised additive models in the spatio-temporal modelling of dengue incidence. Although, the results of the study show a mixed picture, hopefully the framework developed will be able to be used as a starting point to investigate further if climate information can valuably be incorporated in an early warning system for dengue in Malaysia.

Acknowledgements

My thanks to my supervisor Professor Trevor Bailey for his support and guidance through my PhD journey in CEMPS at the University of Exeter. Thanks also to Professor David Stephenson and Dr Christopher Ferro as, respectively, my second supervisor and postgraduate mentor in CEMPS. Special thanks to all Exeter Climate System academics and administrative support staff who helped me in various ways including Dr Catherine, Dr Rachel, Dr Alemtsehai, Dr Theo, Dr Stan and Dr Yuwei. Lastly, to all friends - Dr Ummu Atiqah, Dr Asma, Dr Alasdair, Dr Lester, Dr Maria, Dr Robin, Siti Hasmah and Noraliza.

Thank you to the Ministry of Education Malaysia (MOE) and Universiti Tun Hussein Onn Malaysia (UTHM) for my financial support in respect of this study. Appreciation to Dr Lokman Hakim, Director of Disease Control Division, for his kind assistance regarding dengue data, also to the Malaysian Meteorological Department (MMD) and the Department of Irrigation and Drainage Malaysia (DIDM) for supplying temperature and rainfall data.

Special thanks to my husband, Yusliandy Yusof, my adorable sons and daughter; Afham Hakim, Ammar Hafiz and Amna Haura for their love and support. Appreciation also to my mother Siti Melia Awang and my late father, Che Him Mamat, for their encouragement in my studies. Lastly, thanks to my family in law and all my brothers and sisters, Mazrah, Zaidi, Al-Zuhari, Azani, Azman, Ariza, Nor-Roshayati, Kamarul Halim, Khairiah Mahayu, Kamarudin and Kamariah, thank you all for your love and your invaluable encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter introduces the key motivations behind this study and the primary research aims which are addressed. The focus of the study is on exploring the relationship between climatic variables and the incidence of dengue fever in Malaysia and also the potential for using any such relationships to assist in providing early warning forecasts of dengue epidemics. The chapter starts by outlining the growing need to better understand the risk factors associated with dengue fever, both globally and, more specifically, in the context of Malaysia. It then goes on to specify research aims for the remainder of the thesis and concludes by setting out a structure for the subsequent chapters of the study.

1.1 Motivation

Dengue fever (DF) is a viral infection characterised by sudden high fever, severe headache, rash, muscle and joint pains. The virus is transmitted by the bite of female *Aedes aegypti* mosquitoes and infection rates of dengue can be as high as 90% among those who have not been previously exposed to the virus (Gubler, 1998). Dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) are more serious (and potentially fatal) complications of the disease. Guzman and Kouri

(2002) have reported that nearly 40% of the world's population is now estimated to be at risk from DF in over 100 endemic countries and 500,000 people are estimated to be hospitalised every year with DHF.

The number of dengue cases in Malaysia continues to rise annually and DF is now recognised as a significant public health problem in that country (Smith, 1957; Aziz et al., 2012; Chew et al., 2012). Efforts to reduce the number of dengue cases is now a high priority of various internal and external agencies in Malaysia, not least the Malaysian Ministry of Health which has the main responsibility in addressing the situation. Dengue fever was first reported in Malaysia by Skae (1902), followed by dengue hemorrhagic fever and dengue shock syndrome epidemics in 1962 in Penang, and dengue fever cases first became officially notifiable in 1971 (Rudnick et al., 1965; Poovaneswari, 1993). Now in Malaysia it is the responsibility of all medical practitioners to report every case of dengue fever to the nearest Local Health Office within 24 hours from the time it was diagnosed (Narwani et al., 2005). As the reporting systems have developed, and particularly since 1980, the Malaysian Ministry of Health has recorded continual rising annual cases of dengue to the extent that Ang et al. (2010) recently highlighted dengue fever as an urgent major public health threat in the highly urbanised states of Selangor and Kuala Lumpur.

In general, climate is known to have the potential to influence human health through both direct and indirect mechanisms. The direct mechanisms include, for example, episodes of heat or cold stress and extreme events (drought and flood), while the indirect mechanisms include, for example, the impact of climate anomalies on the risk of vector borne infectious diseases such as malaria and dengue through changing environmental conditions for the vector (Connor et al., 2010). According to Gage et al. (2008), prediction of the relative impact of sustained climate change for vector borne diseases is difficult and will require long-term studies that need to look not only at the effects of climate change but also the contributions of other agents of global change. That said, several studies worldwide have revealed relationships between climatic variables and dengue fever and how these interact with other known risk factors such as socio-economic conditions. Such studies have typically used statistical modelling methods of varying sophistication including time-series analysis, multivariate regression, generalised linear models (GLM), generalised linear mixed models (GLMM), generalised additive models (GAM) or generalised additive mixed models (GAMM). Better understanding of how socio-economic and climatic factors can affect the transmission of dengue fever may help in developing early warning systems (EWS) for epidemics of dengue and so widen the effectiveness of responsive measures (surveillance and prevention). Khun et al. (2005) emphasised the importance of developing systems for early identification for dengue epidemics to help health authorities in surveillance and prevention. To be effective such EWS need to be able to target forecasts geographically within a country into smaller areas such as districts, states or regions (Hu et al., 2012).

In Malaysia, the dengue incidence rate (DIR) in the country as a whole in recent years has fluctuated from as low as 27.5 cases per 100,000 population in 1995 to the high level of 132.5 cases per 100,000 population in 2004 (Kumarasamy, 2006). Epidemics of dengue have occurred roughly every four years with major outbreaks recorded in 1974, 1978, 1982 and 1990 (Lam, 1993b) and with a generally similar pattern with increasing incidence since then. Significant work has been introduced since the 1970s on prevention and control programmes to eliminate the *Aedes* mosquitoes and larval breeding habitats and on public education and law enforcement. However, there have been relatively few modelling studies on the relationship between dengue and climatic and other risk factors in Malaysia as a whole and even less work on the practical development of EWS. That background provides the primary motivation for the work developed throughout the subsequent chapters of this study.

One key challenge in pursuing that agenda is the availability of limited information at the local geographical scale in Malaysia, for example models considered in Racloz et al. (2012) were unable to sufficiently account for the spatio-temporal features of the disease because of the limited geographical resolution of available covariates. However, although there has been little work on dengue EWS in Malaysia, there has been more progress in the neighbour of Malaysia to the south. Dengue fever in Singapore was first recognised as an important public health issue in the early 1960s and *Aedes* control programs have been in place since 1969. Forecasting models of dengue fever have also been developed in Singapore. Ma et al. (2008) carried out a study to look at the association between socio-economic variables and dengue incidence in Singapore for five years from 1998 to 2002 and identified significant association between dengue cases and socio-economic or demographic variables, with areas of higher proportion of disadvantaged residents having more dengue cases. Another more recent study in Singapore used Poisson time series modelling including climate factors such as rainfall and temperature up to 16 weeks or 4 months in advance (Hii et al., 2012). Such studies could be the benchmark for encouraging further research, such as that intended in this study, into the relationship between dengue and climatic variables and other risk factors in Malaysia and the potential for developing EWS for dengue based upon such relationships.

Having established that basic motivation, it is perhaps useful at this point to provide more specific detail on some of the issues so far raised; firstly, in relation to dengue on the world stage and, secondly, in the specific context of Malaysia.

As said previously, dengue fever is a vector borne viral infection transmitted by the bite of female *Aedes aegypti* mosquitoes. The number of dengue cases has increased dramatically around the world in recent years due to the absence of vaccines and drugs (WHO, 2012b). The illness is caused by one of four strains of the dengue virus (DENV-1 to DENV-4). All four strains leave multiple symptoms including headache, rashes and increased body temperature. Infection and recovery from one strain of the virus can lead to immunity from that particular strain and that issue complicates the modelling of the disease because information on the serotype of infections is rarely available on any wide scale.

Potential individual and ecological risk factors for the disease are varied including both socio-economic and environmental conditions. On the socio-economic front factors such as age, income, population density, sanitation, drainage and water supply are potentially important. Amongst the environmental considerations, dengue fever is strongly believed to be influenced by climate variability in temperature and precipitation. One useful related measure is also what is commonly referred to as the 'El Niño Southern Oscillation' (ENSO) which refers to variations in sea surface temperature (SST) of the tropical eastern Pacific Ocean and in air surface pressure in the tropical western Pacific. ENSO, or equivalently the Oceanic Niño Index (ONI) has three different levels; El Niño, Neutral and La Niña. ONI is a global set of anomalies, and is a useful tool to define patterns of climate change. The most heavy and strong ENSO was reported to occur in the years 1997 to 1998. This El Niño was associated with disasters such as drought, flooding and forest fires around the world (Mark, 2005). Understanding links between ENSO and infectious diseases, particularly those transmitted by insects such as dengue, could provide improved long range forecasting of an epidemic or epizootic (Anyamba et al., 2006). The extent to which ENSO can be linked to epidemics of dengue is still not clear, but there are strong recommendations that it could be investigated in future epidemic forecasting for public health preparedness (Mathuros et al., 2009).

Studies associating ENSO and other climatic variables to dengue are reviewed in detail in a later chapter of this study; however it is worth making some brief preliminary reference to a selection of some of those here. Studies in Singapore close to Malaysia have been referenced earlier. Farther afield, in Venezuela, Aura and Alfonso (2010) found a significant association between high dengue incidence and lower values of ONI, but lower dengue incidence with higher value of ONI. Meanwhile, in Puerto Rico, Jury (2008) and Earnest et al. (2012b) concluded that the variability of dengue cases was positively related to temperature but weakly associated with local rainfall and ENSO. Hurtado-Diaz et al. (2007) reported every degree increase in SST leading to a 46% increase in dengue cases in San Andre's Tuxtla and 42% in Veracruz for 16 and 20 weeks respectively. Adriana et al. (2012) used Poisson and Negative Binomial GLMs to investigate the effect of seasonal factors and the relationship of climatic variables to dengue counts in Rio de Janeiro in Brazil. The results indicated significant relationships with the minimum temperature and precipitation at lag one month before, with a 1 °C increase in a month's minimum temperature leading to a 45% increased in dengue cases in the following month and a 10-millimeter increase in precipitation leading to a 6% increase in dengue in the following month. Other research in the South East of Brazil (Lowe et al., 2013) used Negative Binomial generalised linear mixed models (GLMM) to relate monthly dengue incidence to climate and non-climate covariates. Their results provided probabilistic predictions of future epidemics of dengue several months ahead and the general modelling framework used could apply to other areas of Brazil and other climate sensitive diseases.

One issue which these studies perhaps emphasise is that whilst there is broad agreement that climatic factors do influence variability in dengue incidence, there is no clear consensus as to the degree of such effects or indeed, in some cases, their direction. How climate contributes to increase or decrease the incidence of vector borne diseases in human populations will depend on local climatic conditions and local non-climatic epidemiologic and ecologic factors (Patz and Olson, 2006). In other words, effects are geographically dependent upon the region of the world in question and are confounded with other non-climatic influences in ways which are possibly also geographically specific. It is clear therefore that one cannot necessarily transfer results from elsewhere in the world directly into the Malaysian context. Rather there is a requirement to explore from scratch climatic and other relationships with dengue in the specific Malaysian context if progress is to be made towards developing dengue EWS in Malaysia.

Turning briefly to that Malaysian context (a topic which is picked up in more detail in a subsequent chapter), there are factors which have been suggested globally as encouraging dengue spread which particularly pertain in that country, such as rapid and relatively unorganised urbanisation and high rates of population growth. The rise in global commerce and tourism, global warming and changes in public health policy could be important factors too (Gubler, 1998). Developing economies such as Malaysia are also often criticised for poor construction planning which then causes floods or droughts through failure to consider climate information adequately.

Studies to assess the level of knowledge, attitude and practices in relation to dengue in Malaysia have been conducted in 2003 and 2006 by Hairi et al. (2003) and Wan Rozita et al. (2006). In such studies the aims are to evaluate dengue control through increasing the health promotion activities and exposure of communities to educational campaigns. Results obtained were mixed but generally not very encouraging. Meanwhile, Shekhar and Huat (1992) have highlighted major weaknesses of current epidemiological research on dengue in Malaysia which include the inadequacy of data and lack of sound statistical methods. They considered available data used so far to be too restricted, collected using methods that are not clearly described, and which lack scientific validity. The public health sector at the international level has recognised geographic information systems (GIS)[1] as a new technology which has an ability to change the health of societies and contribute to public health policy investigation, development and execution. The WHO has reported that GIS are potentially valuable tools in data compilation and presentation especially for environmental data linked to health services. In Malaysia, this has been explored in relation to dengue by Shaharudin et al. (2002) with results showing no significant difference in the geographical distribution of dengue cases between 1999 and 2000.

As regards the few studies more directly relevant to this study, Lam (1993b) believed that it is possible to predict the severity of a dengue epidemic by the strain of the circulating serotype, but Chew et al. (2012) makes it clear that the situation in Malaysia is complex – although his study showed the predominance of dengue virus in the capital city of Malaysia (Kuala Lumpur) was DENV-4, it was also the case that all four dengue serotypes were in circulation. Ibrahim et al. (2011) used five years data of dengue (2007-2011) to simulate a dynamic system to predict the spread of dengue outbreak in Hulu Langat, Selangor, Malaysia and the results showed that mean temperature, total amount of rainfall and the total of dengue cases in the previous period were highly significant in predicting the possibility of a dengue outbreak.

[1]http://www.gis.com/

In summary, there has been relatively little work on systematic modelling of dengue in relation to climatic and other risk factors across the whole of Malaysia. Key questions remain to be investigated — to what extent can relationships be established, to what extent are there regional differences in such influences and to what extent does any of this have the potential to be incorporated into developing EWS for dengue epidemics in Malaysia? This is the background motivation for the work described in the remainder of this study and in the next section this is laid out in the form of more specific research aims.

1.2 Research Aims

Dengue incidence has been found to be statistically significantly linked to climatic factors, such as ENSO, temperature and precipitation in various studies worldwide. However, there has been only a limited amount of such work (both in the variables considered and in geographical coverage) in the particular context of Malaysia. The primary research question to be considered in this study is therefore 'to what extent can climate forecasts be used to assist in making predictions of dengue fever incidence in Malaysia?'.

It is clear that to do this, the study will need to consider as much meteorological data as is readily available in Malaysia as potential multiple covariates, such as rainfall, temperature, humidity, number of days with rain, sea surface temperature etc. and appropriate time lagged values of these variables. It will also need to allow for non-climate factors such as population size and population density. To allow for the dynamic epidemic behaviour temporally lagged values of reported dengue incidence rates may need to be included. All of this will need to be considered at an appropriately practical spatial and temporal resolution - e.g. district or state levels and annual versus monthly. Previous studies (Racloz et al., 2012) have reported the use of various level of spatial scale such as community, district, municipality or city either at daily, monthly or annual collation times and how these may help in producing better and significant results. The intended study will also need to adopt appropriate, contemporary statistical modelling frameworks based, for example, on generalised linear models (GLM) or generalised additive models (GAM) for count data on dengue cases. The data used in the study should be as comprehensive as possible so as to accord with the aims expressed above. Depending upon the results obtained in the study, then the most viable predictive models need to be selected and validated both on training and out-of-sample data. The implications of all of this is that if models with predictive validity for dengue in Malaysia can be established based upon the data considered then results of this work can hopefully guide decision makers in Malaysia at both national and local level in a better understanding of factors significantly contributing to dengue across Malaysia, the extent to which they may or may not be predicted and to develop appropriate public health response strategies.

Narrowing that agenda to specifics and taking into account practical data limitations (as will be discussed in subsequent chapters), the following are the key associated questions intended to be addressed in this study:

- Based upon reasonably extensive data, are there significant relationships between local and global climate variables and the extent of dengue incidence for the 12 states in Malaysia. What are these relationships, what temporal lags are involved? To what extent are these effects state specific or common to Malaysia as a whole? Are any such relationships changing over time or space in Malaysia?
- How are these effects confounded by non-climatic factors - e.g. basic seasonality in the disease incidence, demographic factors etc.?
- How can any significant relationships established in response to both of the above bullet points be built into developing a practical realisable spatial-temporal model to predict dengue incidence for future dengue incidence in Malaysia and on what spatial and temporal resolution?

- What is the predictive validity of any such model? What are the degrees of uncertainty involved? Are these of any use in terms of targetting preventative measures and public health response?
- Depending on the results of the above, what are the issues that relate to improving the predictive validity of such models in Malaysia if that is possible?

1.3 Summary

This chapter has outlined the background and the motivation for the research considered in the subsequent chapters of this study. Accordingly, the layout of subsequent chapters is as follows. The next, Chapter 2, provides important information about Malaysia and its profile in terms of geography, demography, poverty and health and also about the climate system of that country and the sources of climate data considered for use in this study. Chapter 3 then focusses on dengue fever and its transmission both worldwide and in Malaysia, including further review of studies relating to the relationship between climatic factors and other risk factors relating to dengue and the sources of dengue data considered for use in this study. Chapter 4 then outlines the exploratory data analysis, where each of the monthly covariates available in Malaysia from the various data sources from 2001 to 2009 for twelve states in Malaysia were considered in relation to corresponding dengue incidence. Chapter 5 then develops a model framework using the most important covariates and lagged covariates informally identified in the previous chapter. This chapter attempts to identify and test, given the available data sources and associated spatio-temporal resolution, the most appropriate probability models for dengue counts in Malaysia, including selection of suitable climate and other explanatory variables. Chapter 6 then focusses on the predictive power of the possible models built in Chapter 5. In particular this chapter identifies the discrepancies in predictive power of the Malaysia wide model within the more urbanised states of Kuala Lumpur and Selangor. It analyses possible reasons for that and makes reference to the need to pursue more detailed data collection and modelling at the more localised district level within these states. Finally, Chapter 7 summarises the conclusions, results, outputs and recommendations made from the whole of this study.

Chapter 2

Malaysia and its climate

This chapter outlines key aspects of the geography, demography, socio-economic and health profile of Malaysia which are relevant to subsequent chapters of this study. The sources of cartographic and demographic data used in subsequent chapters are also described. The chapter then goes on to discuss the climate of Malaysia and explains the major seasonal variations in climate including the nature of the monsoon cycle. The sources of climate data relevant to subsequent chapters of this study are then described.

2.1 Introduction to Malaysia

Malaysia is a federal constitutional monarchy in Southeast Asia with a total landmass of 329,847 square kilometres, separated by the South China Sea into two regions: Peninsular Malaysia (West) and East Malaysia. The country is situated close to the equator between 1° to 7° north and 99° to 105° east, with Peninsular Malaysia lying north of Singapore, south of Thailand and east of the Indonesian Island of Sumatra, while East Malaysia is situated on the island of Borneo and shares borders with Indonesia and Brunei (see Figure 2.1[1] which indicates the rela-

[1]http://www.globalsecurity.org/military/world/malaysia/maps.htm

tive locations of Peninsular Malaysia (West) and East Malaysia and how Malaysia shares land borders with Brunei, Indonesia and Thailand and maritime borders with the Philippines, Singapore and Vietnam. The country is composed of highland, floodplain and coastal zones. In particular, the Titiwangsa mountain range forms the backbone of Peninsular Malaysia, from southern Thailand running approximately south-southeast over a distance of 480 kilometres and separating the eastern from the western part of that of Peninsular Malaysia. Surrounding these central high regions are the coastal lowlands.

The country is administratively divided into thirteen states (Perlis, Penang, Kedah, Perak, Selangor, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan, Sabah and Sarawak) and three federal territories (Kuala Lumpur, Labuan and Putrajaya). This study will exclude Sabah and Sarawak and focus on the remaining eleven states which comprise Peninsular Malaysia because dengue fever is of significantly less concern in East Malaysia than in the west where the incidence rate is much higher. Because of its unique urban characteristics, the territory of Kuala Lumpur in Selangor will also be treated effectively as a 'state' on its own in subsequent chapters separate from the rest of Selangor. Therefore, the eleven states and the territory of Kuala Lumpur become a total of twelve 'states' (see Figure 2.2) considered in this study with the understanding that 'Selangor' in that context refers to the area in Selangor state outside of the territory of Kuala Lumpur.

The history of the formation of current day Malaysia is somewhat complex. Originally there was a federal system comprising four of the current states of Peninsular Malaysia, namely Pahang, Perak, Selangor and Negeri Sembilan. This system was implemented by the British in 1895. In 1946, it was merged together with the Strait Settlements first established in 1826 (Penang and Melaka) and with the non-malay States comprising Johor, Terengganu, Kelantan, Kedah and Perlis to form an 11 states Malayan Union. This Malayan Union (current day Peninsular Malaysia) was restructured as the Federation of Malaya in 1948 and achieved independence as Malaysia on 31 August 1957. On 16 September 1963, Malaysia united with the eastern region of Sabah, Sarawak and Singapore but in 1965 Singapore was re-

Figure 2.1: *Map of Peninsular Malaysia (left) and East Malaysia (right).*

moved from the federation so forming the 13 states Malaysia of today. The capital city of Malaysia is Kuala Lumpur which is located in the center of Selangor state in Peninsular Malaysia. The constitution of Malaysia declared Islam as the main religion with protected freedom of religion to others. The heads of the Malaysian government are the Prime Minister and the King (known as the Yang di-Pertuan Agong) who is an elected monarch, being chosen by the hereditary rulers of the Malay states every five years. Malaysia's current Prime Minister is Najib Razak and the King is Tuanku Alhaj Abdul Halim Mu'adzam Shah Ibni Almarhum Sultan Badlishah.

2.1.1 Demographic, Socio-economic and Health Profile

The 2010 Population and Housing Census of Malaysia (known as Census 2010) was the fifth decennial census to be conducted since the formation of Malaysia in 1963. The previous censuses were conducted in 1970, 1980, 1991 and 2000. Census 2010 revealed that the total population of Malaysia was 28.3 million, compared with 23.3 million in 2000. The proportion of the population of Malaysia below the age of 15 years decreased to 27.6% compared with 33.3% in 2000. In contrast, the proportion of working age population (15 to 64 years) increased to 67.3% from 62.8%. The proportion of population aged 65 years and over also increased to 5.1% as compared with 3.9% in 2000. Meanwhile, the median age increased from 23.6 years in 2000 to 26.2 years in 2010. The trend of these indicators is in line with the global expected transition towards an aging population albeit in its early stages in Malaysia. The state with the highest population growth rate for the period 2000-2010 was Kuala Lumpur (17.8%), followed by Selangor (2.7%) and Melaka (2.6%). Among the states which experienced lower growth rates were Terengganu (1.4%), Perak (1.4%) and Perlis (1.2%).

The rapid development of Malaysia shows the proportion of urban population increasing to 71.0% in 2010 compared with 62.0% in 2000. Apart from Kuala Lumpur with a 100% urbanisation level, the other states with high level of urbanisation are

Selangor and Penang with 91.4% and 90.8% respectively. States with lower levels of urbanisation are Kelantan (42.4%), Pahang (50.5%) and Perlis (51.4%) (Wan Abd Raof, 2010).

Malaysia as a whole is a multi-ethnic country. The total population was 28.3 million in 2010 — 91.8% Malaysian citizens and 8.2% non-citizens. Amongst the citizens, the principal ethnic groups are Malay (Bumiputera), Chinese and Indian. Other significant groups are the indigenous people of Sabah and Sarawak, including Kadazan, Dusun, Bajau, Murut, Iban, Bidayuh and Melanau. In Peninsular Malaysia, the conventional ethnic divisions of the population are Malay, Chinese, Indian and Other. This 'official' classification was defined so as to reflect the popular conception of race. Malaysian citizens consist of the ethnic groups Bumiputera (67.4%), Chinese (24.6%), Indians (7.3%) and Others (0.7%). Bumiputera have experienced an increasing trend due to high fertility rates while Chinese and Indians have showed a decreasing trend due to low fertility rates (Wan Abd Raof, 2011).

Hirschman (1987) reports the meaning of a Malay as a person who was born locally, habitually speaks Malay, follows Malay customs and professes Islam. Meanwhile, the Chinese and Indian communities consist of descendants of immigrants from China and the India subcontinent. Other is a open category for the small number of Thais, Europeans and other people who do not fit into the three major categories. The Malaysian population has been a blend of varied cultures since early times. About fifteen hundred years ago, Indians and Chinese entered as traders in the Malay Kingdom. Their entry marked the arrival of gold and silks followed by Hinduism and Buddhism. After a thousand years, principles of Islam also marked their entry with Arab Traders in Melaka, followed with the arrival of Portuguese. Although the Malaysia population encompasses several cultures, the old Malay culture is the most prominent, followed by Chinese and Indian influences (Abu Bakar, 1996).

The poverty line in Malaysia was defined in the 1970s after the Malaysian Government brought an explicit poverty eradication principle into national policy. It is based on assessments of the minimum consumption requirements of an average-sized household for food, clothing, shelter and other non-food needs. Small differences exist in the definition of the poverty line between the three main regions of Malaysia (Peninsular Malaysia, Sabah and Sarawak) in terms of mean household size and cost of living but not for differences between rural and urban location. These poverty lines were adjusted for inflation and changes in mean household sizes from 1976 to 2004. However, this situation was revised again by the Malaysian Economic Planning Unit (EPU) along with the United Nations Development Programme (UNDP) and a new poverty line was defined based on each household and averaged to each state and to rural or urban location together with cost of living, household composition and size (Economy, 2010). A study by Muhamed and Haron (2011) revealed the poverty eradication programmes have resulted in considerable reduction of poverty, decreasing the income inequality alongside achieving rapid economic growth especially in Johor. They reported that in Malaysia as a whole the success of the poverty eradication programmes is evidenced by the sharp decline in the incidence of poverty, which decreased from 52.4% in 1970 to 12.4% in 1992 and further decreased to 3.8% in 2009. Meanwhile, in Kelantan as one of the states located in the North East of Peninsular Malaysia, the proportion of the poor households in rural areas remains higher than that of urban poor. That said, a greater portion of the urban households are vulnerable to poverty compared to the rural households of that state (Siwar et al., 2013).

As in many countries, public health and associated health care services have high priority at both local and national government levels in Malaysia. Within the Malaysia Health Care System there are two sectors, public and private. The public sector is divided into Federal Government and State-Local Government. The Federal Government contains Ministry of Health, Armed Forces, Department of Aborigines, Ministry of Home Affairs and Ministry of Education. While, for the State-Local Government, there is Public Health and Prevention, Hospitals, Clinics, Special Institutions, Maternal and Child Health, Nurse and Paramedic Education, Enforcement and Supervision and Licencing. Health care services consist of tax-funded and government-run primary health care centres and hospitals, but there are also fast-growing private services mainly located in physician clinics and hospitals in urban areas. Meanwhile, public sector health services are administrated by the Ministry of Health through its central, state and district offices.

In general terms, Malaysia shares similar major health risks with its neighbours in the same region. Non-communicable diseases now account for most mortality and morbidity but communicable diseases remain a significant concern (Jaafar et al., 2013). According to the WHO (World Health Organisation) report on the health profile of Malaysia[2], the 10 highest causes of mortality in Malaysia in 2010 were as listed in Table 2.1, the top three being: coronary heart disease, stroke and influenza/pneumonia. These top 10 causes accounted for 22.18% of the total fatalities recorded.

Table 2.1: *Malaysia top 10 causes of death.*

Causes of Death	Rate (Cases per 100,000)	Rank (172 Countries)
Coronary Heart Disease	138.75	57
Stroke	75.81	114
Influenza and Pneumonia	65.08	68
Road Traffic Accidents	34.53	20
HIV/AIDS	23.15	57
Lung Disease	19.09	108
Diabetes Mellitus	18.99	128
Lung Cancers	17.93	74
Tuberculosis	17.82	76
Breast Cancer	15.83	100

The health risk from major infectious diseases in Malaysia, such as bacterial diarrhea, dengue fever and leptospirosis is classified as intermediate level. For example in the WHO report mentioned above dengue is ranked 45th as a mortality cause.

[2]http://www.worldlifeexpectancy.com/country-health-profile/malaysia

However, diseases listed in relation to mortality do not necessarily represent those of most relevance in terms of morbidity and infectious diseases such as diarrhea and dengue are very significant in the picture of the disease burden experienced by the local population[3].

2.1.2 Sources of Cartographic and Demographic data

This subsection describes the sources of the cartographic and demographic data used in later chapters of this study.

Polygons of the twelve states (see Figure 2.2) together with their areas and the latitude and longitude of their centroids (see Table 2.2) were identified from Malaysia map shapefiles[4]. Recall that for the purposes of this study, only 12 of the 14 states of Malaysia will be considered (i.e. those of Peninsular Malaysia and not including Sabah and Sarawak). Also that the data used for the state of Selangor will exclude the data for Kuala Lumpur which is treated as a separate state.

Figure 2.2: *Map of the 12 states in the Peninsular of Malaysia.*

[3]http://www.indexmundi.com/malaysia/major-infectious-diseases.html
[4]http://www.diva-gis.org/gdata

The Department of Statistics Malaysia (DSM) is responsible for carrying out the population and housing census once every 10 years, the last census being conducted in 2010. The method used in this census is face to face interview and the information collected includes the number of persons in households together with a wide range of demographic, social and economic characteristics. For the purposes of this study, estimated annual populations[5] were used based on census 2000 and 2010 for each state in the study area with numbers adjusted for under-enumeration, collated from Department of Statistics Malaysia (see Table 2.2). Where monthly figures are used these are obtained from simple linear interpolation from the relevant annual estimates. Population density is taken as number of people in the state per unit area of the state as defined by Hafiz et al. (2012).

Table 2.2: *Distribution of area, latitude, longitude and population in Malaysia.*

State Names	Area (km^2)	Latitude	Longitude	Pop. 2000	Pop. 2010
Perlis	795	6.433	100.200	198,288	227,025
Kedah	9,425	6.116	100.366	1,571,077	1,890,098
Penang	1,031	5.416	100.333	1,231,209	1,520,143
Perak	21,005	4.583	101.083	1,973,368	2,258,428
Selangor	7,960	3.033	101.433	3,941,316	5,411,324
K.Lumpur	243	3.166	101.700	1,305,792	1,627,172
N.Sembilan	6,644	2.716	101.933	829,774	997,071
Melaka	1,652	2.200	102.250	605,239	788,706
Johor	18,987	1.466	103.750	2,584,997	3,233,434
Pahang	35,965	3.800	103.333	1,229,104	1,443,365
Terengganu	12,955	5.333	103.133	880,234	1,015,776
Kelantan	15,024	6.133	102.250	1,287,367	1,459,994

[5]http://www.statistics.gov.my/portal/index.php

2.2 Climate in Malaysia

Climate and associated seasonality are important determinants in the incidence of various diseases both worldwide and in Malaysia. Climate is a key variable in managing the overall burden of health, especially for developing countries where the ability to control climate-sensitive diseases is constrained. This situation will affect most populations in the future and put the lives and well-being of billions of humans at increased risk. To reduce it, the health sector needs to understand and quantify the specific effects of climate variability and change both on the overall disease burden and on the opportunities and effectiveness in the public health response. The aims are to ensure the future adaptation strategies and understanding of the climate impact on the existing disease burden and current interventions. This applies for air-borne diseases, such as asthma and other respiratory infections, also for vector borne diseases such as dengue fever. The next effects of global climate change on such diseases are difficult to forecast. For example, an increase in temperature may increase the formation of ground-level ozone, a pollutant with well-established adverse effects on respiratory health, on the other hand an increase in cold years with the absence of specific interventions, may encourage mosquito population breeding and rising incidence of dengue, but at the same time an increase in warm years with periods of drought will decrease the mosquito population and reduce the incidence of dengue (Costello et al., 2009). Heat and heat waves are also very likely to increase in severity and frequency with increasing global average temperatures (Tanggang et al., 2010). These conditions can be expected to influence human health and well-being in proportion to the degree of heat stress. Heat stress can cause mild cardiovascular problems to severe tissue damage and, in extreme cases, death. These effects are concentrated among vulnerable groups of people such as the elderly, the very young, the malnourished and those with pre-existing respiratory and cardiovascular conditions. The impact of extreme heat on the elderly takes on particular significance in light of the growing increase in the elderly proportion of the population worldwide as we move towards 2050.

It follows that climatic conditions imply health impacts — both direct and indirect. In respect of dengue fever which is the particular focus of this study, simulation model studies (e.g. Mary Ann, 2009) have considered total population and interactions between climate variables and concluded that predicted climate change will make the dengue problem more acute, especially if current control measures concentrated on *Aedes* mosquito vectors prove to be ineffective. It is also possible that previous trends in incidence could reverse in certain locations. Dry spells may favour transmission as they may disrupt normally running streams and leave standing water during drought which could provide a suitable place for mosquito breeding. Rowley and Graham (1968) have reported optimal temperature and humidity levels for adult mosquito longevity and biting activity and went on to determine the optimal temperature and relative humidity for tethered flight activity in female *Aedes aegypti* and the range for both factors for possible sustained flight. In the future, hopefully the health sector will be able to adopt climate information as an effective tool in epidemic early warning systems for dengue. Then, seasonal forecasts of temperature and rainfall, which are useful indicators of the likely occurrence of dengue outbreaks, could be applied in the implementation of a programme of heightened epidemic surveillance. Meanwhile, real-time temperature and rainfall estimates could be used to initiate selective interventions and to support early detection of disease outbreaks.

All of the above implies it is important to understand the climate context in studying a vector borne disease such as dengue in Malaysia and in developing any intervention strategy and this is the topic of the subsequent sections of this chapter.

2.2.1 General Description of Climate

As said earlier, Malaysia is located near the equator and thus the climate is categorised as ‘tropical’. In summary, hot and humid throughout the year is the best description of Malaysian weather with daytime temperature averaging 30°C and overall relative humidity level ranging between 70% and 90% (Wong et al., 2009).

That said, the climates of Peninsular Malaysia and East Malaysia are somewhat different, with Peninsular Malaysia receiving weather predominantly from the mainland while East Malaysia experiences maritime weather instead. There are two monsoon seasons in Malaysia, firstly, the Southwest monsoon from late May to September and secondly, the Northeast monsoon from November to March every year. As the names suggest, the Southwest monsoon distributes more rainfall to the west coast of Peninsular Malaysia, while the Northeast monsoon which comes from the South China Sea and the North Pacific does the same on the east coast of Peninsular Malaysia.

The mountain ranges throughout Malaysia also influence local climate separating the weather into three zones: lowlands, highlands and coastal regions. Basically, the coastal regions have a sunny climate, with temperatures ranging between 23°C and 32°C, and rainfall ranging from 10 centimetres to 30 centimetres a month. The average monthly temperature for lowlands areas ranges from 24.8°C to 30.1°C while for highlands areas it ranges from 15.0°C to 25.4°C.

As said, climate in Malaysia is also characterised by two monsoon regimes. The Northeast monsoon from November to March brings heavy rainfall, particularly to the east coast states of Malaysia such as Kelantan, Terengganu, Pahang and Johor. The Southwest monsoon, which is recorded from late May to September, normally signifies relatively drier weather for the west coast states of Malaysia especially Selangor, Perak and Kedah. However, with two major oceans surrounding — the Pacific Ocean to the east and the Indian Ocean to the west — climate variability is also influenced by conditions in both oceans (Tanggang and Bahari, 2002). Suhaila et al. (2010a) has described the patterns and trends of five selected rainfall indices in Peninsular Malaysia, based on daily rainfall data from 1975 to 2004. They identified that the eastern areas of the Peninsular were strongly influenced by the Northeast monsoon, while the Southwest monsoon had the greatest impact on the western part of the Peninsular, particularly the Northwest.

The El Niño Southern Oscillation (ENSO) is an oceanic-atmospheric phenomenon which is characterised by sustained fluctuations between unusually warm and cold conditions in the tropical Pacific Ocean. As ENSO cycles, the path of the Pacific Jet Stream and other global climate drivers change causing variation in local temperature and precipitation worldwide. The warm condition is referred to as El Niño and the cold condition is referred to as La Niña. Among the various definitions for El Niño, there are some common characteristics of El Niño such as an anomalous warming of surface water, a warm southward-flowing current off the coast to Peru and a duration of 12-18 months. La Niña could be characterised by having criteria such as cooling of the surface water of the eastern and central Pacific Ocean, which occurs somewhat less frequently than El Niño events but causes generally opposite disruptions to global weather patterns. It tends to happen when the Pacific trade winds blow more strongly than usual, pushing the sun-warmed surface water further west and increasing the upward movement of cold water in the eastern regions (Glantz, 2001).

Sea surface temperatures (SST) in the Pacific Ocean are monitored in regions identified as Niño 1 to Niño 4 as shown in Figure 2.3. Each of these regions provides different information about El Niño, La Niña, Neutral and ENSO. SST anomalies are defined as deviations for a specified region from the averaged climate for 1961 to 1990 (refer to World Meteorological Organisation, WMO).

Indonesia, Malaysia and most of the Philippines, are amongst the first areas to experience ENSO-related impacts. For example, sea surface temperatures anomalies in the equatorial east Pacific Ocean increased significantly during July to October 2006 indicating the typical development of El Niño conditions. Positive Outgoing Longwave Radiation (OLR) anomalies, indicative of severe drought conditions were observed across all of Indonesia, Malaysia and most of the Philippines. This dryness continued until the early part of 2007 (Anyamba et al., 2006). The El Niño effect was also strongly felt in Southeast Asia in 1997. The prolonged drought contributed to the development of forest fires and this, coupled with the existing wind pattern, caused widespread haze over the Southeast Asia region.

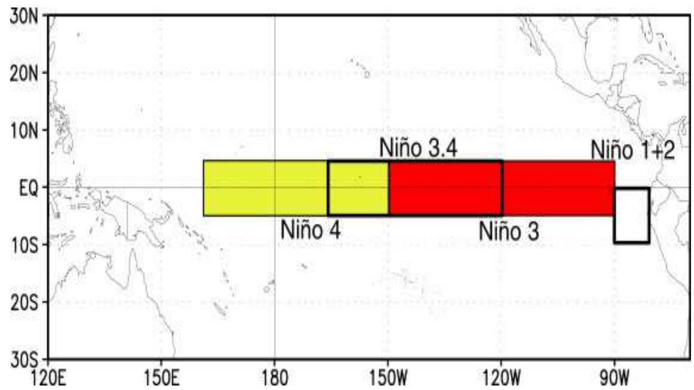

Figure 2.3: *Map showing five regions (referred to as Niño 1, Niño 2, Niño 3, Niño 3.4 and Niño 4) in the Pacific identified as important locations for monitoring winds, sea surface temperatures and rainfall activities, changes that may be associated to varying degrees with El Niño process.*

The relationships between Malaysian temperature and rainfall anomalies, sea surface temperature and ENSO have been discussed in a study by Fredolin and Liew (2004). Such relationships mean that ENSO can have considerable social, environmental and economic impacts in Malaysia. A prolonged drought associated with El Niño frequently causes severe water supply crises, disrupts agricultural activities and destroys rain fed crops besides creating environmental hazards such as haze episodes and forest fires (Juneng and Tanggang, 2008). Ministry of Health (Mohd Ismail, 2007) records show that there was an increase of complaints related to conjunctivitis, bronchitis and asthma among the local population during the haze episodes of 1990, 1991, 1994 and 1997 in Malaysia (Nicol, 1997 and Heil and Goldammer, 2001).

2.2.2 Sources of climate data

Climate data are used in subsequent chapters of this study to investigate the existence of relationships with the incidence of dengue fever. The climatic factors considered are monthly rainfall, number of rainy days and monthly mean temperature over the nine year study period between 2001-2009.

The monthly amount of rainfall was taken from Department of Irrigation and Drainage Malaysia[6] records and based on selected hydrology stations and their locations as supplied. Data on observed monthly rainfall, number of rainy days and monthly mean temperature were also obtained from the Malaysian Meteorological Department[7] for each of the 108 months considered in subsequent chapters.

In addition precipitation 'ncdf' files were obtained from the NOAA Earth System Research Laboratory website via downloading from the Global Precipitation Climatology Project (GPCP) V2.1 which includes a monthly precipitation dataset from 1979 to present that combines observations and satellite precipitation data into 2.5° x 2.5° global grids (Adler et al., 2003). Monthly mean air temperature and monthly mean relative humidity were also downloaded from the National Center for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) Reanalysis datasets which use a state-of-the-art analysis/forecast system to perform data assimilation using past data from 1948 to the present (Kalnay et al., 1996). Further, three datasets were extracted for the period from January 2001 to December 2009 using `R` software (R Core Team, 2010) considering the latitude and longitude of study areas in Malaysia.

Niño 4 is an index used to measure the strength of El Niño and La Niña events relevant to Indonesia, Malaysia and the Philippines and is defined as the departure in monthly sea surface temperature from its long-term mean averaged over the Niño 4 region. The Niño 4 region is in the central Pacific, straddling the dateline and it goes from 160 East to 150 West, and from 5 South to 5 North. The time series of the Niño 4 index was obtained from the Climate Prediction Center (CPC)[8] for the nine year period of the study.

[6]http://www.water.gov.my/

[7]http://www.met.gov.my

[8]http://www.cpc.ncep.noaa.gov/data/indices/

2.3 Summary

This chapter has introduced a geographic, demographic, socio-economic and climatic profile of Malaysia. Sources of related data for the period 2001-2009 on a monthly basis which will be used in the subsequent chapters of this study (cartographic, demographic and climatic) have also been described.

Chapter 3

Dengue Fever and Dengue Fever in Malaysia

This chapter describes the aetiology and transmission of dengue fever and discusses the distribution of the disease and recent trends in that both globally and in Malaysia. The sources of dengue data for Malaysia used in subsequent chapters of the study are then described. In order to provide the context for the exploratory analysis and modelling in later chapters, the chapter then goes on to discuss the potential impact of climatic factors such as temperature and rainfall on the incidence of dengue and reviews the previous research which has been conducted on such relationships both worldwide and, in particular, in Malaysia.

3.1 Dengue Fever and its transmission

The origins of the word dengue are not clear, but one theory is that it came from the Swahili phrase 'Ka-dinga pepo', meaning 'cramp-like seizure caused by an evil spirit'. The Spanish word 'dengue' meaning fastidious or careful could describe the gait of a person suffering the bone pain of dengue fever. The use of the Spanish word may be linked to the similar-sounding Swahili. Slaves in the West Indies who contracted dengue were said to have the posture and gait of a dandy and the disease was known as 'Dandy Fever'. The first recorded case of dengue fever can be found in a Chinese medical encyclopedia from the Jin Dynasty which referred to the water poison associated with flying insects. Dengue epidemics occurred for the first time almost simultaneously in Asia, Africa and North America around 1780s, shortly after the identification and naming of the disease in 1779. However, the first confirmed case reported dates from 1789 by Benjamin Rush, who named the disease 'breakbone fever' because of the symptoms of myalgia and arthralgia. The viral etiology and the transmission by mosquitoes were only discovered in the 20th century.

As discussed briefly in Chapter 1, dengue is caused by infection from one of the four serotypes of dengue virus (DENV) which are known as DENV-1, DENV-2, DENV-3 or DENV-4 (and collectively as *flavivirus*). It is transmitted by the bite of the female *Aedes aegypti* mosquito, which at the same time also spreads the chikungunya and yellow fever viruses (to a lesser extent *flavivirus* can also be transmitted by *Aedes albopictus* and *Aedes polynesiensis* mosquitoes). The *Aedes aegypti* mosquito is small in size measuring around 4 to 7 millimetres and is usually a dark colour with typical white markings on the legs and in the form of a lyre on the thorax. Female mosquitoes are larger than male mosquitoes, and can be differentiated by small palps tipped with silver or white scales. The mosquitoes generally acquire the virus while feeding on the blood of an infected person. The biting activity occurs in the early morning or late afternoon/evening although mosquitoes may feed throughout the day e.g. in darkened interiors and during overcast weather. The mosquitoes become infected by the blood meal from a viraemic person and became infective after an incubation period of 10 days to 12 days. After the mosquito is infected, it may transmit dengue by taking a blood meal or by simply probing the skin of a susceptible person. After the virus has incubated an infected mosquito is capable of transmitting the virus for the rest of its life; but, at the same time, the life expectancy of the adult mosquito also clearly has a considerable infuence on incubation period completion. The virus circulates in the blood of infected humans for another 2 to 7 days, at approximately the same time they have a fever. *Aedes* mosquitoes may acquire the virus when they feed on an individual during this period. Hence, infected humans are the main carriers and multipliers of the virus, and serve as a source of the virus for uninfected mosquitoes.

Repeated infections with different serotypes of dengue can lead to the serious complication referred to as dengue hemorrhagic fever which can prove fatal. Early symptoms of dengue hemorrhagic fever are similar to those of dengue, but after several days the patient become irritable, restless and sweaty. Monath (1994) presented evidence that the risk of sequential infections and consequently an incidence of dengue hemorrhagic fever has risen progressively starting with Asian areas and continuing to the Americas. Ibrahim et al. (2007) revealed the severity of dengue risks could be based on three criteria from blood samples; platelet count (PLT), haematocrit (HCT) and either aspartate aminotransferase (AST) level or alanine aminotransferase (ALT) level. Dengue severity has also been related to the two factors of obesity and dengue virus type II (DENV-2) (Natchaporn et al., 2006).

Aedes mosquitoes prefer to breed in water-filled receptacles, close to human habitation. Gubler and Rosen (1976) found that *Aedes aegypti* larvae thrive in artificial containers that contain water, such as in discarded tyres, buckets, paddling pools and blocked rain gutters. Strickman and Kittayapong (2003) chose Chachoengsao Province, Thailand to count all containers in 10 houses per month which contain mosquito larvae and pupae. They measured the wings of female *Aedes aegypti* and the number of pupae with size of emerging females in these containers. Because *Aedes aegypti* are container-breeders, container management is one of the best approaches to reducing their breeding places. Unused containers should be eliminated and containers which remain open due to frequent usage should be subjected to proper larviciding treatment to prevent *Aedes* from laying their eggs.

Control of dengue transmission and its incidence worldwide is a complex question involving effective surveillance, emergency response, mosquito control and effective use of both vaccines and anti-viral drugs when, and if, they become available. Many of the issues involved are similar to those for any infectious disease and stress the importance of implementing effective intervention and early identification of epidemics to control the disease and reduce morbidity (Khun et al., 2005). However, there are also specific issues associated with dengue. Currently, dengue vaccine development is complicated because to incorporate all four virus sero-types into a single formulation and get approval of such a vaccine will require time — there is currently no 'magic bullet'. Hence, the key to preventing dengue transmission now is reduction of the population of its principal vector *Aedes aegypti* (Ooi and Gubler, 2008). Dengue vector control relies mostly on how larval populations of the mosquito vector are managed (eliminating container habitats or using insecticides). Examples of such schemes include: the attempted 'eradication' of *Aedes aegypti* in Brazil during the 1930s which followed a highly organised programme of surveillance and larval control; also a variety of larval control programs with more modest goals since which have resulted in some reduction of dengue transmission in Australia, Indonesia, Thailand and Brazil (Tren and Bate, 2001; Strickman and Kittayapong, 2003; Kusriastuti and Sutomo, 2005).

Carbajo et al. (2001) used the past history of dengue spread in Argentina to produce a risk map of dengue in order to assist in planning prevention strategies and gain a better understanding about the transmission dynamics in areas which are at the southern geographical distribution limit of the vector. They considered four factors in building the thematic maps which were population density, the entrance of the virus, the conditions of the vector and extrinsic incubation period. The results concluded that the maximum risk of dengue transmission is in the northern and north-eastern part of Argentina year-round and in the central regions during the summer. Similar risk map studies have been carried out in other parts of the world. Monath (1994) is a strong reference for more discussion about dengue, the virus introduction and spreading methods.

3.2 Dengue Fever worldwide

Dengue has primarily emerged as a major world health problem because of changes in human demography and behaviour, as well as unchecked populations of, and increased exposure to, *Aedes aegypti* mosquitoes which spread the virus. Later virus-specific serotype factors then influence the epidemiology of dengue. According to Gubler (2002), dengue fever became distributed worldwide in the tropics during the 18th and 19th centuries when the shipping industry and commerce were expanding. At that time, the principal mosquito vector, *Aedes aegypti*, and the viruses responsible for dengue fever were spread via sailing ships because the mosquitoes used the stored water on the ships as a breeding site and maintained the transmission cycle. Both the mosquito and the virus were introduced when such a ship called at a port but because of the slow mode of transportation, epidemics were infrequent, with intervals of 10 to 40 years.

The frequency of dengue fever epidemics has steadily increased worldwide in recent years and endemic transmission has been established over a geographically expanding range of places. Monath (1994) has discussed the specific countries or places that dengue appear most at risk and the reasons for this. Countries or areas where dengue incidence or risk of dengue has been reported can also be viewed on the website of International Travel and Health Interactive Map[1] prepared by WHO. Figure 3.1 outlines the key areas at risk in 2010.

Gubler (1998) highlighted that about 2.5 billion people or 40% of the worlds population live in areas with high risk of dengue transmission. Dengue fever has become a reemergent disease endemic to most of the tropical and sub-tropical regions of the world, with frequent and cyclical epidemics. Nowadays, dengue has spread to more than 100 countries in Asia, the Pacific, the Americas, Africa and the Caribbean. The most important factor recognised is unplanned urbanisation which is believed to have the largest impact on disease amplification for individual countries, whereas travel is believed to have the largest impact on global spread. Community knowl-

[1] http://apps.who.int/ithmap/

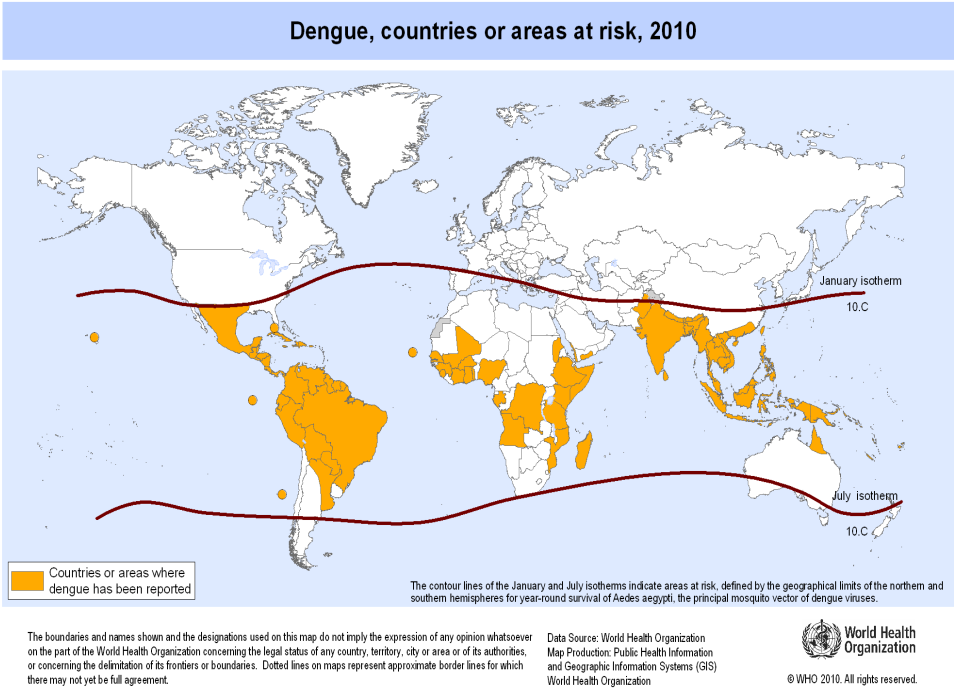

Figure 3.1: *Countries at risk of dengue in 2010.*

edge and practice concerning the disease is also important. COMBI (Communication for Behavioral Impact) is a strategic approach to control diseases all over the world. Reports from India, Kenya and Bangladesh evidence that COMBI had been used to control Tuberculosis (TB), Lymphatic filiariass in Zanzibar, Nepal and Sri Lanka and leprosy in Mozambique (WHO, 2012a). COMBI has also been effective in controlling dengue for example in improving environmental sanitation. A study in the sub-urban residential areas of Taman Desa Kolej, Nilai, Negeri Sembilan, Malaysia, Rozita et al. (2013) concluded that COMBI activities were really effective in helping control dengue cases and produced significant change to opinion, knowledge and practices about dengue among the residents. Parks et al. (2004) wrote a report about the importance of social mobilisation and communication to sustainable dengue prevention and control emphasising key features obtained from 12 national case studies of dengue-related social mobilisation and communication initiatives. Most of the case studies were originally commissioned to illustrate key points in a WHO guide on planning social mobilisation and communication for dengue prevention and control. The report expressed the hope that social mobilisation and communication would continue alongside improvement of public health infrastructure, epidemiological and entomological surveillance, effective clinical management and emergency preparedness. Mosquito control is known as the only good alternative to prevent dengue epidemic influenced by human behaviour and climatic conditions. To maintain this situation, constant effort is needed to be combined with some expensive methods of control. However, the successes of prevention and control are very rarely reported because of the continuous reintroduction of virus or vector from outside and sometimes because of the growing resistance of mosquito populations to insecticides.

Dengue hemorrhagic fever has also steadily increased. For example, the first outbreak of dengue hemorrhagic fever in Singapore occurred in the 1960s, and since then this epidemic has recurred annually and become a primary disease for the urban human population i.e. in areas of the highest population density (Chan et al., 1971). The lack of existing theoretical models involving social and demographic factors encouraged Hales et al. (2002) to develop a model based on vapour pressure (measure of humidity) to assess the geographical limits of dengue fever transmission. They found that the current geographical limits of dengue fever transmission can be modelled with 89% accuracy on the basis of long-term average vapour pressure. They also estimated that future climate change could expose some 5 to 6 billion people to risk of dengue transmission in the longer term.

Changes in the global epidemiology of dengue fever have been observed in recent years for North and South America as well as in the Pacific region and in Southeast Asia. This may be due to climatic changes and to the failure in controlling the mosquito vector but changes in social factors are also important. For example in the 1940s large dengue outbreaks were documented in the United States reaching places as far north as Boston, but today, the situation has changed significantly and outbreaks are rare despite the fact that suitable climate and mosquito vectors and susceptible human hosts are all still present in the continental United States and dengue viruses are frequently reintroduced by infected travellers. Studies on the US-Mexico border have suggested that the restriction of transmission is due to the limitation of contact between human hosts and mosquito vectors that comes with low housing density and the use of air conditioning and screens[2]. There has been an increase in the number of travellers worldwide, including to tropical areas. Schwartz (2002) report that in the period 1995 to 2002, 149 cases among Israeli travellers were acquired in Thailand mostly from four locations; Ko-Phangan, Ko-Samui, Ko-Tao and Ko-Phi Phi. In South America, Lowe (2010) have reported on trends in monthly dengue counts and epidemic cycles in micro-regions of Brazil in the period 2001-2009. Successive epidemics of dengue have been occurring in Brazil since 1986 and almost three million cases of dengue fever and 2,229 cases of dengue hemorrhagic fever had already been recorded by 2002. The introduction of the three serotypes in circulation (DENV-1, DENV-2 and DENV-3) has always started in Rio de Janeiro. Approximately 47,370 and 89,394 cases of dengue due to DENV-1 were recorded in 1986 and 1987 respectively, corresponding to a risk rate of 34.5 and 64.63 cases per 100,000 population. The two following years were characterised by low occurrence of dengue fever. However, the introduction of DENV-2 in 1990 was also followed by an epidemic reaching close to the magnitude of previous epidemics (27.29 and 71.1 cases per 100,000 population in 1991 and 1992 respectively). From 1994 onwards, the transmission rapidly progressed to many Brazilian cities and this wave of epidemics remained constant for four consecutive years, reaching a peak in 1998 (326.4 cases per 100,000 population). It is very clear that the decline of this latest epidemic did not attain the inter-epidemic levels of the two previous waves, when the risk varied from 1.13 cases per 100,000 population in 1988 to 4.87 cases per 100,000 population in 1993, as the rate always remained greater than 127 cases per 100,000 population. The fourth wave began in 2001, shortly after the DENV-3 was detected, and was characterised by increased rates of both dengue fever and dengue hemorrhagic fever, considerably higher than the total accumulated over the entire previous decade (Teixeira et al., 2006). Aura and Alfonso (2010) has reported on similar trends for one region of Venezuela for

[2] http://www.cdc.gov/dengue/entomologyEcology/climate.html

the period 2001 to year 2008. In other work in South America based in Brasilia, Belem, Fortaleza and Boa Vista, Favier et al. (2006) have investigated methods for determining the reproductive number for dengue early in any epidemic taking into account incubations both in the vectors and in the host. The results indicated higher estimates of the reproductive number than that suggested in previous work.

Turning to Australia, dengue transmission is currently restricted to the Queensland area, where *Aedes* is established. The most likely factor influencing the distribution of dengue in that country is the increase of dengue activity in the Asian and Pacific areas which could increase the rates of virus importation especially by travellers. Interestingly, Russell et al. (2009) cites Australia in stressing the need for projections of future dengue spread to consider carefully local historical, cultural and demographic data. They quote evidence that the dengue vector and viruses arrived in Australia before European settlement with visitors to Northern Australia from Malaysia and Indonesia, however the semi-nomadic way of life of the indigenous population at that time was not conducive to the establishment of *Aedes aegypti* and the disease.

In an Indian context, Singh et al. (2005) reported on clinical and laboratory data for 185 cases of both dengue fever and dengue hemorrhagic fever collected from Lok Nayak Hospital of New Delhi in 2003. After analysis, 2.7% of the mortality rate was recorded in an outbreak that started in September, reached a high peak in the next two following months and lasted until December of the same year. Suggestions were made to strengthen the vector control measures including disposal of water containers and improving sanitation.

In Southeast Asia generally one of most current issues in surveillance for dengue is a lack of uniformity in the case definitions used (Ooi and Gubler, 2008). The latter reported that different Asian countries classify dengue fever differently and hence there are variations in the numbers of dengue cases that are included in surveillance reports amongst countries adopting different criteria for classifying dengue cases. They also reported that a confusion between dengue fever and dengue hemorrhagic fever existed sometimes in dengue fever surveillance. Results by Chuang et al. (2010) showed that the distribution of the onset-to-confirmation time for the positive cases was also different. The same paper proposed a dynamic statistical model to estimate the daily number of new cases and the daily cumulative number of infected cases and demonstrated that the daily new cases and cumulative epidemic curves estimated by the proposed method have a lower bias than the values estimated solely based on the available daily-confirmed cases. During years of normal transmission in Thailand cases are seen in the rainy season which is from July to November, but during outbreak years in 1998 and 2002, maximum cases occurred during the dry season from December to June. Mammen et al. (2008) examined data on dengue infection and mosquito density within Thai villages to determine the spatial and temporal dimensions of dengue transmission. Results showed significant spatial and geographical clustering in dengue transmission within a study area which was a rural area of Thailand where dengue was hyper-endemic. A thirteen month study with a total of 271 samples from patients suspected of having dengue infections were selected from clinics and hospital in Brunei. Brunei is located on Borneo Island bordered by the Malaysian States of Sarawak to the west and Sabah to the east. Through three phases of testing procedures, 45 people suspected positive for dengue-specific were investigated and overall the predominant infected serotype was DENV-2 followed by DENV-1 (Osman et al., 2007).

In Cambodia, the continuing contribution of dengue fever to the hospitalisation and deaths in hospitals of infants and small children has been associated with delays in presentation for medical attention, diagnosis and appropriate care. It is important to identify the reasons that influence these delays, in order to develop appropriate interventions to re-address the impact of dengue. Sokrin and Lenore (2007) used ethnographic data which was collected in two villages in the eastern province of Kampong Cham, Cambodia in 2004. Interviews were conducted with mothers whose children had been infected with suspected dengue fever, or who had been sick for other reasons, in 2003 and 2004. The results concluded that women selected a therapeutic option based on perceptions of the severity of the child's condition, confidence in the particular modality, service or practitioner, and affordability of the therapy. While they knew what type of health care was required, poverty in combination with limited availability and perceptions of the poor quality of care at village health centers and public referral hospitals deterred them from doing so. Women initially used home remedies, then sought advice from public and private providers, shifting from one sector to another in a pragmatic response to the child's illness. The lack of availability of the financial resources for poor people and their continuing lack of confidence in the care provided by government centre resulted in a combination of a delay in help seeking and inappropriate treatment of the child's illness.

Elsewhere in the world, Ashford et al. (2003) reported the first outbreak of dengue fever with DENV-4 virus which occurred between January and June 1995 in Palau, an island nation with 32,000 inhabitants in the Western Pacific. They established active surveillance at the national hospital and private clinics, reviewed available clinical records and conducted serologic and entomologic surveys to determine the magnitude of the outbreak and risk factors to guide control strategies in that country. Over the duration of study, they found 817 patients with acute febrile illness with body or joint aches and one of the following signs: either headache, rash, nausea, vomiting or hemorrhagic manifestations, presented to health facilities in Palau. Potential vectors included the introduced mosquito species *Aedes aegypti* and *Aedes albopictus* as well as the native species *Aedes hensilli*. A public education campaign, improved solid waste disposal, continued monitoring of febrile illness, early detection and diagnosis of potential dengue fever outbreaks and programmes of mosquito control were suggested in order to decrease dengue outbreak.

In the Middle East, Khormi and Kumar (2011) conducted a study in Jeddah County in Saudi Arabia to model areas at risk of dengue fever, based on the spatial relationship between dengue fever cases and different socio-economic parameters. High resolution satellite images were used to classify neighbourhoods based on width of streets, roof area of house and density of houses. Geographically Weighted Regression (GWR) was then used to relate dengue cases to neighbourhood classification, population size and density and other socio-economic factors. Strong positive associations were found between dengue cases and some of these factors e.g. overall prevalence among Saudis was higher than non-Saudis especially and there are significant differences in age groups for adults between the ages of 16 and 60 years. In another paper relating to the same place, a study by Khormi et al. (2011) used GIS with the aim of improving the monitoring and surveillance of the *Aedes* vector. Five years of data were used to produce spatio-temporally maps of dengue risk. Monthly hotspots were mainly concentrated in central Jeddah districts but the pattern was found to change considerably with time. The paper proposed following the monthly dengue fever pattern to facilitate the allocation of resources for the treatment of the disease, preventing its prevalence and monitoring its vector.

On the general front worldwide, mosquito control is known as the single most effective intervention to prevent dengue epidemics. To maintain this situation, constant effort is needed. However, the successes of prevention and control are very rarely reported because of the continuous reintroduction of virus or vector from outside and sometimes because of the growing resistance of mosquito populations to insecticides. At the same time, climate variability and global warming are other factors which may favour epidemics of dengue. In one effort in a Claris EC project (Degallier et al., 2010) developed a model for the transmission of dengue to serve as a tool for estimating the risk of epidemics under different climatic change scenarios so that it could be used as an early warning system with meteorological forecasts as inputs — a topic which is taken up in more detail in subsequent sections of this chapter.

3.3 Dengue Fever in Malaysia

Most early cases of dengue fever that were recorded in Malaysia came from Penang; however, the first nationwide outbreak started in Kuala Lumpur. Since then, dengue has become a major public health problem in Malaysia. Shekhar and Huat

(1992) reported that dengue has been endemic in Malaysia since the 1960's and a major issue from 1973 onwards. Upward trends in dengue incidence from 1988 are reported by Narwani et al. (2005). Most recently the states with the highest DIR (per 100,000 population) were Selangor, Kuala Lumpur, Penang, Perak and Negeri Sembilan (Mohd Ismail et al., 2007, 2009). Most of the increase in dengue morbidity and mortality has proceeded in parallel with the rapid economic development, expansion of urban areas and corresponding increases in population density in different locations. Dengue is one of the most common mosquito-borne diseases in Malaysia, so Ministry of Health Malaysia maintains a current dengue report on its homepage[3] which is updated weekly. The next subsection discusses the distribution of dengue cases across Malaysia, subsequent subsections consider surveillance and control of dengue in Malaysia and the sources of dengue data used in later chapters of this study.

3.3.1 Distribution of dengue in Malaysia

The earliest reported case of dengue fever in Malaysia occurred in 1902 when it reached Penang from Singapore and was identified as being the DENV-1 serotype (Skae, 1902). The first case confirmed as dengue-cased haemorrhagic fever was reported in 1962 in Georgetown City, also in Penang.

The first wider scale reports of the disease in Malaysia were prepared by Rudnick et al. (1965), but the disease was observed only sporadically until 1973 where the first major outbreak of dengue fever and dengue hemorrhagic fever occurred in Malaysia with a total of 1,487 cases. Of these, 969 cases were dengue hemorrhagic fever with a fatality rate of 5.6 cases per 100,000 population. The main epidemic focus was in Johor and DENV-3 was identified as the prevalent serotype. Another major epidemic, again focussed in Johor, occurred in the subsequent year with a total of 2,200 cases and 104 deaths reported. The next major outbreak was in 1982 with 3,005 cases notified of which 28.4% were cases of dengue hemorrhagic fever.

[3]http://www.moh.gov.my/

There were 35 reported deaths. The majority of these cases were in young adults and children, and mortality was often the result of multiple-organ failure, a typical feature of dengue hemorrhagic fever. In the next decade, a fairly low incidence rate was reported with an average of about 500-900 cases of dengue fever and dengue hemorrhagic fever reported each year with peaks in 1987 and 1989. A study by Lam (1993b) found that during the decade of 1973 to 1982 there were a total of 12,077 dengue cases with a case fatality rate of 3.38%. In the following decade of 1983 to 1992, the number of reported cases increased to 26,361 but the case fatality rate dropped to 0.55%. This increase was attributed to the rising economy, rapid industrialisation and urban migration at that time, with the reduction in the fatality rate thought to be due to better response of patients seeking early medical treatment as well as better case management.

The Department of Medical Microbiology at the University of Malaya was designated as a WHO Reference Center for Dengue Fever and Dengue Hemorrhagic Fever in 1982 and since then countrywide epidemiological surveillance for dengue has been conducted in close collaboration with the Malaysian Ministry of Health. By 1993, Poovaneswari (1993) reported dengue to be endemic in Malaysia, especially in the major towns with an overall median incidence for dengue fever and dengue hemorrhagic fever of 27.49 cases per 100,000 population. All of the 12 states in Malaysia were reported as being affected; although the majority of cases were confined to the highly populated states such as Kuala Lumpur, Johor and Penang. Subsequently the number of dengue fever and dengue hemorrhagic fever cases in Malaysia has continued to increase with the dengue incidence rate increasing four times in 8 years from 27.5 cases per 100,000 population in 1990 to 123.4 cases per 100,000 population in 1998 when the highest peak yet seen was recorded. In that year, 27,381 cases were recorded nationally (Abu Bakar and Shafee, 2002), the outbreak affected the whole country but was particularly serious in Terengganu State — 1,907 confirmed dengue fever cases and 153 confirmed dengue hemorrhagic fever cases were recorded by the Vector Borne Diseases Control Unit, Terengganu State Health Department. They found the dengue outbreak peaked in Terengganu

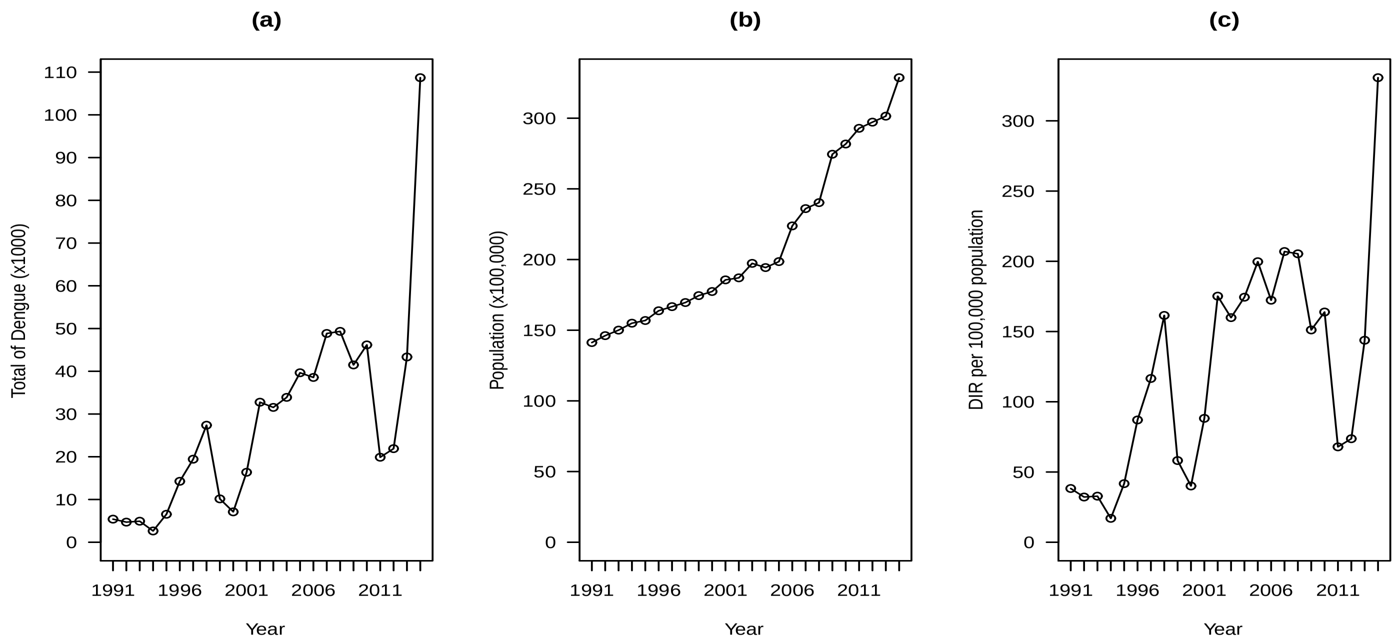

Figure 3.2: *(a)The total number of dengue cases, (b) The total population size and (c) DIR per 100,00 population in Malaysia from 1991 to 2014 (Source : Ministry of Health Malaysia, 2014).*

around August to October 1998 and declined in February 1999. Nor Azimi (2000) has reported that the national case fatality rate was 0.3% in 1998.

The high rate in 1998 fell for the year 2000 to 31.99 cases per 100,000 population but has since risen again and nowadays, dengue is a leading cause of severe illness and hospitalisation in Malaysia and a serious public health issue. According to the 2008 'Health Facts' issued by the Malaysian Ministry of Health, the incidence rate of dengue and DHF was 167.76 and 10.16 cases per 100 000 population with a mortality rate of 0.02 and 0.38 respectively (Ministry, 2008). A study by Mia et al. (2013) found that during the period 2000 to 2010, the number of dengue cases and number of deaths increased on average by 14% and 8% per year respectively. The proportion of the Malaysian adult population exposed to the dengue virus was examined in research by Muhammad Azami et al. (2011). The results indicated 916 or 91.60% positive for dengue out of 1,000 people, with 541 female and 375 male dengue seropositive. The conclusions were that there were similar sero-prevalence rates between urban and rural samples, implying dengue has spread beyond the urban areas in Malaysia and is now confirmed as being endemic across the whole of the country. The most abundant strain of the disease in the country has changed over time, for instance, DENV-2 was the predominant strain in 1989-1991 (Chee and Abu Bakar, 2003) and was replaced by DENV-3 from 1991 onwards (George, 1992). Most recently, the Malaysian Ministry of Health (Ministry, 2014) reports the total number of dengue cases Malaysia in 2014 as the highest ever, standing at 108,698 (see Figure 3.2). Interesting, whilst the number of dengue cases in Malaysia almost tripled from 2013 to 2014, neighbouring Thailand and the Philippines saw decreases in total dengue in that period.

Examination of the historical outbreak data suggests that a major dengue fever and dengue hemorrhagic fever outbreak occurs in Malaysia in an irregular cyclical pattern of every few years. Little is known about the reasons for such cycles. More is known about the cycle within any particular year which corresponds to relatively lower levels during the period from January to April when incidence begins to rise reaching a peak in July or August and then declining. This annual cycle may well relate to the monsoon seasons discussed in the previous chapter i.e. the Northeast monsoon and Southwest monsoon. Storage of water during the dry season from January to April and the drizzling rainfall before the heavy monsoons arriving would create suitable breeding places for the disease vector.

Various studies have looked into risk factors and clinical outcomes for dengue in Malaysia. The lack of much previous discussion about the application of serological techniques to dengue diagnose, encouraged Smith (1957) to perform an early study looking at clinical and epidemiological considerations related to dengue outbreak. Much more recently, Seng et al. (2005) went through a geostatistical modelling, analysis and mapping approach in Johor state to understand better the correlation between dengue fever prevalence, population distribution and meteorological factors, also the characteristics of the space-time clusters. By mapping the spatial variation of dengue incidence using geostatistical analysis and space-time scan statistics, they found a strong positive spatial association between dengue fever prevalence and population distribution. However, the assumption that dengue prevalence must be higher when population density is higher was contradicted by their results possibly due to the positive impact noted in an earlier paper by Chee and Abu Bakar (2003) from dengue control and prevention programmes in high population density areas. Seng et al. (2005) also concluded that accumulation of rainfall over 10 to 14 days is quite enough to support the mosquito's breeding cycle and the dengue virus incubation period. Narwani et al. (2005) looked at the relationship between dengue serology results and observed symptoms and socio-demographic and clinical variables. The study aimed to determine the characteristics of dengue fever and dengue hemorrhagic fever in Malaysia in the years between 1998 and 2003. The results showed significant differences in dengue fever and dengue hemorrhagic fever incidence between age groups and also relationships with systolic blood pressure. Associations were also found between the type of dengue and geographical area; whilst for symptoms the only associations found with any of the factors studied was in severity of joint pain. Mohamad Ismail et al. (2011) performed a small retrospective study from year 2000 to 2004 to look at the dengue infection among pregnant women. They focused on maternal presentation, complications of patient and fetus and pregnancy outcomes. Out of 16 cases, three cases died from dengue shock syndrome, one case dropped out because of an abortion and four cases disappeared during the study period. Amongst the remaining 8 cases, four babies arrived by premature birth, three babies went to intensive care unit and another one was recorded as an early neonatal death. Pouliot et al. (2010) reviewed 30 published studies in assessing the impact of dengue infection during pregnancy on birth outcomes. The results indicated an increase in cesarean deliveries and pre-eclampsia in women who have dengue infection during pregnancy. An increase in low birth weight among infants born compared to non-infected women was also found in the study. Aziz et al. (2011) carried out an investigation to identify the contribution of a range of environmental parameters to dengue outbreaks. They found ten environmental parameters which influence the dengue transmission and distribution such as housing types, land surface temperature, elevation, soil moisture, humidity, rainfall, temperature, population density, greenness and land use. There have been several other studies looking at how climatic variability may impact epidemics of dengue and these are picked up in a later section of this chapter.

3.3.2 Surveillance and control

The Malaysian government has identified dengue control as a national priority. Until now, there is still no effective vaccine or specific treatment for dengue fever and current control methods (e.g; larviciding, space spraying insecticides or 'fogging', public education, legally enforced breeding site reduction) have not stopped the spread of the disease, so there is an urgent need to evaluate promising new technologies in Malaysia at the moment.

The primary goal of public health surveillance in relation to dengue is to monitor transmission of the disease in the community so as to guide an effective program to prevent future occurrence and spread. It is also to identify the cost-effectiveness of public health prevention programmes and estimate the burden of dengue for the community. Such efforts should help authorities to monitor dengue cases more accurately and hopefully gain early predictions of epidemics using established risk factors. As in most of the tropical countries around the world, dengue places a heavy burden on public health systems in Malaysia given that surveillance and emergency response is inevitably constrained by lack of the infrastructure and functional support systems.

As discussed earlier, there are currently no specific medications to treat dengue and there is no vaccine commercially available against dengue today. Prevention is therefore the only step to reduce the risk of dengue infection in Malaysia. Several approaches are available; either mosquito control (larval control or adult mosquito control) or by reducing mosquito biting especially during daylight hours. Many studies have considered control of dengue outbreaks in Malaysia such as anti-larval and anti-adult measures, health education and legal enforcement, but dengue still continues to be major problem. Environmental factors such as rainfall, temperature, living conditions, domestic waste management and population distribution and demographic structure are important in identifying the mosquito survival and reproduction.

Lam (1993a) considered requirements to reduce the incidence rate of dengue to less than 6 cases per 100,000 population and the case fatality rate to 0.04%. Yap et al. (1994) discussed necessary future planning in terms of vector control approaches such as source reduction, environmental management, larviciding and adulticiding. Poovaneswari (1993) discussed methods in prevention and control of dengue in Malaysia such as anti-larval activities (house inspection, the use of larvicide, Enforcement of the 'Destruction of Disease-Bearing Insects Act' of 1975, anti adult activities fogging), health education activities and community participation. Teng and Singh (2001) identified there are four major sources of *Aedes* breeding including construction sites, solid-waste dumps, open spaces and factories. Teng and Singh (2001) reported on ways in which surveillance methods have been upgraded to strengthen dengue control in Malaysia. These included reprioritising *Aedes* surveillance aimed at new breeding sites, strengthening information system for effective disease surveillance and response, changing insecticide fogging formulation, legislative changes for heavier penalties and strengthening community participation and collaboration as well as clinical efforts to reduce case fatality. Meanwhile, the Ministry of Health Malaysia has clarified favourable areas for mosquito breeding such as; in construction sites, rubbish dumping sites, parks, vacant land, cemeteries and public infrastructure areas.

Egg traps are a safe, economical and environmentally friendly method for the surveillance of mosquito populations first introduced in the United States. Their sensitivity is such that they can efficiently estimate the population of *Aedes* even when that is low. A decade ago, egg traps become a popular tool to catch mosquitoes in Malaysia. Dhang et al. (2005) conducted a study to determine the distribution and abundance of *Aedes aegypti* and *Aedes albopictus* in two urban residential areas and settlement areas in Selangor state. They found both mosquitoes appeared in both places, but *Aedes aegypti* was definitely the most found dengue vector. Chen et al. (2006) followed this study up by using egg traps in the same state as previously but with the addition of Kuala Lumpur as a new location. The results were similar and reinforced the conclusions that *Aedes aegypti* was found at a higher frequency than *Aedes albopictus*. The next egg trap implementation reported in surveillance of the dengue vector (Rozilawati et al., 2007) was for two selected urban and sub-urban areas in Penang state for a 14 months period. The main dengue vector found was *Aedes albopictus* in this case. Wan Norafikah et al. (2009) used egg traps on the University of Malaya campus in Kuala Lumpur. Results showed a correlation between the mean number of larvae per egg trap of *Aedes albopictus* and rainfall meaning that the most populous dengue vector in the university campus was most likely to be *Aedes albopictus*.

Narwani et al. (2005) reported on dengue in Kota Bharu one of the districts in the state of Kelantan with a sharply increasing trend in dengue cases going from an incidence of 8.5 cases per l00,000 population in 1988 to 88.6 cases per 100,000 population in 2003. Overall there were 4,476 dengue fever cases and 240 dengue hemorrhagic fever cases in these six years. Overall, the monthly peak season of recorded dengue was in January and the lowest recorded dengue was in May. Narwani et al. (2005) also went on to look at requirements of an effective prevention and control programme to provide early warning of dengue epidemics in the state. They concluded that virologic surveillance should be considered an important element in any such system. Dengue virus transmission must be monitored to identify which serotypes are present, their distribution and the type of illnesses associated with each. Juni et al. (2015) in a study of the outbreak of dengue in the rural population of Negeri Sembilan has recently reported on the risk behaviours which need to be addressed in programmes for the control and prevention of dengue.

As regards education programmes more generally, the Ministry of Health in Malaysia has applied variety of mass media interventions and community-based actions to prevent and control dengue fever in the past with mixed results. Of particular note is a study conducted by Suhaili et al. (2004) to plan and implement social mobilisation (also known as Communication-for-Behavioural-Impact or COMBI as mentioned earlier in this chapter) in Johor Bahru District of Johor state in 2001. Johor state is located at the Southern end of Peninsular Malaysia and shares a common boundary with Pahang and Melaka to the North. The state's capital city is Johor Bahru, the second largest city in Malaysia after Kuala Lumpur. This study was supported by WHO in recognition of the difficulties faced by Malaysia in prevention and control of dengue fever. The approach produced positive behavioural results and because of that the COMBI methodology was adopted as the national method for social mobilisation and communication to help in controlling dengue in Malaysia. A similar kind of study was carried out in 2002 in rural areas of Kuala Kangsar District in the State of Perak. A survey on 200 people was conducted to assess the level of knowledges, attitudes and practices in relation to dengue. Two thirds of those surveyed reported having received information on dengue coming from television and radio and most of them were supportive in trying to control *Aedes* populations. There was a significant association between knowledge of dengue and attitude toward *Aedes* control. However, it was also found that even good knowledge about the disease is not guaranteed to lead to good practice because of extant traditions in community life (Hairi et al., 2003).

In summary, considerable work has been done on prevention and control of dengue in Malaysia, however, it remains a complex and problematic task beset by both simple issues, such as a delay in fogging activities or the presence of an abandoned housing project contributing to potential mosquito breeding sites, or more complex ones such as inadequate public compliance in combative measures. The Ministry of Health in Malaysia continues to strive in supporting strategies to combat dengue, in constant monitoring of their implementation and in emphasis on the importance of future preventive and control activities to move dengue down the scale of the major public health problems in Malaysia.

3.3.3 Sources of dengue data

As discussed in Chapter 2, this study is concerned with dengue incidence in twelve 'states' of Peninsular Malaysia (as defined in 2.2). Research ethics approval for this study was granted by the National Medical Research Register[4], Ministry of Health Malaysia[5]. Monthly numbers of dengue cases used in subsequent chapters of this study are those for the period of nine years from January 2001 to end December 2009 for each of these twelve areas obtained from the Ministry of Health Malaysia[6] (Narwani et al. 2005). The dengue cases referred to are the total of confirmed dengue fever cases and confirmed dengue hemorrhagic fever cases in each of the months concerned. Some more detailed dengue data at district (county) level is also available through the website of the State Health Department of Selangor[7] (available online from 2009 upwards Rosnah et al., 2009).

[4]https://www.nmrr.gov.my/fwbLoginPage.jsp

[5]http://www.moh.gov.my/

[6]http://www.moh.gov.my/

[7]http://www.jknselangor.moh.gov.my/index.php?lang=ms

3.4 Climate and dengue

Climatic conditions and climate change affect people, plants and animals. It is common knowledge that scientists are working towards better understandings of future climate change and how the effects will vary by region and over time and on the observed changes that are already occurring. Such effects amongst many include sea level rise, shrinking glaciers, changes in the range and distribution (biodiversity) of plants and animals, trees blooming earlier, lengthening of growing seasons, ice on rivers and lakes freezing later and breaking up earlier, and thawing of permafrost.

Amongst such considerations it is clear that human health can, in part, be affected directly and indirectly by climatic conditions and climate change e.g. through extreme periods of heat and cold, storms, and through the dynamics of climate-sensitive diseases such as those identified by Inter-Govermental Panel on Climate Change (IPCC) (IPCC, 2014). In order to manage this climate sensitivity more effective working relationships between the health sector and the providers of climate data and information are required. In short, climate is an important variable in managing the overall burden of disease, especially in developing countries where the ability to control climate-sensitive diseases is constrained. To reduce its adverse effects, the public health sector must understand and quantify the specific effects of climate variability and change both on the overall disease burden and on opportunities and effectiveness in the public health response. This applies to future adaptation strategies to understand fully the impact of the climate on the existing disease burden and current interventions so that the public health sector can use climate information effectively in epidemic early warning systems. This provides new challenges for the health sector which historically has not usually been engaged in climate and environmental monitoring. Acquiring and using this type of information successfully depends on developing partnerships between health practitioners and the gatherers and providers of climate and environmental information such as National Meteorological Services (Rogers et al., 2008).

As suggested in earlier sections, dengue is very much a climate sensitive disease. The virus is transmitted by *Aedes* mosquitoes whose prevalence is highly sensitive to environmental conditions. Climatic factors such as temperature, precipitation and humidity may be critical to mosquito survival, reproduction and development which thus affects mosquito presence and abundance e.g. higher temperatures reduce the time required for the virus to replicate and disseminate in the mosquito.

The next two subsections consider global and Malaysian perspectives in respect of the relationship between climatic factors and dengue.

3.4.1 Global perspective

The global epidemiology and transmission dynamics of certain vector-borne diseases, such as malaria and dengue have changed considerably since the middle of the 20th century. Some of that is undoubtedly in response to global climate change.

Malaria is, of course, generally spread by the *Anopheles* mosquito, rather than the *Aedes* which is the vector in the dengue case; but, nevertheless studies on the relation between climatic factors and malaria do have some relevance for dengue as some similar issues and involved. It is therefore useful to briefly review some of the work in this area on malaria before moving on to that specifically concerned with dengue. Loevinsohn (1994) in a study looking at malaria epidemics in Rwanda found that changes in malaria incidence were associated with temperature and rainfall and had responded to the steady increase in temperature in Rwanda over the period 1961 to 1990 has experienced before. Ebi et al. (2005) developed a model of future climate suitability for stable malaria transmission using geographic distributions of malaria for 16 projections of climate in 2100. Preliminary results showed that the changes in temperature and precipitation could transform the geographic distribution of malaria in Zimbabwe. The highlands becoming more suitable for transmission, while the low veld and areas with low precipitation showed varying degrees of change, depending on the climate sensitivity and greenhouse gas emission stabilisation scenarios incorporated in the general circulation model used. A study by Martens et al. (1995) looked at mathematical modelling of the effect of anthropogenic global climate change and associated projections of long term changes in temperature and precipitation on mosquito and parasite characteristics and their potential impact on malaria risk. The simulation results indicated a widespread increase in transmission potential of the malaria mosquito population and an extension of the areas conducive to malaria transmission. Scenarios involving a global mean temperature increase of several degrees in the year 2100 increased the epidemic potential of the mosquito population in tropical regions two-fold and more than 100-fold in temperate zones. Other work (e.g. Ghebreyesus et al., 2009) has looked at how seasonal forecasts of temperature and rainfall, which are useful indicators of the likely occurrence of malaria outbreaks, can be used to implement a programme of heightened epidemic surveillance; while real-time temperature and rainfall estimates could be used to start selective interventions and to support early detection of disease outbreaks.

Turning to work specifically related to dengue, the effects of climate change on the distribution of that disease have become an area of increasing research interest over recent decades due to the significant increase in the global incidence of dengue (Gubler, 2002). As regards the global picture, Patz et al. (1998) looked at climate data from 1931 to 1980 to investigate the potential added risk posed by global climate change on dengue transmission. They performed simulations linking temperature output from three climate general circulation models (GCMs) to relationships concerning dengue vector potential for transmission. The three models predicted an average projected temperature elevation of 1.16 °C in the year 2050 and the simulations indicated that under such conditions even reduced numbers of mosquitoes could maintain the same level of endemicity of the disease in areas where the dengue virus was present leading to increased epidemic potential even if control programmes on *Aedes* populations had some success.

Hopp and Foley (2003) also attempted to look at climatic relationships to the dengue vector at the global scale. A numerical model was developed to simulate the response of *Aedes* to climatic variations and to examine how modelled mosquito populations may be related to dengue and dengue hemorrhagic fever cases worldwide. They found that global scale variations in climate could induce large variations in modelled *Aedes* populations which are historically strongly correlated to reported dengue and dengue hemorrhagic fever cases. Data for 240 months between 1978 and 1997 from 73 provinces in Thailand were considered in identifying the effects of climatic factors such as rainfall, temperature and relative humidity on the occurrence of dengue. Degallier et al. (2010) has developed a general model for the transmission potential of dengue and used this to forecast the risk under different climate change scenarios. The paper suggests that such a model might be incoporated into an early warning system with meteorological forecasts as inputs.

Another worldwide issue was highlighted by WHO who reported that the 1997/98 El Niño might have been the cause of the dengue fever epidemics in many tropical countries. This is because of the interaction between the atmosphere and the ocean, the warm El Niño and the cold La Niña phases of the ENSO engender significant temperature and precipitation anomalies around the world. Gagnon et al. (2001) presents the results of a correlation analysis of past ENSO events with dengue epidemics across the Indonesian archipelago and Northern South America. The analysis showed correlation between El Niño and dengue epidemics in French Guiana and Indonesia and to a lesser extent in Colombia and Surinam. These regions experience significantly warmer temperatures and less rainfall during El Niño years. Public health officials could therefore possibly benefit from El Niño forecasts and they should emphasise control activities such as insecticide sprayings and media campaigns concerning the potential breeding sites of dengue mosquitoes during these years.

Turning to more geographically focussed studies, there has been various work on dengue and climate in South America. In Cuba, Puerto Rico and Southeastern Brazil the last decade has seen some increases in rainfall during the dry season. These have contributed to the year round permanence of the mosquito population. Even vegetated watersheds now flood under intensifying rains throughout the dry season, enhancing the risk of outbreaks of dengue throughout the year. Also a greater mosquito abundance at the beginning of the rains leads to higher subsequent growth in the population and hence more likelihood of epidemics. As a result, the South American continent is experiencing an increase in dengue with greater mortality from the more severe hemorrhagic cases. Greater incidence has not only been seen in traditional dengue affected regions, but warmer and moister seasons in neighbouring regions have caused dengue to spread there too. The historical once-a-year prevention campaigns for dengue have needed to be replaced by year-round vigilance. Research by Aura and Alfonso (2010) on the potential associations between climatic variation and dengue cases in Western Venezuela presents analyses based on an 8 year period from 2001 to 2008. The results indicated a significantly higher dengue incidence with lower values of ONI (El Niño periods) and lower dengue incidence with higher values of ONI (La Niña periods). The models are expected to be useful to anticipate and mitigate dengue incidence rate through the implementation of mosquito eradication and determination of the optimum time for fogging activity. Lowe (2010) used generalised linear mixed models (GLMM) to look at the relationship between monthly dengue incidence and a range of climatic and socio-economic variables in microregions of South East Brazil for the period 2001 to 2007. Significant relationships were found with temperature and rainfall lagged by 1-3 months and with ENSO lagged by 6 months. The model estimated on the 2001-2007 data was then tested for predictive validity by using it to make probabilistic predictions for 2008 and comparing these to the observed data in that year. The results showed that an epidemic alert was successfully issued for 94% of the 54 microregions that recorded high dengue incidence rates in South East Brazil during the peak dengue season of February to April 2008. Using data collected in household surveys in 2001 and 2002 in the city of Goiania in central Brazil, Siqueira et al. (2008) used generalised additive models (GAM) to generate smoothed risk maps for dengue adjusted for socio-demographic, climatic and temporal covariates. They found significant spatial heterogeneity in dengue risk across different areas of the same city.

In the Carribean, Gharbil et al. (2011) has looked at the impact of temperature and other climate measures on dengue incidence on the island of Guadalupe using a Seasonal Autoregressive Integrated Moving Average (SARIMA) model applied to data from 2000 to 2006. The model fitted was then used to forecast dengue incidence rate from year 2007 onwards compared to observed data using three different approaches; one year-ahead, three months-ahead and one month-ahead. The three months-ahead approach proved the most suitable forecasting model to adopt for effective operational public health response. Three variables were detected as having positive significant influence; average temperature at lag 11 weeks, relative humidity at lag 7 weeks and minimum temperature at lag 5 weeks. This result makes sense as temperature was believed to influence the dengue outbreaks forecasts more than using humidity and rainfall. Global warming should increase the range of the mosquito and reduce the size of larva and adults. Smaller adults must feed more frequently to develop their eggs. So warmer temperatures boost the incidence of double feeding and thus increase the likelihood of transmission. The time the virus must spend incubating inside the mosquito is shortened at higher temperatures, and this can mean a potential higher transmission rate of disease. The main problem in countering this in Guadalupe is inadequate utility services meaning residents must store water in jars and tanks and these are the preferred breeding grounds for *Aedes*. The increase in air travel is another factor, with infected fliers acting as sources for the virus. Close by on the Island of San Juan in Puerto Rico, Schreiber (2001) has used data for 1988 to 1993 to investigate relationships between dengue incidence and hydrological flow in order to identify and quantify the specific climate conditions and the associated lag periods which determine disease variations. They concluded the mean seasonal variation in dengue has a strong relationship with mean seasonal climate variation, but that drainage is also an important factor. They also assessed the ability of the models developed to predict dengue incidence and managed to develop an early warning model to predict increases dengue incidence with a three week lead time.

Moving to Asia, Lu et al. (2009) investigated the impact of weather variability on the transmission of dengue fever in Guangzhou, China with the aim of proposing early public health responses that would help to minimise morbidity and mortality. Poisson time series models were fitted to monthly notified cases of dengue fever and weather variables for six years from 2001 to 2006. Results showed that the best predictive model for dengue incidence was one that included minimum temperature and minimum humidity both lagged by one month and also wind velocity. Autoregressive integrated moving average (ARIMA) models were used by Wu et al. (2007) to evaluate the impacts of weather variability on the occurrence of dengue fever in a major metropolitan city, Kaohsiung, in Southern Taiwan. This study found somewhat surprisingly that the incidence of dengue fever was negatively associated with monthly temperature and an inverse association was also found with relative humidity.

In Thailand, Thammapalo et al. (2005) developed models for dengue incidence including trend, cyclic effects and climatic factors. Results showed that an increase in temperature was associated with a rise in the incidence of dengue hemorrhagic fever in nine provinces, and an increase in rainfall was associated with a decreased incidence of dengue hemorrhagic fever in seven provinces. The overall picture was that dengue hemorrhagic fever incidence was negatively associated with extra rainfall in the Southern region, but positively associated with elevated temperatures in the Central and Northern regions. Mathuros et al. (2009) used global ENSO records, dengue surveillance data and local meteorological data from two geographically diverse regions in Thailand to assess the temporal relationship between El Niño and the occurrence of dengue epidemics, then constructed Poisson autoregressive models for incidences of dengue cases. The result revealed that at time lag of between 1 and 11 months the strength of El Niño was a significant predictor for occurrences of dengue epidemics.

Althouse et al. (2011) applied linear and generalised linear statistical models to predict incidence of dengue in Bangkok and Singapore based upon climatic and other covariates. Best fitting models for each of the two places differed in some of the covariates selected and also in the size of effects. Loh and Song (2001) looked at clusters of dengue cases in Singapore for the years 2000 and 2001 where a cluster was defined as: at least two cases located within 200 metres of each other with dates of the onset of symptoms within three weeks of each other. They identified 102 clusters and used non-linear regression to relate cluster size with various entomological and climatic covariates. They found significant positive relationships with the detected number of mosquito habitats in the vicinity and the average amount of rainfall one week before the cluster period.

So the overall picture on dengue and climate is a complex one. There is consensus that climatic variations have strong influence on dengue incidence, but more mixed results from different regions of the world on the relative importance of ENSO, rainfall, temperature or humidity and associated time lags. The next section looks at the relatively small amount of research that has been carried out on these issues specifically related to Malaysia.

3.4.2 Malaysia perspective

Although some of the work on climate and dengue described in the previous section in neighbouring countries in South East Asia is pertinent to Malaysia, there are few studies on this issue specifically in Malaysia. Indeed that is exactly one of the main motivations for the work throughout this study.

The relationships between dengue hemorrhagic fever and rainfall were examined by using dengue hemorrhagic cases, precipitation and temperature data from the states of Selangor and Johor during the period of 1973 until 1977. In that investigation, Aiken et al. (1980) found an increase in dengue hemorrhagic fever cases following the March to May wet season and the size of the increase was positively related to the size of the moisture surplus. Besides that, an apparent lack of association between dengue hemorrhagic fever cases and rainfall appeared during the second wet season which is between September and November every year.

Terengganu is one of the states in East Malaysia. This state experiences flooding every year, during the Northeast monsoon which occurs from November to March and brings heavy rainfall. Nor Azimi (2000) described a pattern of dengue in this state which is somewhat different from other states in Malaysia. Before 1997, the dengue situation in Terengganu was under control with annual dengue cases fairly low with the highest number in 1992 at only 214. But, in 1997 and 1998 the dengue situation in Terengganu changed greatly as described earlier in this chapter. A large number of dengue cases have been reported since then. The number of dengue deaths has also increased even though the case fatality rate of 7.14% in 1997 was reduced to 2.25% in 1998.

Wan Fairos et al. (2010) studied relationships between DF and DHF cases and climatic and other variables in Malaysia using daily data for the period July 2006 to December 2008 and found that daily temperature and wind speed significantly influence the incidence of dengue fever after a lag of some 3 weeks, but that humidity has a weaker relationship. Choy et al. (2011) looked at dengue cases collected from the Seremban District Health Office and the Ministry of Health in Malaysian and conducted interviews with 15 key informants or experts on climate change and public health. Relative humidity and rainfall data were obtained from the Meteorological Department and the Department of Irrigation and Drainage. A positive significant relationship was found between mean maximum temperature and relative humidity to the number of dengue cases. For precipitation, the results were more mixed with data from only two out of the four rainfall stations showing a significant relationship to the local dengue cases. In a data mining approach, Abu Bakar et al. (2011) have developed predictive models for dengue outbreak detection using multiple rule-based classifiers based on environmental data.

Mazrura et al. (2010) found positive associations between climate variability and the *Aedes* population in a study carried out in 2009 in the Ledang District of Johor State to assess community vulnerability to dengue and to promote COMBI (as mentioned earlier in this chapter) as a methodology for encouraging community responses in controlling dengue. Trends on *Aedes* population, dengue cases and community surveys at pre and post-interventions, the processes for dengue control activities were analysed. Other similar reports on COMBI activity and programmes can be found in the following papers: Suhaili et al. (2004); Rozhan et al. (2006); Rozita et al. (2013); Azmawati et al. (2013).

3.5 Summary

This chapter has discussed dengue fever and the mechanisms for its transmission and then reviewed the burden of dengue, its geographical distribution and associated trends both in the world in general and in Malaysia in particular. Issues concerned with surveillance and control of the disease were also considered. The sources and extent of the Malaysian dengue data used in subsequent chapters was also explained. The chapter then went on to consider the relationship between climate and dengue globally and in Malaysia and previous associated studies were reviewed.

With this background the next chapter goes on to explore the question of climate, dengue and Malaysia in more depth beginning to focus down on the key research aims of this study as introduced in Chapter 1.

Chapter 4

Exploratory data analysis

This chapter is concerned with preliminary exploratory analyses of the datasets collated for this study so as to inform the statistical modelling of the dengue incidence rate (DIR) in subsequent chapters. The chapter begins by describing the structure of the collated datasets and the variables included and clarifying any associated definitions and provisos. Subsequent sections then proceed to summarise trends and possible relationships in that data, starting with annual trends in DIR and the intra-annual cycle and then moving on to look at relationships between DIR and demographic and climatic factors.

4.1 Description of collated datasets

The sources of data for this study have already been described in Chapter 2 (demographic and meteorological data) and in Chapter 3 (dengue data), for convenience they are also summarised in Table 4.2. Two datasets collated from these sources are used in the remainder of this study.

The first is simply used for background context and not for the main modelling, it comprises annual numbers of dengue cases for the period 1991 to 2009 for each of the 12 states of Peninsular Malaysia as defined in Table 2.2 along with the total population of those states. The second, and main modelling dataset refers to monthly dengue cases during the 108 month period between January 2001 and December 2009 for each of the 12 states of Peninsular Malaysia (reliable monthly figures for all states are not available for earlier years). So the structure of that dataset is therefore a spatio-temporal series of $108 \times 12 = 1296$ records with each record containing the following basic variables; state, latitude and longitude of state capital, land area of state, year, month, total monthly number of confirmed DF and DHF cases, estimated state population pertaining to the year (as projected by the Department of Statistics Malaysia), population density of state, maximum, minimum and average monthly rainfall, maximum, minimum and average monthly temperature, monthly number of rainy days and Niño 4 average sea surface temperature (SST) pertaining to the month. A number of derived variables were then added to each record. First, values of the climatic variables at various preceding monthly lags. Second, the Dengue incidence rate (DIR) where this is defined (Hafiz et al., 2012) as the number of new confirmed cases of DF and DHF, y_{st}, diagnosed in state s $(s = 1, \ldots, 12)$ in month t, $(t = 1, \ldots, 108)$ divided by the total estimated population of the state p_{sj} (in 100,000s) for the year j $(j = 1, \ldots, 9)$ in which month t falls. So the DIR is the monthly incidence per 100,000 persons at risk i.e.

$$DIR = \frac{y_{st}}{p_{sj}} \times 100,000 \tag{4.1}$$

In some of the subsequent discussion in this chapter, annual rather than monthly DIR is used, which is simply the same calculation with the monthly cases replaced by the annual cases for the whole year in question. Third, the region of Malaysia to which the state belongs where this is defined on the basis of sub-divisions of Malaysia used in various previous studies which reflect broad regional differences in demography and climate in the country and where the twelve states are divided into four regions referred to subsequently in this study as 'North East', 'South East', 'North West' and 'South West' (see Table 4.1).

The names used only roughly correspond to the geography they imply. North

Table 4.1: *The division of the states in Malaysia.*

North East	South East	North West	South West
Pahang	N.Sembilan	Perlis	Perak
Terengganu	Melaka	Penang	Selangor
Kelantan	Johor	Kedah	K.Lumpur

East refers to most of the East of Peninsular Malaysia and includes the states of Kelantan, Terengganu and Pahang; the South East region consists of the states of Johor, Melaka and Negeri Sembilan which are actually located more in the South of the Peninsular; North West refers to the Northern part of Peninsular Malaysia which contains Kedah, Penang and Perlis. Finally, the South West region includes the capital of Malaysia, Kuala Lumpur along with Selangor and Perak. The inclusion of a regional grouping of states in Malaysia as a potential impact factor for dengue incidence follows the findings of Johansson et al. (2009b) in Puerto Rico and also those in previous work by Wan Fairos et al. (2010) in Malaysia.

Various caveats and limitations could be raised in relation to the collated datasets described above, but two merit particular comment. First, under-reporting cases of dengue is a potential problem in Malaysia as elsewhere in the world. The problem has decreased in recent years with the introduction of new technologies; however, it remains an issue in less developed countries which still employ the old style of recording dengue data and that includes Malaysia. It has been suggested that to avoid under-reporting cases in modelling dengue, the collated dengue data should come from multiple sources and at different levels such as at the national, regional and state levels (Donald et al., 2012). The data collation should also be broken down by the setting from which the case is reported, classification of severity and the patients age (Yara et al., 2011). Lowe (2010) discusses similar issues in relation to the Brazillian Health System including collecting all details on each patient (basic demographic data, dates of symptom onset, case classification etc.) directly into the computer systems 'on the spot'. Unfortunately, little of this can be assured in the Malaysian context and the potential for under-reporting in the datasets used subsequently in this study needs to acknowledge.

Second the spatial resolution of the data (i.e. the State) is not ideal. It would be useful to obtain dengue cases by district within State for the whole of Malaysia as this would help in explaining and understanding the situation in the specific areas where particular epidemics have been reported, for example in different districts of higher population density areas such as Kuala Lumpur (Mohd Ismail et al., 2009). However, national coverage at the district level is simply not reliably available for anything other than the most recent few years due to complications and inconsistencies in data handling at the district level. In addition, accurate population data at district level is hard to obtain. Hence, for the time period of this study, the State level is the only realistic spatial resolution at which to analyse dengue across the whole of Peninsular Malaysia.

4.2 Patterns in Dengue Incidence

The total population of Malaysia has doubled over the last 30 years rising from 13.7 million in 1980 to 28.3 million at the end of 2009. However, the upward trend has moderated considerably during the nine year primary period of this study from 2001-2009 as indicated in Figure 4.1 which derives from Malaysian Government population estimates and shows only a slight upward trend in population estimates for the four regions of Malaysia during these years. What is also clear from this diagram is the marked difference between the population in the South West region as compared to the other three regions for the 108 months. This demographic pattern needs to be borne in mind as a backdrop when looking at trends in dengue incidence subsequently in this section.

Table 4.2: *Source and original resolution of dataset.*

Data	Spatial resolution	Temporal resolution	Source
Dengue cases	State	Monthly count	NMRR http://www.nmrr.gov.my/
Area	State	None	DSM http://www.statistics.gov.my/portal/index.php
Population	State	Yearly estimate	DSM http://www.statistics.gov.my/portal/index.php
Rainfall	State	Monthly observed	DIDM http://www.water.gov.my/
Number of rainy days	State	Monthly observed	DIDM http://www.water.gov.my/
Temperature	State	Monthly observed	MMD http://www.met.gov.my/
Precipitation	2.5° x 2.5° grid	Monthly mean	ESRL http://www.esrl.noaa.gov/psd/data/gridded/
Temperature	2.5° x 2.5° grid	Monthly mean	ESRL http://www.esrl.noaa.gov/psd/data/gridded/
Humidity	2.5° x 2.5° grid	Monthly mean	ESRL http://www.esrl.noaa.gov/psd/data/gridded/
Oceanic Niño Index	2.5° x 2.5° grid	3-month running mean	CPC http://www.cpc.ncep.noaa.gov/

National Medical Research Register (NMRR), Department of Statistics Malaysia (DSM), Department of Irrigation and Drainage Malaysia (DIDM), Malaysian Meteorological Department (MMD), Earth System Research Laboratory (ESRL), Climate Prediction Center (CPC)

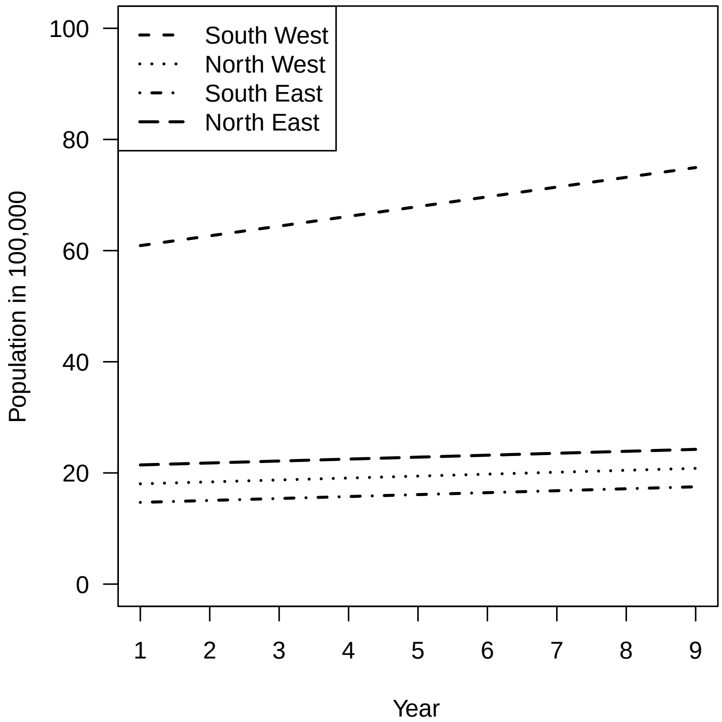

Figure 4.1: *Annual population estimates for the four regions of Malaysia from 2001 to 2009.*

4.2.1 Annual Patterns in Dengue Incidence

Turning more directly to dengue incidence, the annual dataset demonstrates the overall trend (see Figure 4.2) in annual DIR over the nineteen years from 1991 to 2009 across all of the twelve states of Peninsular Malaysia. Also included in this diagram is superimposed loess or localised regression smoother (e.g see Cleveland et al. (1992) and Cleveland and Devlin (1998)). During this period a total of 406,359 confirmed dengue fever cases were recorded, the highest peak of 46,703 cases being in 2008. Clearly, there are peaks and troughs in DIR over this period with a sharp drop in 2000 but the general pattern is one of an increasing trend over the 19 year period as reinforced by the superimposed localised regression (loess) scatter plot smoother. There may be some support for the proposition that epidemics occur roughly every four years (Lam, 1993b) as there were increases in dengue counts in 1998, less so in 2002 and then in 2007 relative to other years.

The different regions of Malaysia are susceptible to different climatic conditions and the geographical characteristics of the regions differ. Hence, it is useful to look at the patterns in DIR within regions (North East, South East, North West, South West) to identify potential variations between these regions. The increasing trend in annual DIR noted earlier is evident in each of the four regions of Malaysia taken separately, as shown in Figure 4.3. However, the value DIR in the South West region is significantly higher compared to the other regions, peaking in the year 2008 at over 350 cases per 100,000 populations compared to the North West,

South East and North East which had DIR values below 200 cases per 100,000 populations in every year. This point is reinforced by superimposing the trends for the four regions on top of each other as in Figure 4.4.

It is evident from this figure is that there are both similarities and differences in the pattern of peaks and troughs in the four regions over these 19 years. For example, the increase in DIR in 1998 and the drop in 2000 is evident in all of the regions, whereas the rise in 2002 in the South West is not so evident in the other regions and

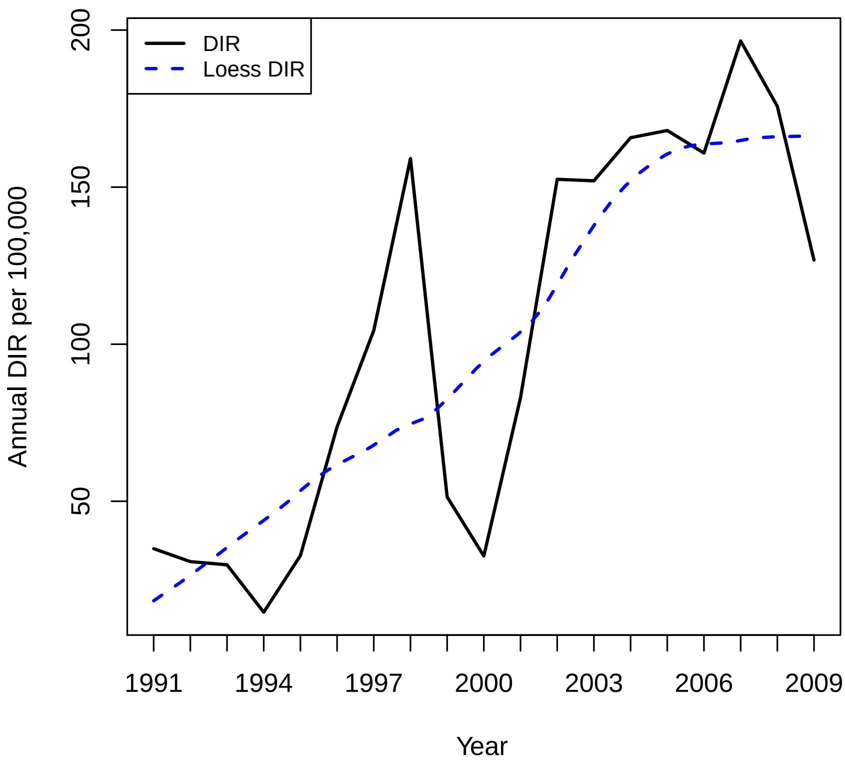

Figure 4.2: *Annual DIR per 100,000 population for the 12 states of Peninsular Malaysia from 1991-2009.*

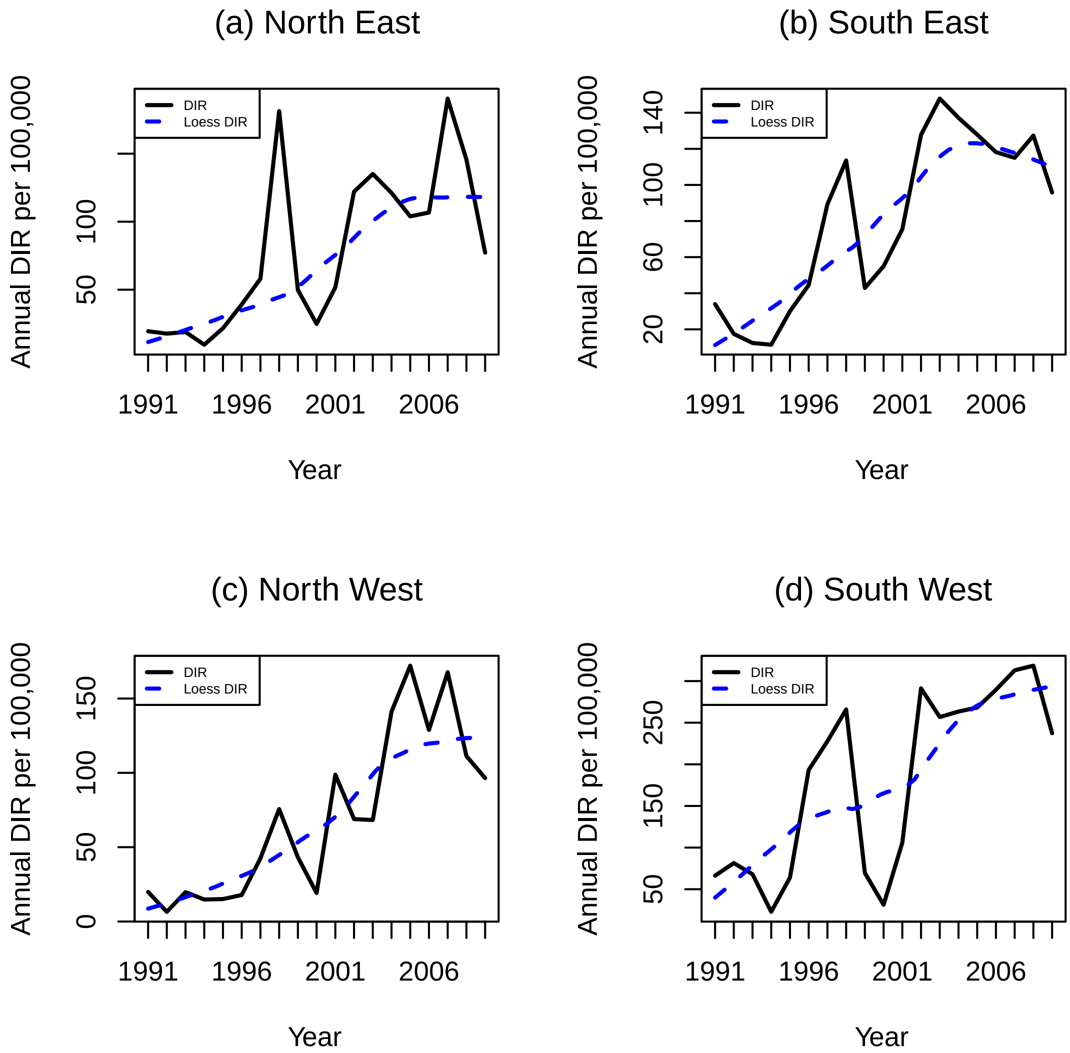

Figure 4.3: *Annual DIR per 100,000 population for the four regions (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 1991-2009.*

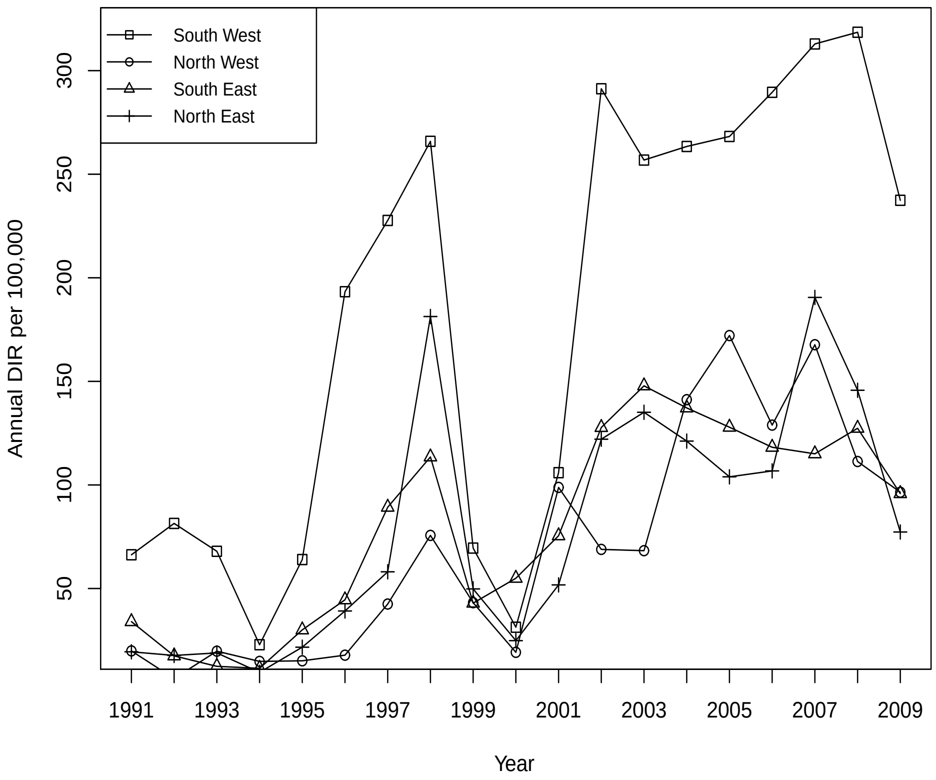

Figure 4.4: *Annual DIR per 100,000 population for main regions of Malaysia considered in this study.*

rises in 2007 in the North East and North West precede those in the South West and South East which occur in 2008. Some of these regional differences may result from differences in the strength of monsoon behaviour in the regions over these years. As discussed in Chapter 2, the Northeast monsoon and the Southwest monsoon have big impacts on climate in Malaysia, the two inter-monsoon seasons also marginally contribute to events throughout Malaysia (Soman and Slingo, 1997). Actual total numbers of dengue cases and the average annual DIR for the four regions for all of the 19 years is given in Table 4.3 which again emphasises the high levels in the South West which recorded 244,241 over the period and an average DIR of 180.75 cases per 100,000 whereas the other regions show broadly similar and much lower figures.

Table 4.3: *Total number of dengue cases and average DIR in each region from 1991-2009 for Malaysia.*

Region	Total Dengue	Average DIR
North East	53,357	78.67
South East	64,542	80.17
North West	44,219	69.90
South West	244,241	180.75

The regional picture discussed above can be further refined by looking at DIR trends in the period 1991-2009 within the individual states of each of the four regions, as shown in Figures 4.5, 4.6, 4.7 and 4.8.

First, it is noticeable that all states within all regions show an overall increasing trend over the 19 years. As to year by year patterns within the period, broadly speaking there is a lot of similarity in the patterns in the different states within a region, but, as might be expected, there is less similarity in patterns in states in different regions. Within that overall picture various differences between states within the same region can be noted.

In particular, in the North East, Kelantan did not experience the 1998 peaks

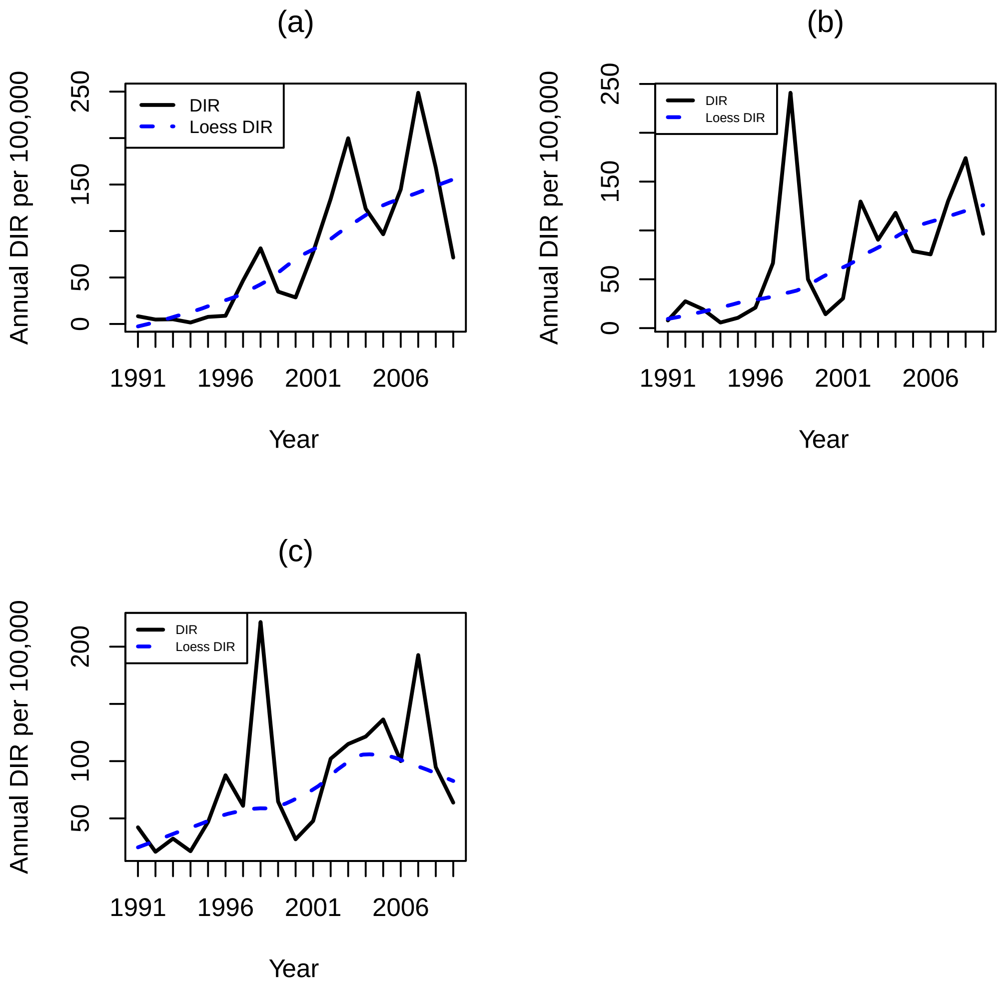

Figure 4.5: *Annual DIR per 100,000 population for 3 states (a) Kelantan, (b) Terengganu and (c) Pahang in the North East region from 1991 to 2009 for Malaysia.*

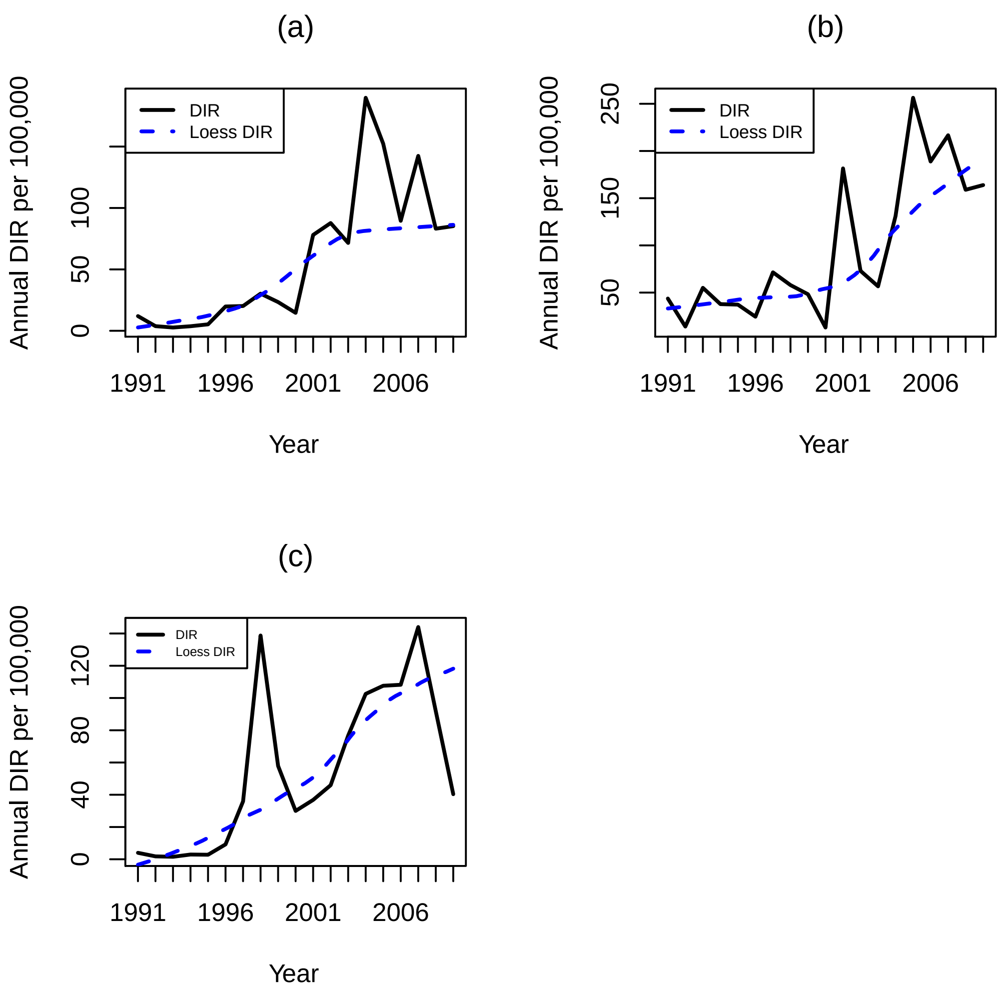

Figure 4.6: *Annual DIR per 100,000 population for 3 states (a) Perlis, (b) Penang and (c) Kedah in the North West region from 1991 to 2009 for Malaysia.*

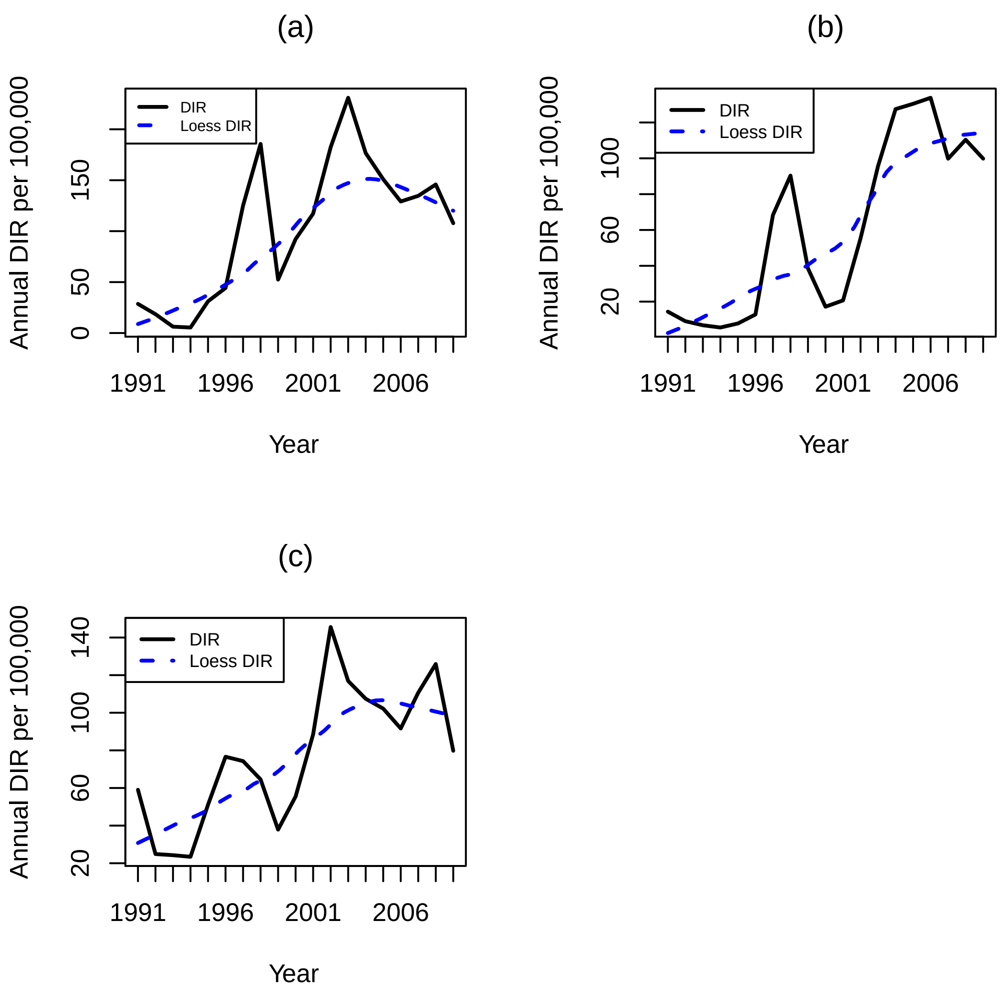

Figure 4.7: *Annual DIR per 100,000 population for 3 states (a) Negeri Sembilan, (b) Melaka and (c) Johor in the South East region from 1991 to 2009 for Malaysia.*

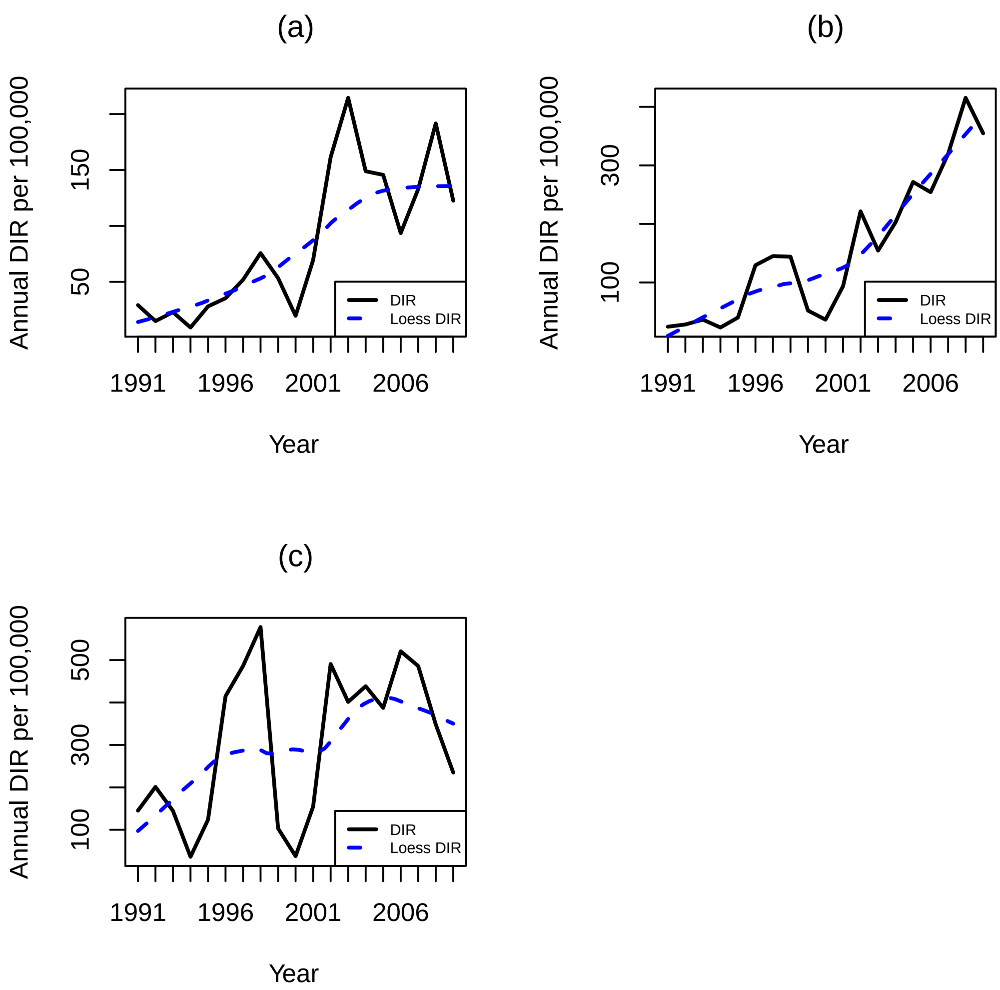

Figure 4.8: *Annual DIR per 100,000 population for 3 states (a) Perak, (b) Selangor and (c) Kuala Lumpur in the South West region from 1991 to 2009 for Malaysia.*

recorded in Terengganu and Pahang. Whereas, in the North West, rates in Perlis are generally higher than those in Penang and Kedah and the 1998 peak in Kedah is not seen in Perlis and Penang. Meanwhile, in the South East little major difference is evident between the three states of Negeri Sembilan, Melaka and Johor. Finally, in the South West, the very high levels in Kuala Lumpur and the strikingly different pattern in Selangor when compared with Perak and Kuala Lumpur are worthy of note. Meanwhile, just focusing on the most recent years from 2001-2009, Table 4.4 gives the state with the highest annual DIR in each of those years and shows Penang in North West for 2001 then followed by Kuala Lumpur in South West for the next six years from 2002 to 2007 but then replaced with Selangor also in the South West for 2008 and 2009.

Table 4.4: *State with highest annual DIR from 2001-2009.*

Year	State	Region	DIR
2001	Penang	North West	171.08
2002	K.Lumpur	South West	456.01
2003	K.Lumpur	South West	374.82
2004	K.Lumpur	South West	411.25
2005	K.Lumpur	South West	364.99
2006	K.Lumpur	South West	493.92
2007	K.Lumpur	South West	463.66
2008	Selangor	South West	419.28
2009	Selangor	South West	360.57

The reasons for these various inter-regional and intra-regional differences between states are undoubtedly complex, but as well as the monsoon/climate influences mentioned earlier, differential demographic changes may also be important.

For example, some states have experienced higher population growth rates and associated uncontrolled urbanisation resulting in the kinds of poor housing and inadequate water supply which then encourages ideal vector habitats to increase in those areas (Emilie et al., 2011).

4.2.2 Monthly Patterns in Dengue Incidence

Seasonal patterns in incidence rate are common in all vector-borne diseases due to the life-cycles of the vector and their dependence on the local climate for breeding areas, sufficient temperature etc. (see Johansson et al. 2009b). It is therefore important to explore seasonal patterns in monthly dengue incidence rate.

The monthly dataset described earlier covering the period from 2001 to 2009 can be used to look at seasonal patterns in monthly DIR in Peninsular Malaysia. The overall picture averaged over the 12 states and over the nine years is shown in Figure 4.9 which shows that there are two months in the year in which DIR peaks, January and July.

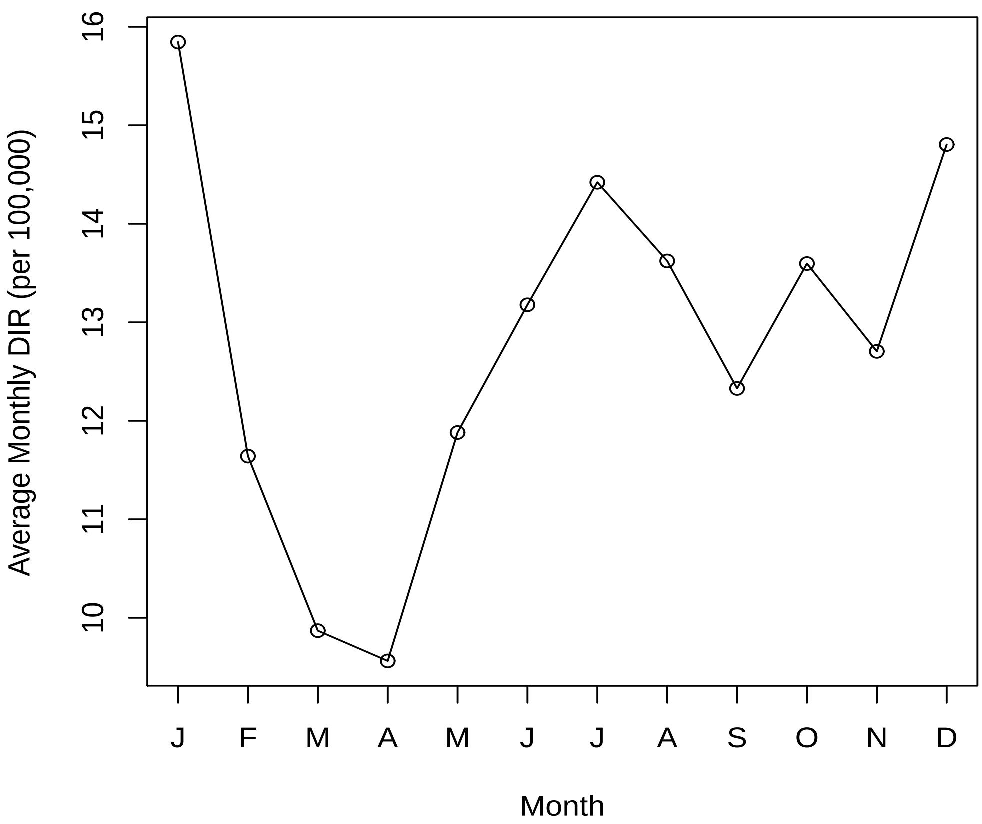

Figure 4.9: *Average monthly DIR values for Malaysia 2001-2009.*

This might be a result of the monsoon seasons — the Southwest monsoon occurs between the months of May to September and the Northeast monsoon occurs between November and March (with then latter carrying more rain). The two peak seasonal periods occur roughly in the central month of each monsoon season, in line with previous reports of the relationship between the monsoon occurrences and higher values of DIR in Malaysia (Soman and Slingo, 1997).

Given potential regional geographical and climatic differences it is also useful to look at the picture separated by the four regions used previously in this chapter. Figure 4.10 shows monthly DIR per 100,000 population for these four regions over the 108 month period. The most obvious aspect of this plot is the generally higher level of monthly DIR in the South West (noted earlier in annual DIR patterns) but it is difficult to extract meaningful differences in the overall seasonal pattern from this plot which confounds both trends over years as well as cycle within years. More informative perhaps, is the average monthly DIR in each region over the nine years as shown in Figure 4.11 which displays the pattern from June through to May in the subsequent year.

Meanwhile, Table 4.5 provides the total number of dengue cases and the maximum recorded monthly DIR for each region over the corresponding period. The main point from Table 4.5 is the simple observation (similar to that noted earlier when looking at annual DIR) of the high numbers and high maximum monthly rate in the South West region when compared to lower (and roughly equal) figures in the other three regions.

Figure 4.11 similarly indicates the notable difference between the South West and the other regions in that monthly DIR remains consistently high throughout the year. However, this figure also shows some similar features in the seasonal cycle in the North East, South East and South West with an annual peak in January and another in July; the latter being more pronounced in the North East, less evident in the South West and even less so in the South East. This pattern is to be expected due to the similar impact of the South West Monsoon on these three regions. This

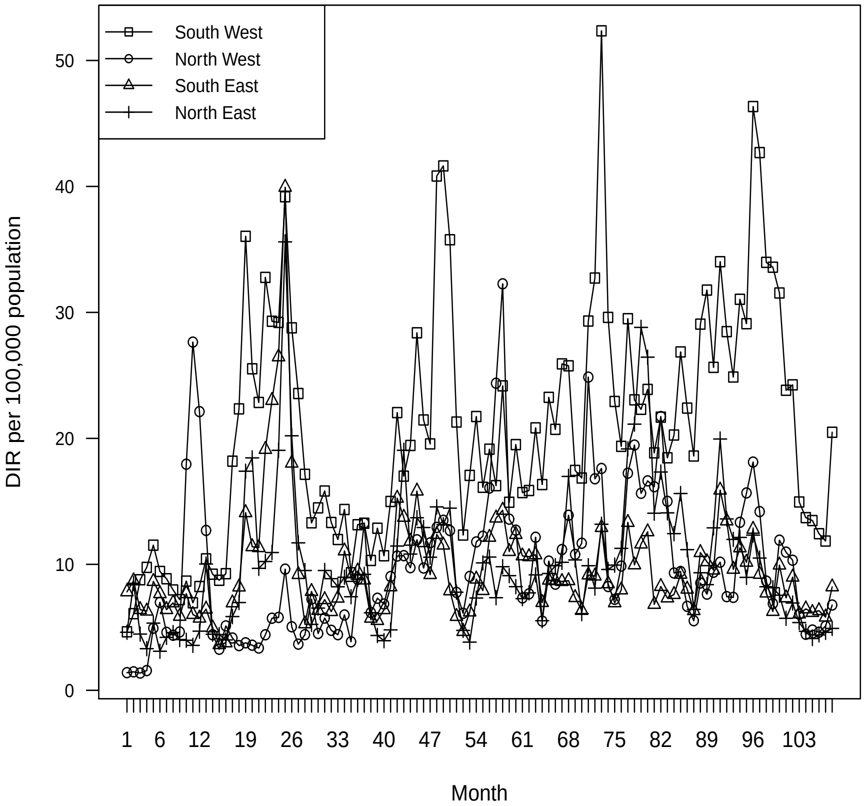

Figure 4.10: *Monthly DIR per 100,000 population for North East, South East, North West and South West of Malaysia from 2001 to 2009.*

Table 4.5: *Notified total of dengue and maximum monthly DIR of main regions in Malaysia from 2001-2009.*

Region	Total Dengue	Max DIR
North East	39,282	35.60
South East	46,779	39.93
North West	35,024	32.27
South West	187,919	52.35

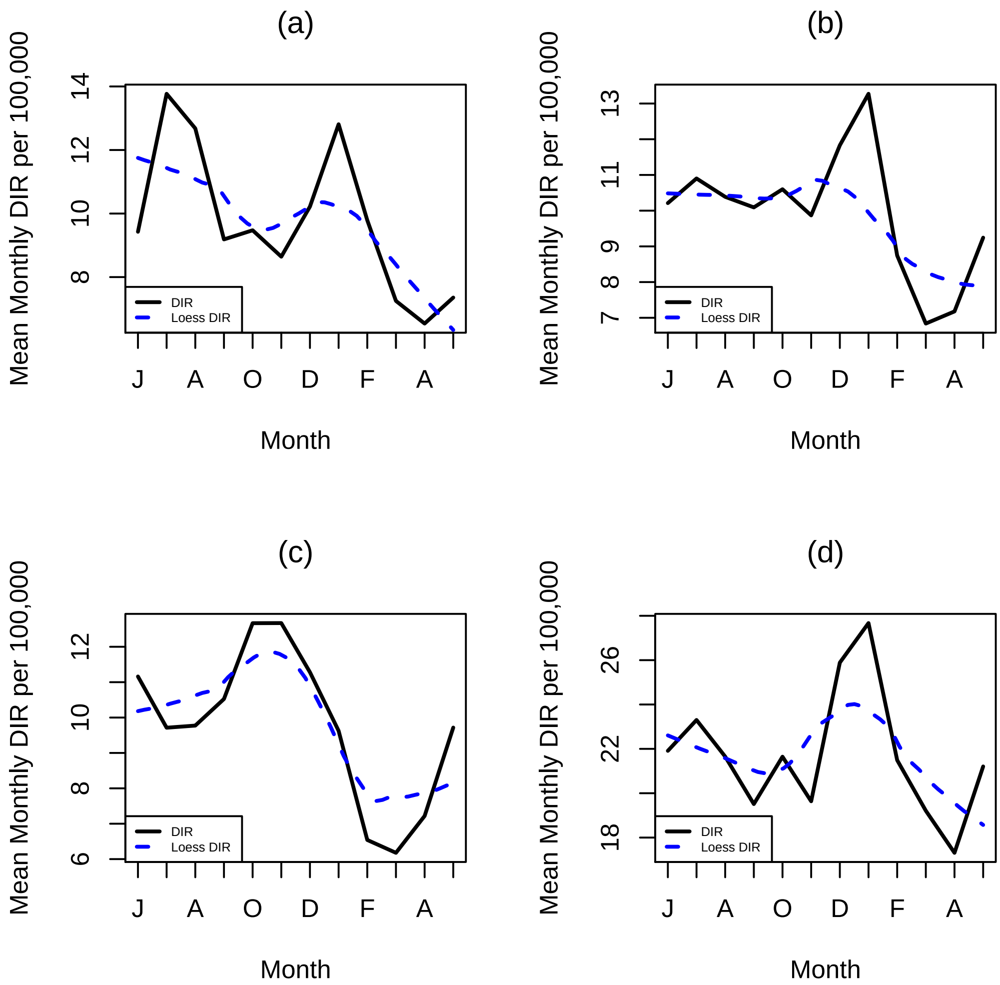

Figure 4.11: *The mean annual cycle of monthly DIR per 100,000 population for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

monsoon mainly affects the states with a Western coastline and which applies to various parts of the North West, South East and South West regions. The South East region has a particularly high DIR value in July compared to January due to the Eastern coastline in the state of Johor and the impact of the Northeast monsoon. As mentioned before, the climate of Malaysia is described by two monsoon seasons and two inter-monsoon seasons. The Southwest monsoon season occurs from May to September while the Northeast monsoon season is from November to March. During the Northeast monsoon it is the exposed areas in the Eastern part of Malaysia which receive heavy rainfall while the Southwest monsoon impacts the Western part of Malaysia as described in Suhaila et al. (2010a). Since monsoons bring heavy rainfall to the local regions this may have strong influences on the seasonal cycle in DIR (Oki and Musiake, 1994; Aiken et al., 1980). Increases in dengue after heavy rainfall particularly occur in urban areas where static rainwater provides mosquitoes with suitable breeding conditions. Warmer temperatures can also affect the transmission of the dengue virus as this allows the mosquitoes vector to survive and reach maturity early than expected (Muhammad Azami et al., 2011). Interestingly, the annual cycle in the North West looks rather different to that in the other three regions with peak monthly DIR occurring in November and then to a lesser degree in May/June. Suhaila et al. (2010b) found that during the Southwest monsoon season the North West region records high levels of rainfall, but the existence of the Titiwangsa Range blocks the Northeast monsoon from strongly affecting the North West region.

It is also useful to look at seasonal cycles within the individual states comprising each of the four regions. The average DIR in each month for 2001-2009 in the states is shown in Figures 4.12, 4.13, 4.14 and 4.15. There is clearly a fair amount of local variation here, but just commenting on some of the most noticeable differences: in the North East the mean monthly cycle in Kelantan and Terengganu is broadly similar and somewhat different to that in Pahang; in the North West the mean monthly cycle in Perlis stands out as different to that in Penang and Kedah in having no evident January peak; in the South East the January peak is evident

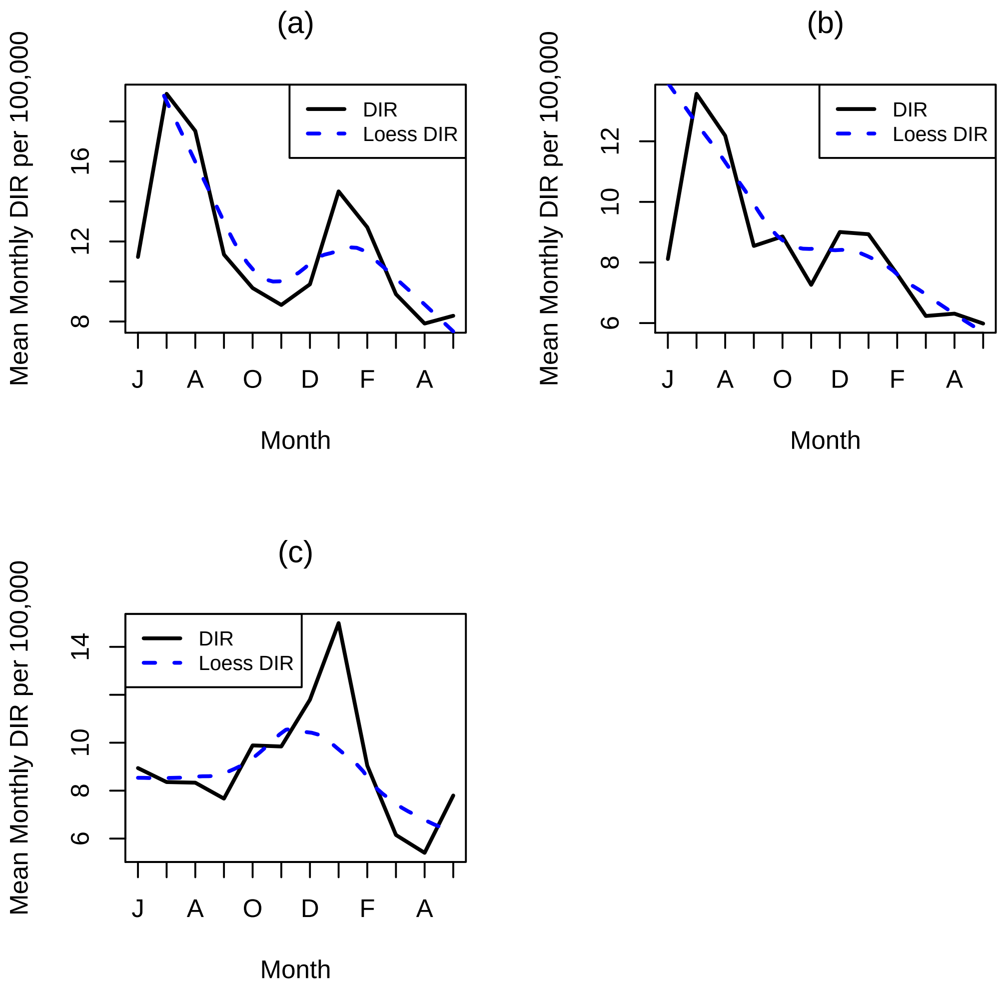

Figure 4.12: *The mean monthly cycle of DIR per 100,000 population for 3 states; (a) Kelantan, (b) Terengganu and (c) Pahang in the North East region in Malaysia from 2001 to 2009.*

in all the three states of N.Sembilan, Melaka and Johor, whereas the behaviour at other times of the year differs somewhat in these states; finally, in the South West Perak and Selangor behave relatively differently to Kuala Lumpur in that the July peak is much more marked in the latter case.

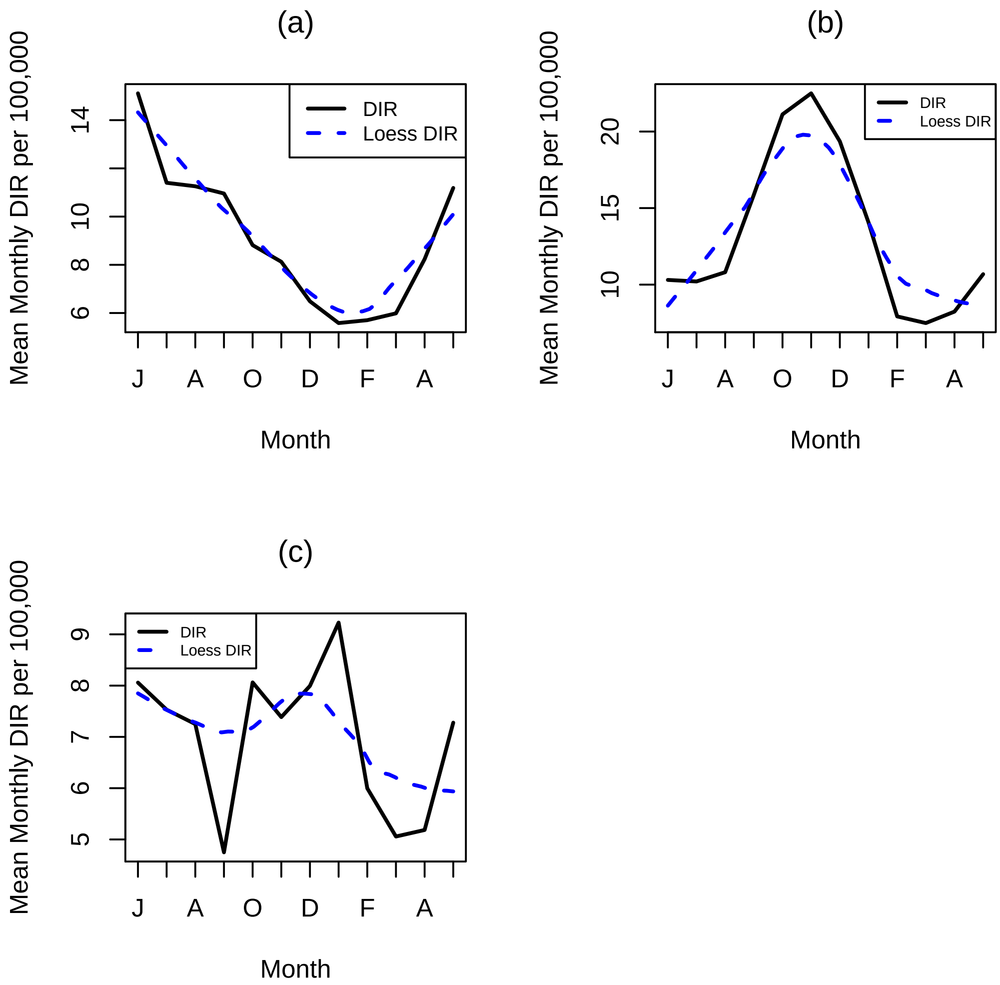

Figure 4.13: *The mean monthly cycle of DIR per 100,000 population for 3 states; (a) Perlis, (b) Penang and (c) Kedah in the North West region in Malaysia from 2001 to 2009.*

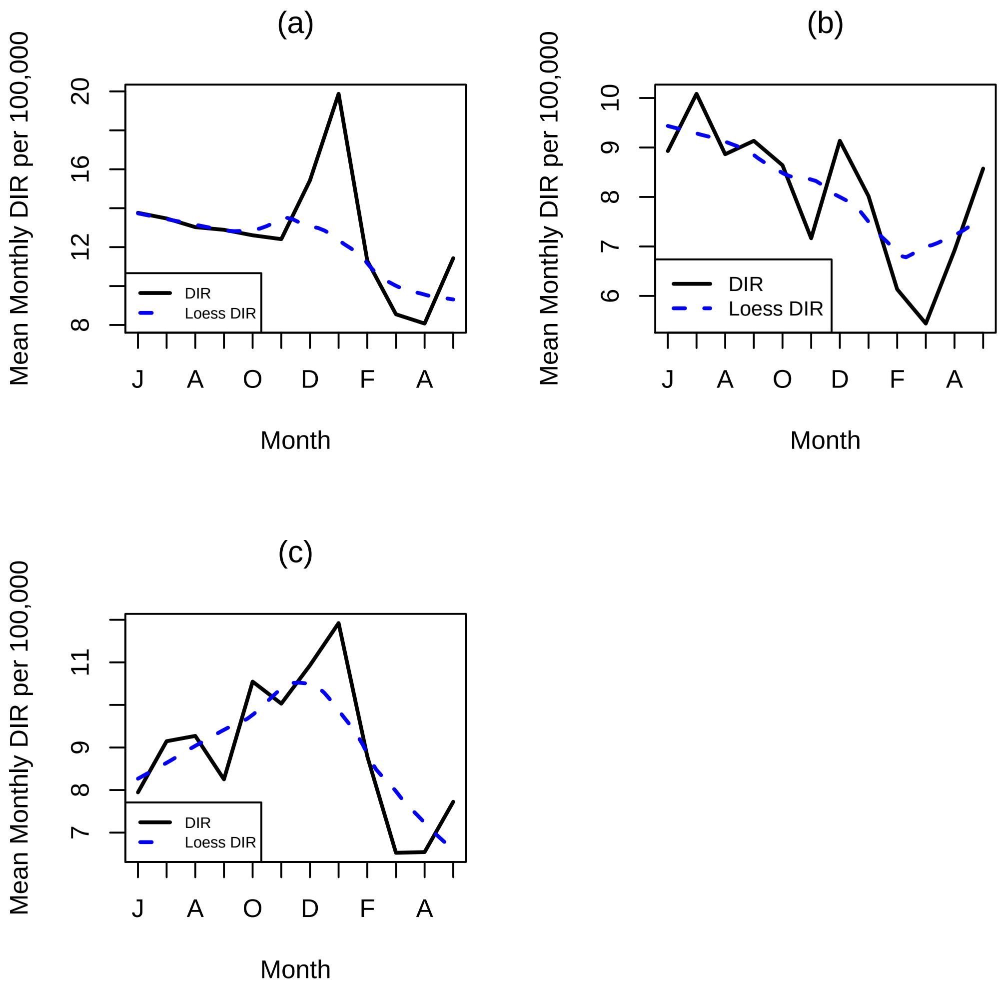

Figure 4.14: *The mean monthly cycle of DIR per 100,000 population for 3 states; (a) N.Sembilan, (b) Melaka and (c) Johor in the South East region in Malaysia from 2001 to 2009.*

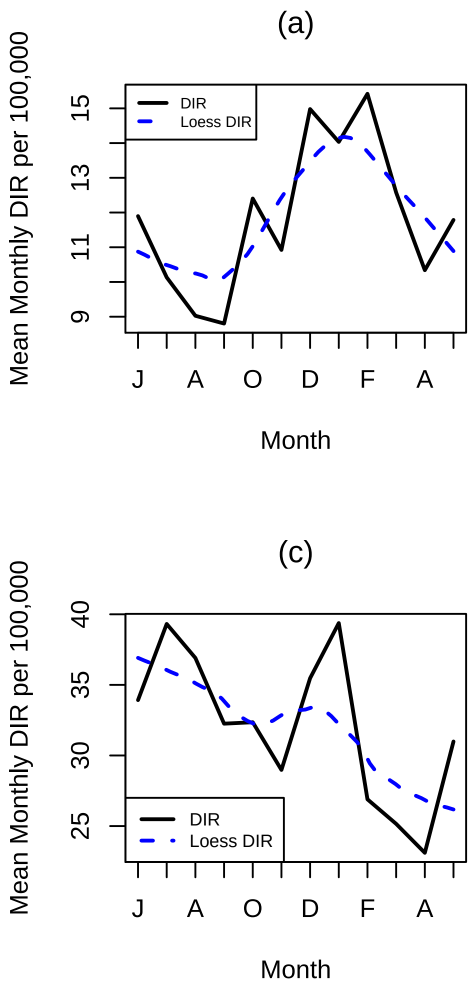

Figure 4.15: *The mean monthly cycle of DIR per 100,000 population for 3 states; (a) Perak, (b) Selangor and (c) Kuala Lumpur in the South West region in Malaysia from 2001 to 2009.*

4.3 Dengue Incidence and Demographic Data

General trends in population growth in Malaysia have already been touched in the introduction to Section 4.2. The 2010 Population and Housing Census of Malaysia (Census 2010) was the fifth decennial census to be conducted since the formation of Malaysia in 1963 (previous censuses being conducted in 1970, 1980, 1991 and 2000) and revealed that the total population of Malaysia was 28.3 million, compared with 23.3 million in 2000. This gives an average annual population growth rate of 2.0% for the period 2000-2010[1]. The increasing trend in DIR over this period has already been discussed and it is of note that this is particularly marked in those states in the South West of the country (such as Kuala Lumpur) where the main urban areas of Malaysia are located and where there is a higher population density. The effect of high population density on the incidence of dengue fever has been noted in several studies. In Brazil, it was reported that 70% of the individuals in urban populations in the country had previously contracted dengue fever, implying positive correlation between DIR and population density (Siqueira et al., 2005); whilst, Gubler (2002) has commented more generally on the role of population growth, increased urbanisation and improved transportation systems as contributors to the increased global incidence of dengue fever. It is also undoubtedly the case that in many developing countries population density acts as a surrogate measure for poor living conditions and social inequalities which are also well known risk factors for dengue (Guzman and Kouri, 2002; Mondini and Chiaravalloti, 2007; Stefan et al., 2008). That said, the relationship between DIR and population density is not necessarily straightforward. Improvements in water supply and vector control may also be associated with increased population density and these will help in control and prevention of dengue transmission. Wolf et al. (2011) have reported that areas having high population density with adequate water supply do not experience severe dengue outbreaks compared to rural areas where there is high risk of dengue due to lack of piped water supply and thus more mosquito breeding sites in water

[1]http://www.statistics.gov.my/portal/index.php

storage containers.

Figure 4.16 shows the relationship between the logarithm of monthly DIR and population density in the 12 states of Malaysia for the 108 months period from 2001-2009. As expected, this plot demonstrates high levels of variability; but, nevertheless, there is some clear evidence of higher DIR being associated with those states with very high population density.

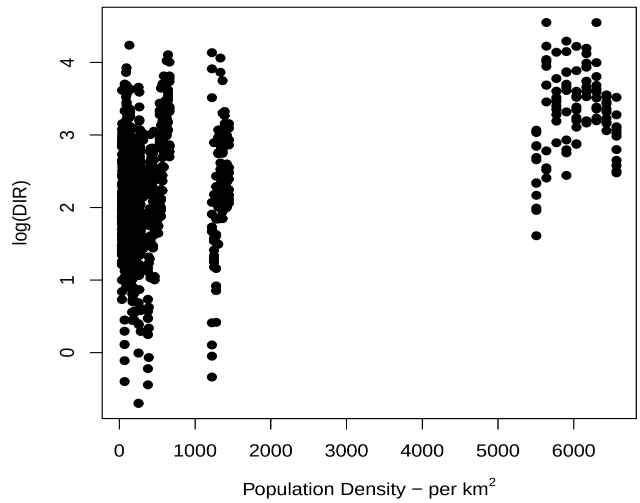

Figure 4.16: *Relationship between log monthly DIR and population density for 12 states of Malaysia 2001-2009*

Another way in which demographics (both population and population density) may indirectly influence DIR arises from the infectious nature of the disease - higher populations provide more hosts for the virus and higher population densities provide for higher chances of transmission. One way to capture this infectious effect is to look at DIR in relation to DIR in the immediately preceding months. Figure 4.17 shows the relationship between the logarithm of monthly DIR and that lagged by one, two and three months in the 12 states of Malaysia for the 108 months period from 2001-2009. There is some evidence of the positive relationship in each case and it is particularly of interest that this extends to DIR lagged by 3 months, since this relationship could be practically useful in developing predictive DIR models in subsequent chapters as opposed to shorter lags where data would not be available in a practical setting.

4.4 Dengue Incidence and Climate Data

As discussed in Chapter 3, studies on the relationship between weather, climate and dengue have attracted serious attention from scientists throughout the world. As mentioned earlier in this chapter, the monsoon seasons play a major role in the Malaysian context (Suhaila et al., 2010b). The pattern has a large effect on rainfall, and is one of the main contributors in dengue epidemics in some regions of Malaysia as highlighted by Aiken et al. (1980). This section explores the relationship between observed monthly DIR in Malaysia and climatic factors such as rainfall, temperature, humidity, number of rainy days and ENSO.

Figure 4.18 shows the relationship between the logarithm of monthly DIR and average monthly rainfall and its lagged values in the 12 states of Malaysia for the period 2001-2009. There appears little relationship in current and lag 1 month (possibly slightly negative) but more of a positive relationship at lags of 2 and particularly 3 months. This aligns with the findings of Souza et al. (2010) which reported positive correlation between the building infestation rate and number of dengue cases and rainfall.

Figure 4.19 shows the corresponding picture with respect to number of rainy days in the month which is potentially useful alternative measure to average rainfall (amount versus intensity). Little relationship is apparent except perhaps at lag 3 where there is perhaps some suggestion of a negative trend. The relationship between the logarithm of monthly DIR and average monthly temperature and its lagged values in the 12 states of Malaysia for the period 2001-2009 is shown in Figure 4.20. Again there is high variability but some indication of positive associations with lagged temperature at 1, 2 and 3 months. The corresponding picture with regard to humidity is shown in Figure 4.21. Here there is little evidence

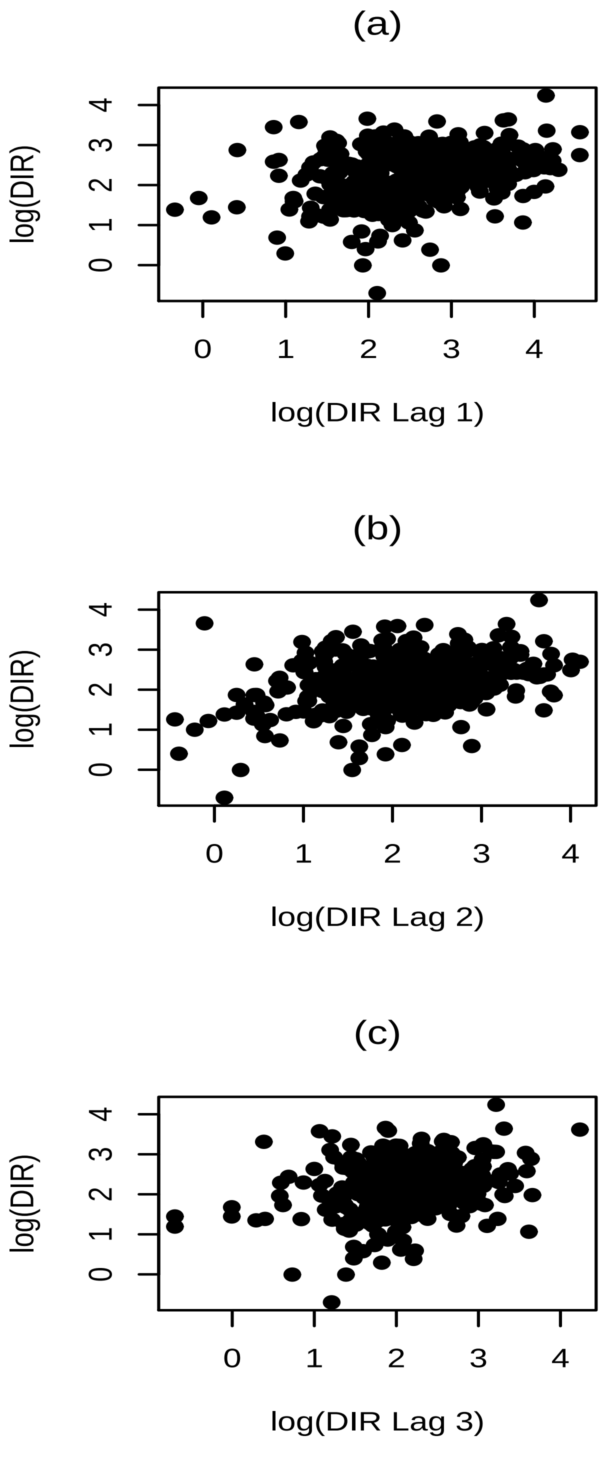

Figure 4.17: *Relationship between log monthly DIR for 12 states of Malaysia 2001-2009 and (a) log monthly DIR lag 1 Month, (b) log monthly DIR lag 2 Months, (c) log monthly DIR lag 3 Months.*

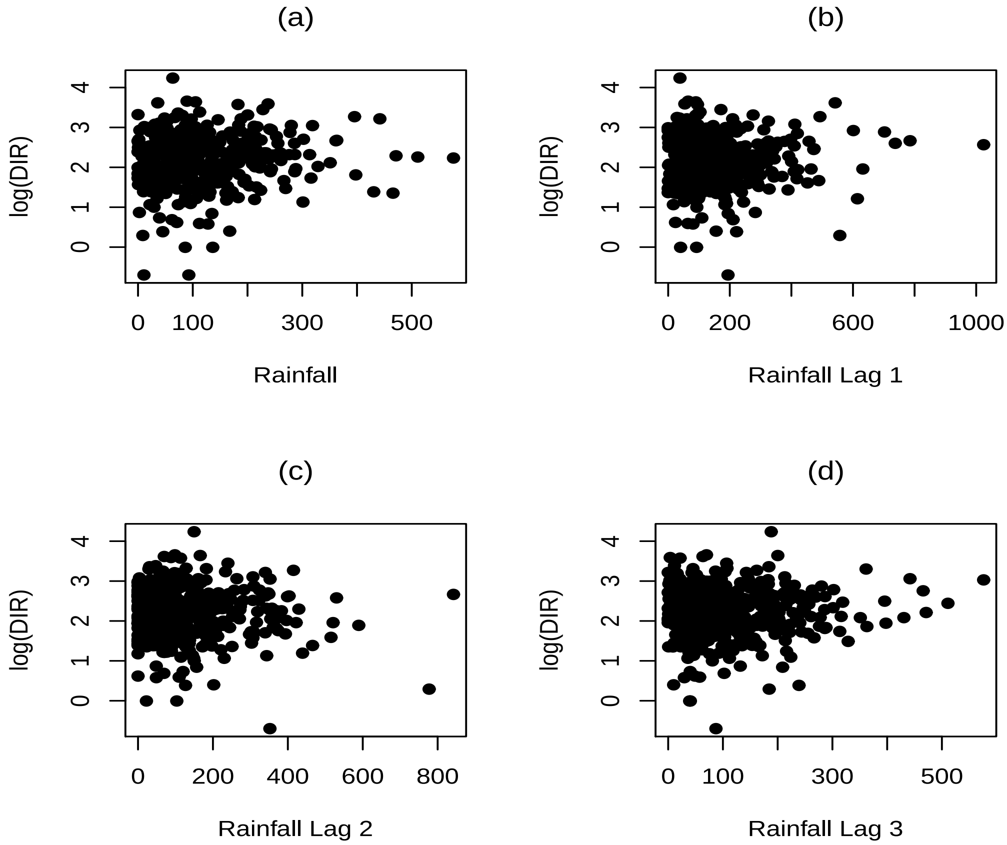

Figure 4.18: *Relationship between log monthly DIR for 12 states of Malaysia 2001-2009 and (a) Rainfall Current Month, (b) Rainfall Lag 1 Month, (c) Rainfall Lag 2 Months and (d) Rainfall Lag 3 Months.*

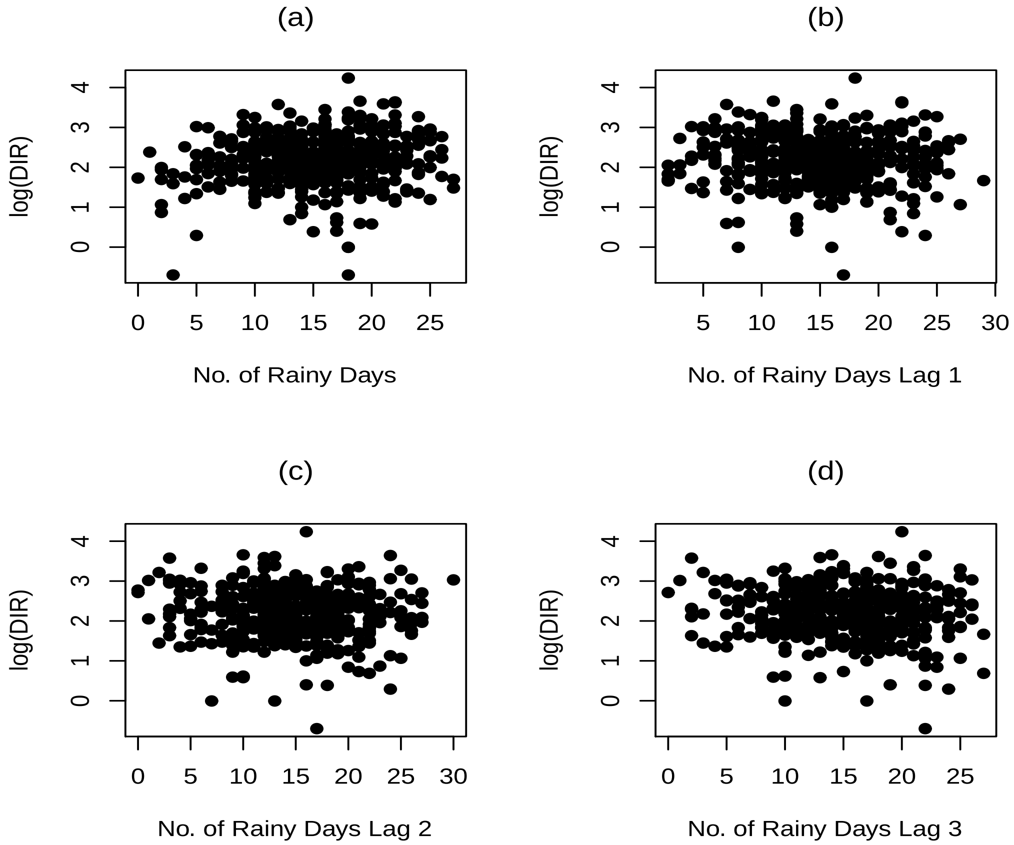

Figure 4.19: *Relationship between log monthly DIR for 12 states of Malaysia 2001-2009 and (a) Number of Rainy Days Current Month, (b) Number of Rainy Days Lag 1 Month, (c) Number of Rainy Days Lag 2 Months and (d) Number of Rainy Days Lag 3 Months.*

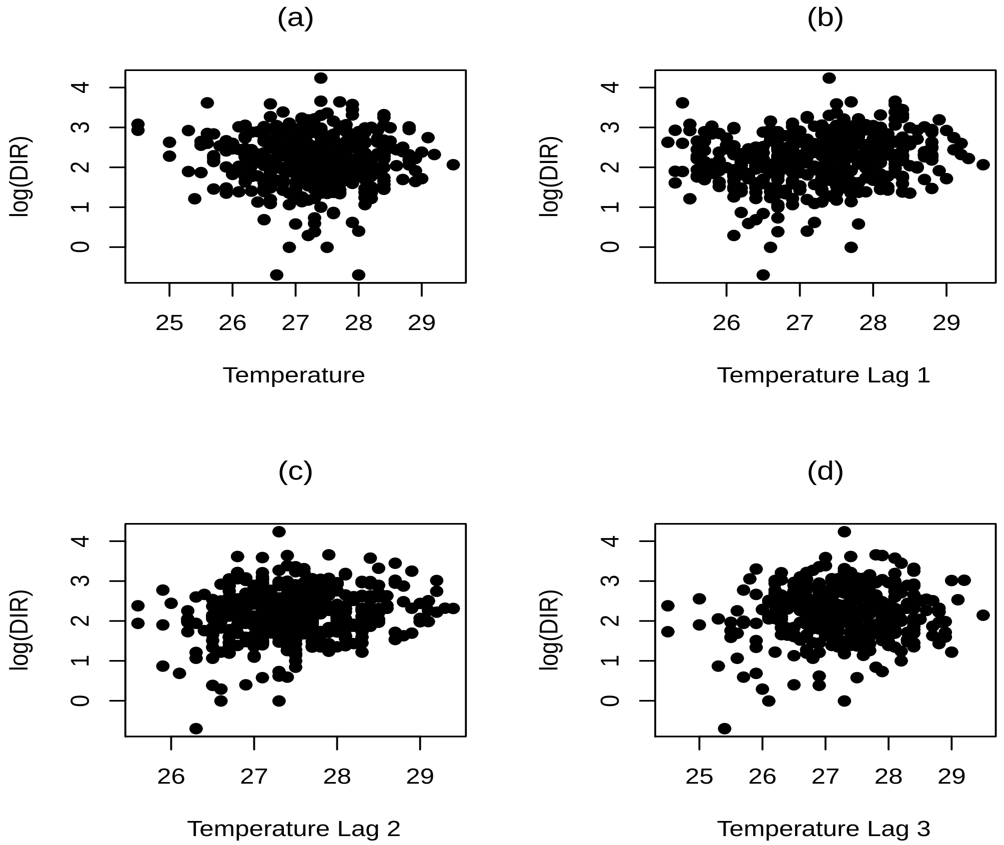

Figure 4.20: *Relationship between log monthly DIR for 12 states of Malaysia 2001-2009 and (a) Temperature Current Month, (b) Temperature Lag 1 Month, (c) Temperature Lag 2 Months and (d) Temperature Lag 3 Months.*

of any relationship with humidity either in the current month or in the preceding three months.

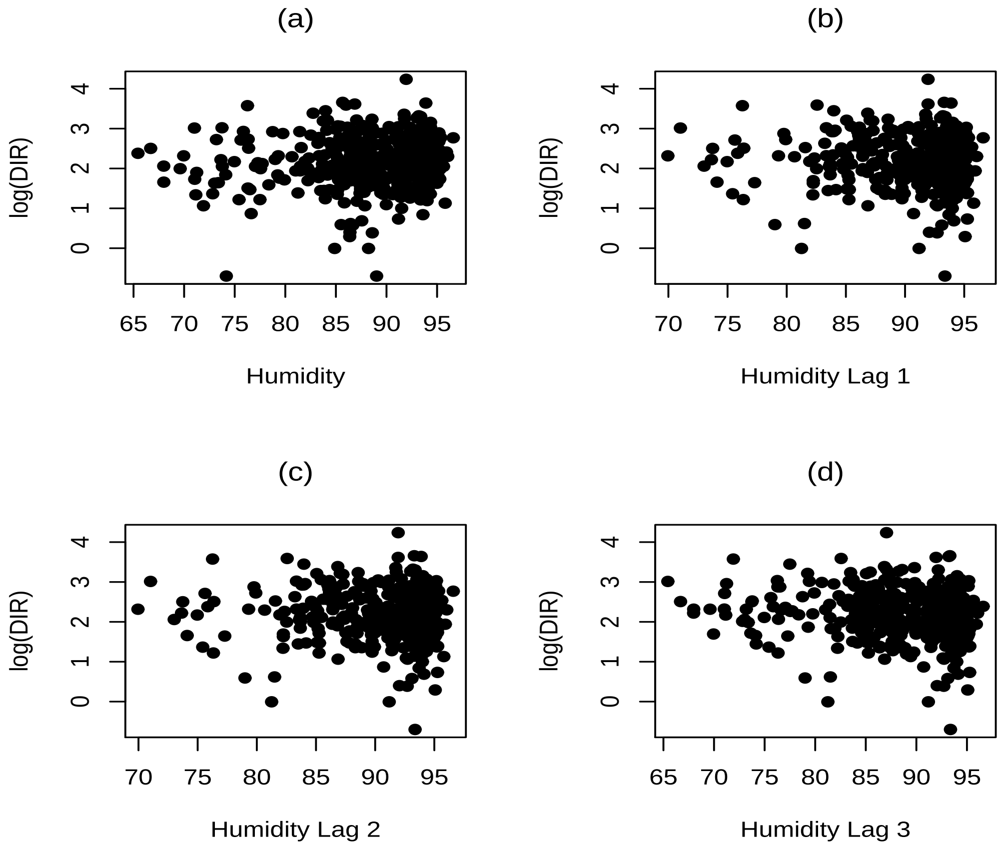

Figure 4.21: *Relationship between log monthly DIR for 12 states of Malaysia 2001-2009 and (a) Humidity Current Month, (b) Humidity Lag 1 Month, (c) Humidity Lag 2 Months and (d) Humidity Lag 3 Months.*

Previous sections of this chapter have identified strong regional differences in seasonal patterns of DIR in Malaysia and so it is perhaps useful to investigate some of the relationships with climate variables at a regional level. Figure 4.22 shows scatter plots of the logarithm of monthly DIR and average rainfall in the same month for the four different regions of Malaysia from 2001 to 2009. Very little convincing relationship is apparent in any of the regions. The corresponding picture for average rainfall lagged by 3 months is shown in Figure 4.23 and here some positive association is apparent in both the South East and particularly the South

West regions.

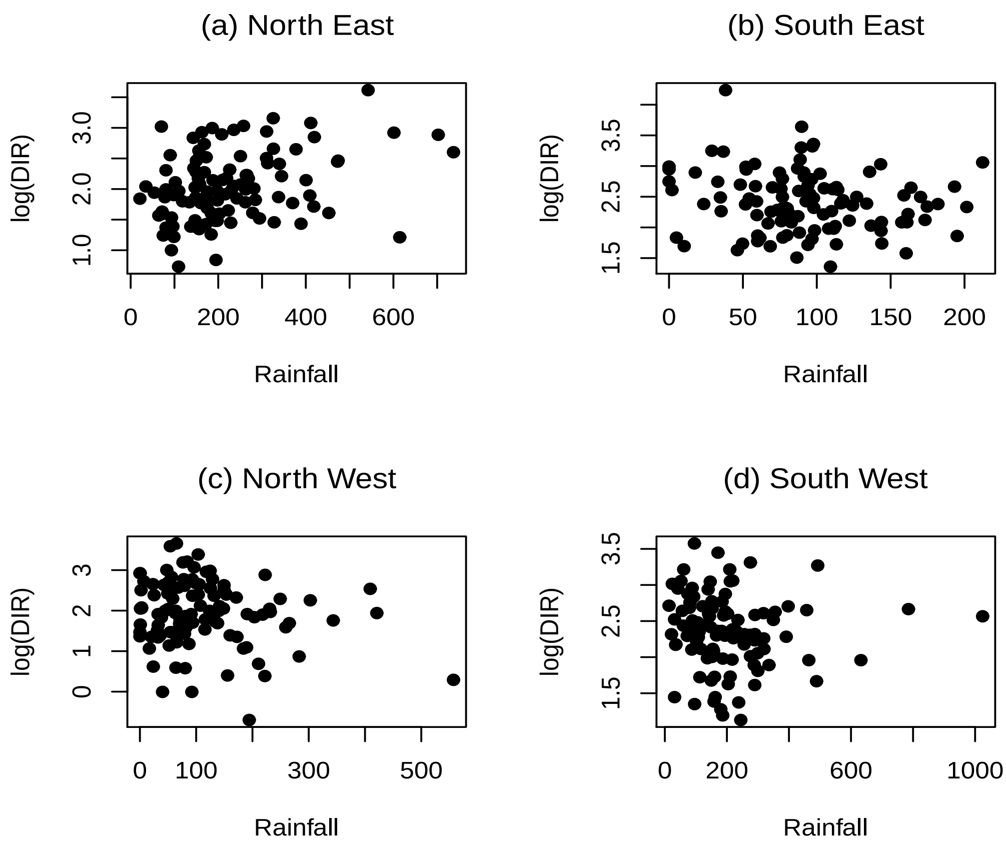

Figure 4.22: *Relationship between log monthly DIR and average rainfall in same month for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

When regional level relationships between log monthly DIR and numbers of rainy days in the month are investigated (see Figures 4.24 and 4.25) the only suggested relationships are with rainy days lagged by 3 months and particularly in the South West.

Regional level relationships between monthly DIR and temperature in the same month and in previous 3 months are shown in Figures 4.26 and 4.27. A negative relationship with temperature in the same month and in previous 3 months is perhaps evident in the North East region, but otherwise relationships are weak.

Some of the complexity of potential interactions in the above relationships between

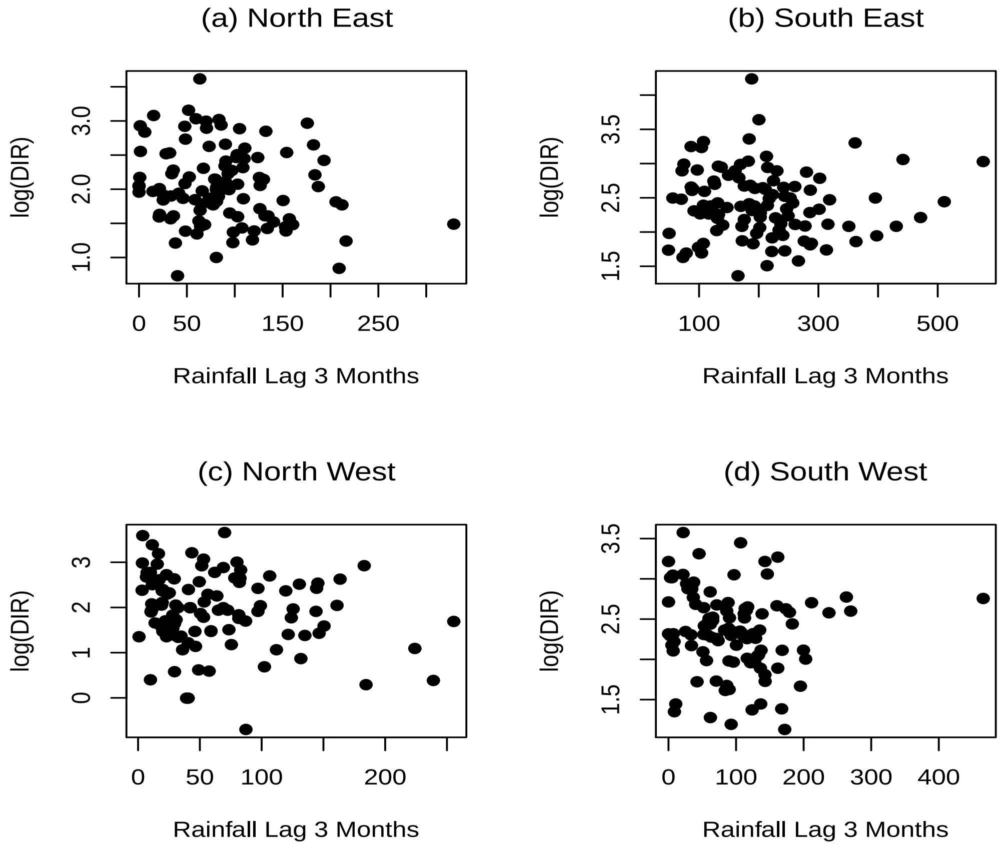

Figure 4.23: *Relationship between log monthly DIR and average rainfall 3 months previously for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

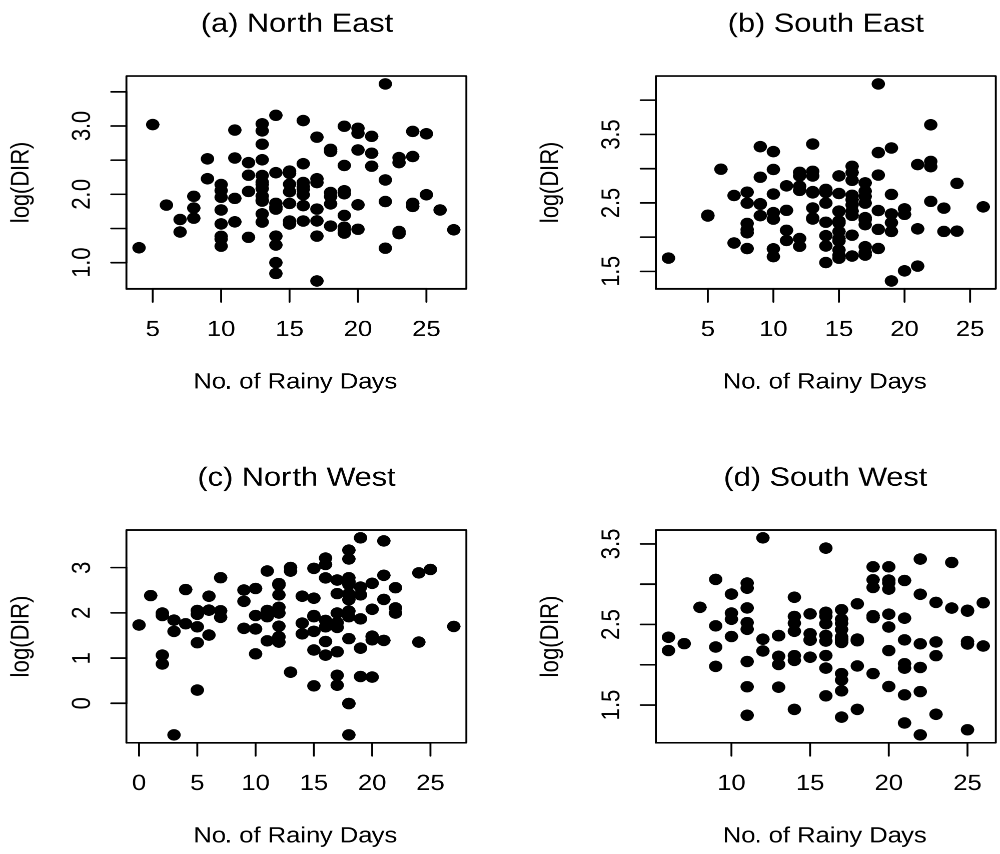

Figure 4.24: *Relationship between log monthly DIR and number of rainy days in same month for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

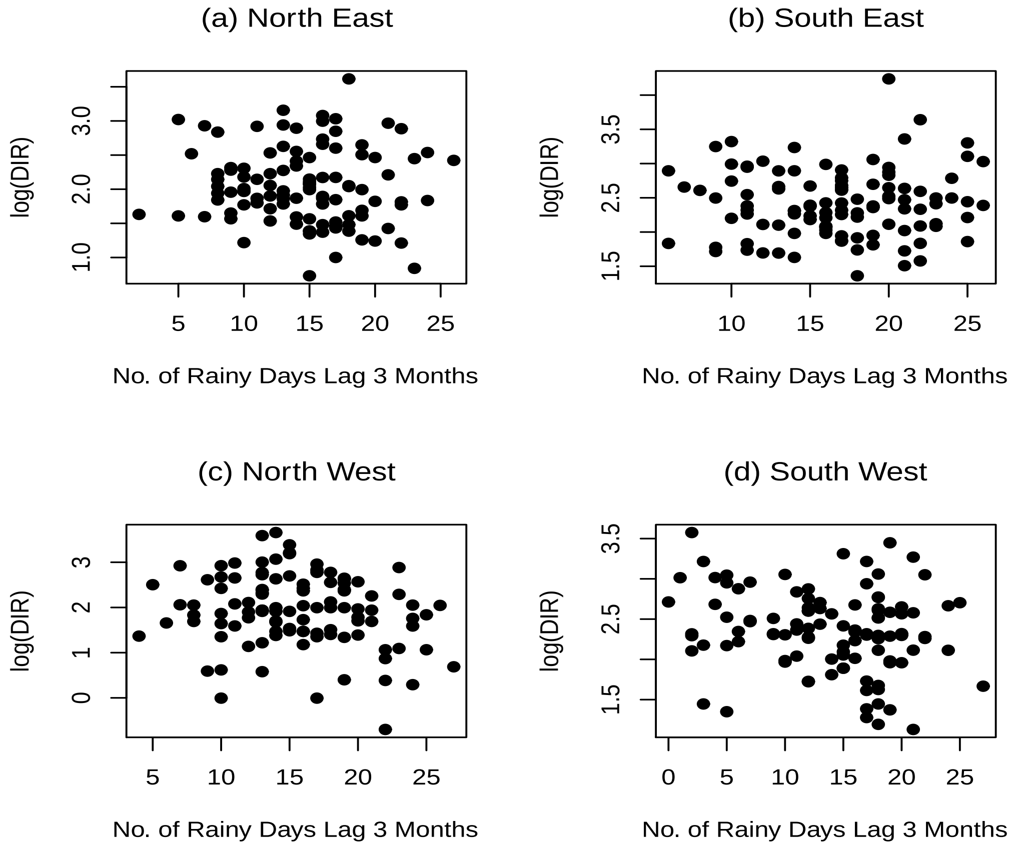

Figure 4.25: *Relationship between log monthly DIR and number of rainy days 3 months previously for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

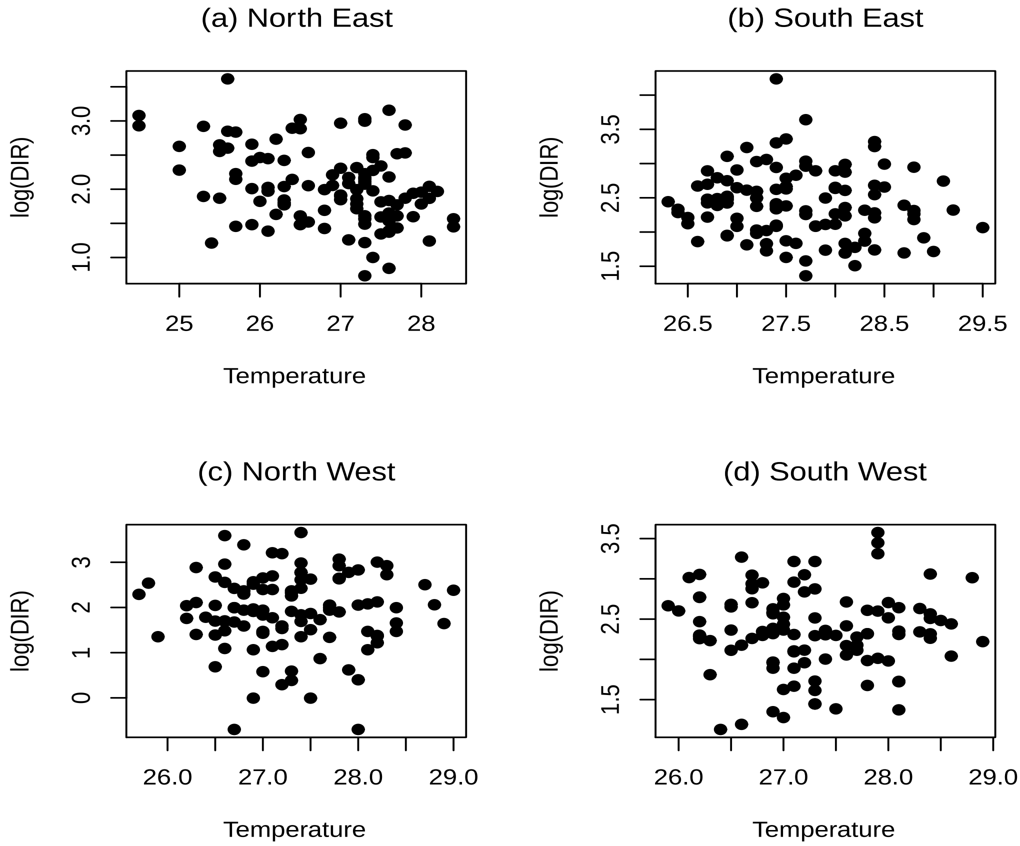

Figure 4.26: *Relationship between log monthly DIR and average temperature in same month for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

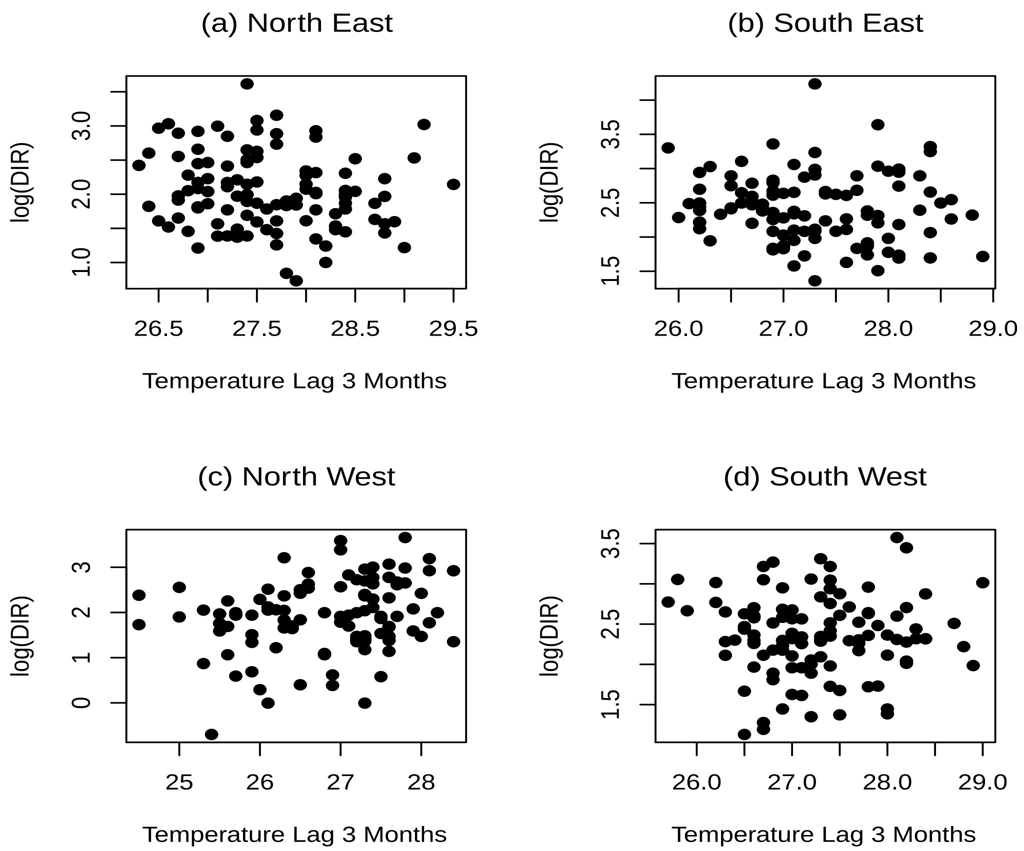

Figure 4.27: *Relationship between log monthly DIR and average temperature 3 months previously for (a) North East, (b) South East, (c) North West and (d) South West of Malaysia from 2001 to 2009.*

climate variables and DIR is perhaps illustrated in Figure 4.28 which shows the mean annual cycle over 2001-2009 of monthly DIR, average rainfall and temperature together with number of rainy days for the South West region for the period January 2001 to December 2009. DIR peaks in February (Figure 4.28-a), while rainfall peaks 3 months earlier in November (Figure 4.28-b) along with the highest number of rainy days (Figure 4.28-c); then, temperature peaks in May (Figure 4.28-d).

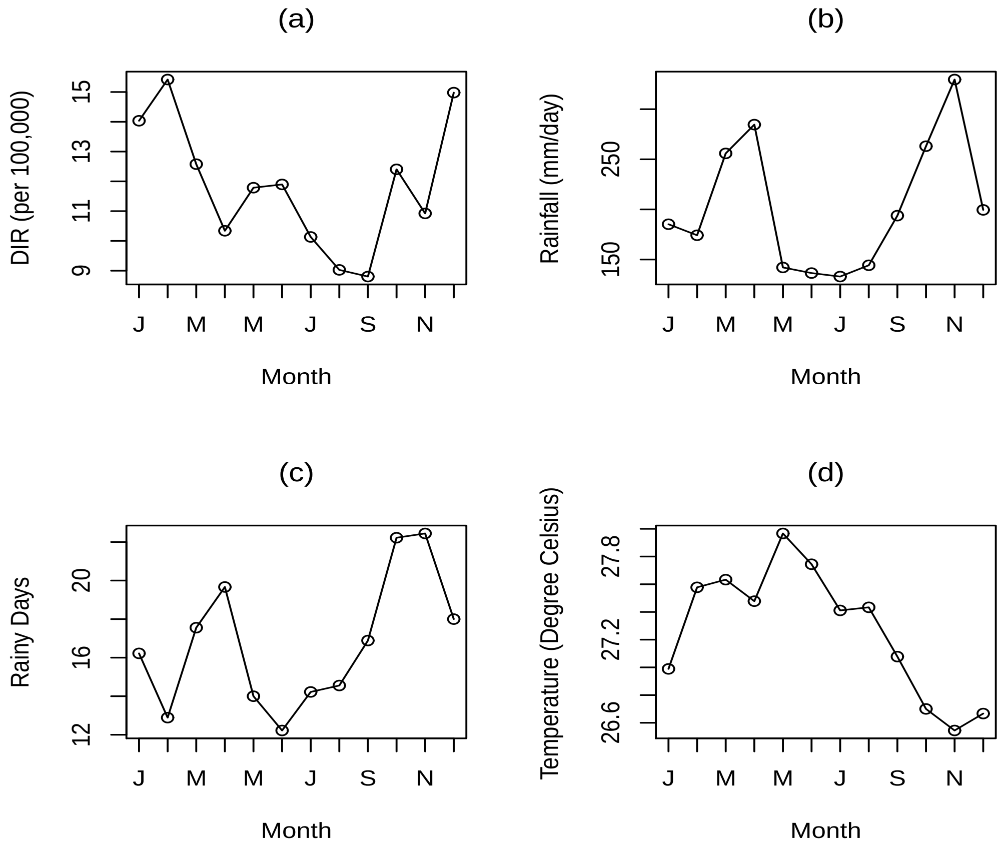

Figure 4.28: *Mean annual cycle of (a) log monthly DIR, (b) Average Rainfall, (c) Average Temperature (d) No. of Rainy Days in South West of Malaysia from 2001 to 2009.*

Finally in this section, we turn to possible relationships between ENSO and dengue incidence in Malaysia. As discussed in Chapter 2, ENSO influences inter-annual climate variability with warm El Niño and cold La Niña phases engendering significant temperature and precipitation anomalies around the world. Both events are sometimes described as very weak, weak, moderate, strong and very strong, depending on their impacts (Glantz, 2001). Other classifications (Webster and Palmer, 1997) focus on the magnitude of ENSO parameters, such as SST deviations or the geographical area covered by the pool of warm water in the Pacific. Very strong events can result in temperatures up to 3.5 °C above average in the eastern Pacific, with localised warming of up to 9 °C.

Consequence of ENSO for dengue transmission and for related infectious diseases are an area of current research and could perhaps become an important contributor to the development of Early Warning Systems (EWS) for dengue in countries such as Malaysia. Epidemics of dengue fever in many tropical countries have potential links with climatic anomalies associated with ENSO (Kovats et al., 1999). Analyses of the relationship of DIR to ENSO and local weather present challenges. Several studies have looked at associations between dengue epidemics and ENSO, such as Johansson et al. (2009a), which reported time-series analyses for Puerto Rico, Mexico and Thailand, and found no systematic association between dengue outbreaks and ENSO; whereas Gagnon et al. (2001) highlighted that there is a statistically significant correlation between El Niño and dengue epidemics in French Guiana and Indonesia where these regions experience statistically significant warmer temperature and less amount of rainfall in El Niño years.

Turning to the current study, a time series of the Oceanic Niño Index (ONI), defined as the 3-month running mean of SST anomalies in the Niño 4 region in the central Pacific was obtained from the NOAA Climate Prediction Center (CPC)[2] from 2001 to 2009 and the behaviour for this is illustrated in Figures 4.29. Using this index, the CPC defined ENSO events are when SST anomalies are $\geq +0.5$ for five consecutive months for warm (El Niño) and ≤ -0.5 for cold (La Niña). SST anomalies are weak $\leq \pm 0.5$, moderate $\geq \pm 0.5$ and strong event for ≤ -1.0 and ≥ 1.0.

Relating this to dengue DIR in Malaysia, Figure 4.30 shows a weak positive relationship between the logarithm of monthly DIR and ENSO at different lags for

[2]http://www.esrl.noaa.gov/psd/gcoswgsp/Timeseries/Nino4/

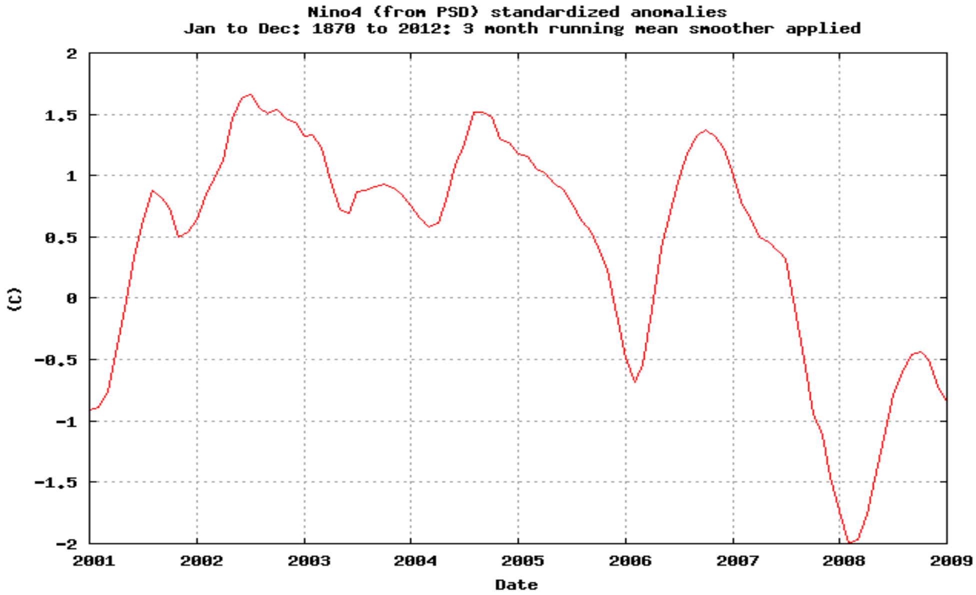

Figure 4.29: *Standardised anomalies of SST (Niño 4 Index) (3 months-running mean) (Jan 2001-Dec 2009).*

the 12 states of Peninsular Malaysia over the 9 years from 2001-2009. Visually these plots look very similar, but actually the strongest relationship (i.e. sample correlation) occurs at the longest lag (i.e. 6 months).

4.5 Summary

This chapter has presented exploratory analyses of possible relationships between annual and monthly DIR and climate and other factors that can be more formally used in model building in subsequent chapters. The variables that were considered included annual trend, in year seasonal effects, population, population density and lagged dengue incidence rate as well as climate factors such as average rainfall and temperature, number of rainy days, ENSO and lagged values of these climate variables.

The analyses presented have deliberately been informal based upon simple scatter plots with superimposed smooth fits in some cases. We have not reported correla-

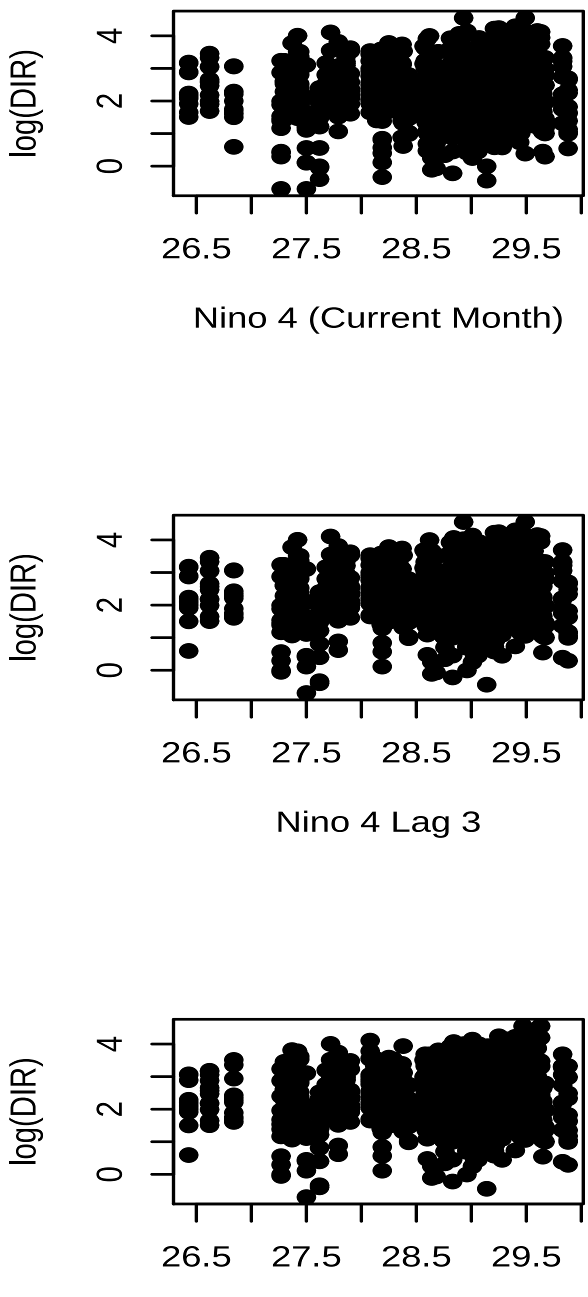

Figure 4.30: *Relationship between log monthly DIR and Niño 4 at different lags in the 12 states of Peninsular Malaysia from 2001 to 2009.*

tion coefficients or any formal tests of relationships because the object was simply to suggest relationships that might need to be incorporated into more formal statistical models in subsequent chapters. The key findings to emerge from the analyses presented are:

- There is some evidence of an increasing annual trend in DIR in all states of Malaysia
- There is a strong in-year seasonal cycle in DIR and that differences in this cycle may need to be allowed for in different broad geographical regions of Malaysia and possibly in different states
- High population density is positively related to monthly DIR as is DIR in the previous month
- Relationships between monthly DIR and climate variables are generally quite weak; nevertheless some relationships may be able to be usefully incorporated into predictive models. These include average temperature and rainfall, number of rainy days and ENSO. However lagged values of these variables need to be considered up to 6 months in the case of ENSO and from 1-3 months in the case of other variables.

In summary, DIR in Malaysia is potentially associated with country wide trend, regional seasonal cycle, population, population density, dengue incidence in preceding months, lagged average temperature, average rainfall, number of rainy days and ENSO.

In the next chapter, a framework will be proposed to model spatio-temporal variations in DIR which can incorporate and more formally assess the relative impact of such factors.

Chapter 5

Model Development

The aim of this chapter is to determine an appropriate modelling framework for monthly dengue incidence in Malaysia and using that framework then develop suitable spatio-temporal statistical models by testing and selecting appropriate explanatory variables from the potential associations identified and described in the preceding chapter. The data set used throughout this chapter will that involving monthly dengue counts for the 12 states of Peninsular Malaysia for the period 2001-2009 as described in Chapter 4.

5.1 Modelling Frameworks

The general approach adopted in this chapter follows that used in a number of other recent ecological modelling studies on dengue and is based on the generalised linear model (GLM) (Nelder and Wedderburn, 1972) and variations thereof. Examples of such studies where dengue count data has been modelled by using the GLM framework include those by Zuur et al. (2009), Hashizume et al. (2012), Krisada and Lily (2013), Lowe et al. (2013) and Cabrera (2013).

The rapid growth to the use of GLM in a variety of fields is due to its flexibility which allows the inclusion of an extensive set of distributions belonging to the exponential family. The GLM is a development of the linear model to accomodate both non-gaussian response distributions and transformations for non-linearity in the systematic model component. In a GLM, there are independent observations $(y_1, \ldots, y_n)$ on a response, where the distribution of y_i is in the exponential family with parameters θ_i and ϕ and with functions $a(\phi)$, $b(\theta_i)$, $c(y_i, \phi)$ chosen to be appropriate for the particular data i.e.:

$$p(y_i; \theta_i, \phi) = \exp\left[\frac{(y\theta_i - b(\theta_i))}{a(\phi)} + c(y, \phi)\right]$$

In order to complete the GLM specification, we have explanatory variables (predictors), $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})$ (could be quantitative or categorical, transformations of predictors, or polynomial terms) whose values may influence the distribution of the response y_i through a linear predictor; $\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$ which affects the mean of the response via a known, smooth and invertible 'link function' $g(\cdot)$ so that $g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$ function. Note the link function $g(\cdot)$ does not transform y_i, but rather its mean μ_i (e.g. a gaussian linear model with response $\log y_i$ is not the same as a GLM with normal error and a log link). For a full account of the theory and application of GLMs, see McCullagh and Nelder (1989).

The GLM specification is loose enough to encompass a wide class of models useful in statistical practice, but tight enough to allow the development of a unified methodology of parameter estimation (model fitting) and associated inference (at least approximate inference) based on general likelihood methodology. Suppose $(y_1, \ldots, y_n)$ are data from a GLM, so the distribution of y_i is in the exponential family with parameter θ_i and ϕ, with unknown functions $a_i(\phi)$, $b(\theta_i)$, $c(y_i, \phi)$ and with link function $g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$. Further assume $a(\phi) = a\phi$ for some constant s (not much of a restriction in practical modelling). Then the likelihood is:

$$L(\theta_1, \ldots, \theta_n; \phi) = \prod_{i=1}^{n} \exp\left[\frac{(y_i\theta_i - b(\theta_i))}{a\phi} + c(y_i, \phi)\right] \tag{5.1}$$

So the log-likelihood is:

$$\ell(\theta_1, \ldots, \theta_n; \phi) = \sum_{i=1}^{n} \left[\frac{(y_i\theta_i - b(\theta_i))}{a\phi} + c(y_i, \phi) \right] \tag{5.2}$$

To obtain MLEs for $(\beta_0, \ldots, \beta_p)$ we need to solve the system of simultaneous equations:

$$\frac{\partial \ell(\theta_1, \ldots, \theta_n; \phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} = 0 \quad \text{for} \quad j = 0, \ldots, p \tag{5.3}$$

where $\ell_i = \frac{(y_i\theta_i - b(\theta_i))}{a\phi} + c(y_i, \phi)$. By using the chain rule:

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i} \times \frac{\partial \theta_i}{\partial \mu_i} \times \frac{\partial \mu_i}{\partial \eta_i} \times \frac{\partial \eta_i}{\partial \beta_j} \tag{5.4}$$

Then, since for the exponential family $\mu_i = b'(\theta_i)$, we have:

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{(y_i - b'(\theta_i))}{a_i\phi} = \frac{(y_i - \mu_i)}{a\phi} \tag{5.5}$$

with $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$ and since this depends on μ_i via $b'(\theta_i)$, we can write $\frac{\partial \mu_i}{\partial \theta_i} = v(\mu_i)$ (often called the variance function of the model) or $\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{v(\mu_i)}$.

Also, since $g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$

$\frac{\partial \eta_i}{\partial \mu_i} = g'(\mu_i)$ or $\frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$

and $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$ (where x_{i0} is taken to be 1). So, putting this all together, the MLEs equations reduce to:

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{a\phi v(\mu_i) g'(\mu_i)} = 0 \quad \text{for} \qquad j = 0, \ldots, p \tag{5.6}$$

Equation 5.6 is a set of equations for $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$, since $\boldsymbol{\beta}$ determines the value of μ_i for $i = 1, \ldots, n$. Note that since $\phi \neq 0$, the solution $\widehat{\beta}$ in Equation 5.6 does not depend on knowledge of ϕ. Although the equation definitely depends on the specific model, a general numerical iterative solution can be derived via a Newton-Raphson approach which (after manipulation) gives the rth iteration as $\widehat{\boldsymbol{\beta}}_{(r)} = (\mathbf{X}'\mathbf{W}_{(\mathbf{r-1})}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_{(\mathbf{r-1})}\mathbf{z}_{(\mathbf{r-1})}$ where:

- $\mathbf{X}$ is the $n \times (p+1)$ 'design matrix' which (as in the normal theory linear model) has row vectors $(1, x_{i1}, \ldots, x_{ip})$.
- $\mathbf{z}_{(r-1)} = (z_1, \ldots, z_n)$ is a so called working vector with elements:

$$z_i = \eta_i + (y_i - \mu_i)\frac{\partial \eta_i}{\partial \mu_i} \tag{5.7}$$

  evaluated at $\boldsymbol{\beta}_{(r-1)}$ and at the data values.

- $\mathbf{W}_{(r-1)}$ is a diagonal weighting matrix with ith diagonal element:

$$w_{ii} = \frac{1}{av(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \tag{5.8}$$

  evaluated at $\boldsymbol{\beta}_{(r-1)}$ and at the data values.

At each stage, this looks similar to Gaussian least squares regression of the response vector $\mathbf{z}$ on the explanatory matrix $\mathbf{X}$, except this regression is 'weighted' by the diagonal elements of $\mathbf{W}$. Therefore, the model fitting method is often referred to as iterative re-weighted least squares (IRLS). IRLS algorithms are available in standard statistical computing software such as `R` (R Core Team, 2010). `R` provides a flexible implementation of the GLM framework in the function `glm` (Chambers and Hastie, 1992). Note also that with a normal error and identity link these GLM expressions for parameter estimates essentially collapse to the usual non-iterative results for the normal theory linear model.

Standard errors for estimates then follow from the standard general likelihood approach which gives:

$$var[\widehat{\boldsymbol{\beta}}] = \phi(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{5.9}$$

with $\mathbf{W}$ being evaluated at the final iteration of the parameter fitting process. We then obtain confidence intervals and associated hypothesis tests from $\frac{\beta_i - \widehat{\beta}_i}{\sqrt{var(\widehat{\beta}_i)}}$ being approximately distributed as $\mathbf{N}(0,1)$. If the scale parameter ϕ is unknown it must be estimated to obtain the previous results and one such estimate is given by:

$$\widehat{\phi} = \frac{1}{(n-p-1)}\sum_{i=1}^{n}\frac{(y_i - \widehat{\mu}_i)^2}{v(\widehat{\mu}_i)} \tag{5.10}$$

where $\widehat{\mu}_i = g^{-1}(\mathbf{x_i'}\widehat{\beta})$ is the predicted mean value of the *ith* response i.e. that predicted at explanatory variable values $\mathbf{x_i} = (1, x_{i1}, \ldots, x_{ip})'$ (Note that this is not the only estimate of ϕ which can be used). Using $\widehat{\phi}$ we have $\widehat{var}(\widehat{\boldsymbol{\beta}}) = \widehat{\phi}(\mathbf{X'WX})^{-1}$ and $\frac{\beta_i - \widehat{\beta}_i}{\sqrt{\widehat{var}(\widehat{\beta}_i)}}$ is then approximately distributed as t_{n-p-1}. Note that with a normal error and identity link these GLM expressions for parameter standard errors and confidence intervals essentially collapse to the usual results for the normal theory linear model.

In testing goodness-of-fit GLM M, the general likelihood ratio statistic $\wedge = \frac{L_M}{L_{M_s}}$ (where M_s is the saturated model) is given by:

$$\begin{aligned} \wedge &= \frac{\exp \sum_{i=1}^{n} \left[\frac{(y_i\widehat{\theta}_i - b(\widehat{\theta}_i))}{a\phi} + c(y_i, \phi) \right]}{\exp \sum_{i=1}^{n} \left[\frac{(y_i\widetilde{\theta}_i - b(\widetilde{\theta}_i))}{a\phi} + c(y_i, \phi) \right]} \\ &= \exp \sum_{i=1}^{n} \left[\left(y_i(\widehat{\theta}_i - \widetilde{\theta}_i) - b(\widehat{\theta}_i) + b(\widetilde{\theta}_i) \right) / a\phi \right] \end{aligned} \tag{5.11}$$

where $\widehat{\theta}_i$ denote MLEs estimates under the model M and $\widetilde{\theta}_i$ denote MLEs estimates under the saturated model M_s (i.e the model in which $\widehat{\mu}_i = y_i$). So, the log likelihood ratio statistic is:

$$-2 \log \wedge = 2 \sum_{i=1}^{n} \left[\left(y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - b(\widetilde{\theta}_i) + b(\widehat{\theta}_i) \right) / \phi \right] \tag{5.12}$$

Thus in the case when $a = 1$ (true for all models we will be concerned with) $-2 \log \wedge$ can be written as $\frac{D_M}{\phi}$ where:

$$D_M = 2 \sum_{i=1}^{n} \left[y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - b(\widetilde{\theta}_i) + b(\widehat{\theta}_i) \right] \tag{5.13}$$

D_M is known as the deviance of the model M and $\frac{D_M}{\phi}$ is known as the scaled deviance. Note D_M depends on the data $\mathbf{y}$ and the estimated parameters $\widehat{\boldsymbol{\beta}}$ in the linear predictor but it does not depend on ϕ. General likelihood theory tells us that the scaled deviance $-2 \log \wedge$ has an asymptotic χ^2_{n-p-1} distribution with expected value $n - p - 1$ under the hypothesis that there is no significance difference in fit between the model M and the saturated model M_s (Note that here p is the number of explanatory variables so the number of parameters is $p+1$).

It follows that when ϕ is known, the goodness-of-fit of a GLM can be assessed by comparing the value of the scaled deviance of the model to the χ^2_{n-p-1} distribution (the scaled deviance should roughly equal to $n-p-1$ for an edequate model). Note where ϕ is unknown the above suggests that a sensible estimate of ϕ is given by $\widehat{\phi} = \frac{D_M}{n-p-1}$ since this will 'make the model fit' (an alternative estimate for ϕ to that suggested previously). A test of differences in fit between nested GLM models, M_1 and M_2 with $a_i(\phi) = \phi$ and numbers of parameters $p_1 + 1 < p_2 + 1$ is also based on log-likelihood ratios. In this case, the relevant log-likelihood statistic is the difference in scaled deviances i.e $\frac{D_{M_1} - D_{M_2}}{\phi}$. General likelihood theory states that if ϕ is known, this difference is approximately $\chi^2_{p_2 - p_1}$ distributed if there is no difference in model fit. On the other hand, if ϕ is unknown, it will need to be replaced by an estimate and the 'best' such estimate is clearly associated with the model with the more parameters i.e. $\widehat{\phi} = \frac{D_{M_1}}{n-p_2-1}$. So we thus plug in $\widehat{\phi}$ for ϕ and we then slightly modify the likelihood ratio statistic for comparing M_1 and M_2 to : $\frac{(D_{M_1} - D_{M_2})/(p_2 - p_1)}{D_{M_2}/(n-p_2-1)}$. This for a GLM, has an approximate $\mathcal{F}_{p_2-p_1, n-p_2-1}$ distribution under no difference in model fit. We can also (more crudely) compare the fit of two models M_1, M_2 by comparing their Aikaike Information Criterion (AIC) values which penalises their fits by the number of parameters used (Akaike, 1973; Sakamoto et al., 1988) i.e. compare $\frac{D_{M_1}}{\phi} + 2(p_1 + 1)$ and $\frac{D_{M_2}}{\phi} + 2(p_2 + 1)$ where the lower AIC is better. The Bayesian Schwartz Information Criteria (BIC), defined as in Equation 5.14 below (Schwartz, 1978) is an alternative comparison criterion between models which penalises for the number of parameters in the model. Similiar to AIC, the lower the BIC value, the better the model is.

$$BIC = -2\ell(\widehat{\mu}; y) + p\log(n) \tag{5.14}$$

Another commonly used descriptive measure of model fit is the pseudo-R^2 value which simply compares the log-likelihood from the null model (contains only an intercept) to the log-likelihood from the fitted model (Breslow, 1996) i.e.

$$R^2 = 1 - \frac{D_M}{D_{M_{null}}} \tag{5.15}$$

R^2 values closer to 1 implying that the model is a better fit to the data. An adjusted pseudo-R_a^2 can also be defined which adjusts for the number of explanatory variables in the model where:

$$R_a^2 = \frac{n-1}{n-p-1} R^2 \tag{5.16}$$

where n is the number of data points, and p is the number of covariates in the model.

Prediction of values from a GLM is similar to that used for the normal theory linear model. Prediction of a response y by its fitted value (predicted mean value), so $\widehat{y_i} = \widehat{\mu_i}$. However, in a GLM the mean μ_i is actually a function of the linear predictor, η_i. Therefore, the prediction for the response at explanatory variable values $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})'$ is $\widehat{y_i} = g^{-1}\left(\mathbf{x_i'}\widehat{\beta}\right)$. It may be shown that the variance of a GLM fitted value $\widehat{y_i}$ is given by $var[\widehat{y_i}] = h_{ii} var[y_i] = \sigma_i^2$ where h_{ii} is the ith diagonal element of the matrix : $\mathbf{H} = \mathbf{W^{\frac{1}{2}}X(X'WX)^{-1}X'W^{\frac{1}{2}}}$ which is the GLM equivalent of the normal theory linear model 'hat matrix' $\mathbf{H} = \mathbf{X(X'X)^{-1}X'}$. Recall that for a GLM, $var[y_i] = \sigma_i^2 = a_i \phi v(\mu_i)$, where $v()$ is the variable function which we can estimate by using $\widehat{\sigma}_i^2 = a_i \widehat{\phi} v(\widehat{\mu_i})$. Finally, the estimated variance of $\widehat{y_i}$ as $var[\widehat{y_i}] = h_{ii}\widehat{\sigma}_i^2$ is obtained.

As for the Gaussian linear model, residuals form the basis for model checking for the GLM. However, various other different kinds of residuals can be defined for a GLM as well as the 'raw' residuals $\widehat{\varepsilon}_i^{(p)} = (y_i - \widehat{y_i})$. In particular, the Pearson residuals are the standardised version of the raw residuals and derived as:

$$\widehat{\varepsilon}_i^{(p)} = \frac{\widehat{\varepsilon}_i}{\sqrt{\widehat{var[\varepsilon_i]}}} = \frac{(y_i - \widehat{y}_i)}{\sqrt{[(1-h_{ii})\widehat{\sigma}_i^2]}} \tag{5.17}$$

where the expression for $\widehat{var[\varepsilon_i]}$ follows from earlier one for $var[\widehat{y_i}]$. A further alternative is the deviance residuals which can be described as the square root of an individual observation's contribution to the deviance with the sign of $(y_i - \widehat{y}_i)$ attached i.e.: $\mathrm{sgn}(y_i - \widehat{y}_i)\sqrt{D_i}$ where $D_i = 2\left[y_i(\widetilde{\theta}_i - \widehat{\theta}_i) - b(\widetilde{\theta}_i) + b(\widehat{\theta}_i)\right]$ with $\widehat{\theta}_i$ and $\widetilde{\theta}_i$ referring to estimated values under the model and the saturated model respectively, or more particularly the associated standardised deviance residuals:

$$\widehat{\varepsilon}_i^{(d)} = \frac{\mathrm{sgn}(y_i - \widehat{y}_i)\sqrt{D_i}}{\sqrt{(1 - h_{ii})}} \tag{5.18}$$

There are other possible types of residuals, but the Pearson and standardised deviance residuals tend to be the most useful for model diagnostic purposes and should approximately follow an $\mathrm{N}(0, 1)$ distribution if the distributional assumptions in the original GLM are valid. In particular standardised deviance residuals pin-point any observations that give a disproportionately large contribution to the deviance (leverage and influence).

So turning to the modelling of observed dengue counts, y_i, the particular version of the GLM that may perhaps be considered first is a Poisson GLM with a log link and a population offset which can be written as:

$$\begin{aligned} y_i \sim P[\mu_i] &= P[p_i \rho_i] \qquad i = 1, \ldots, n \\ \log \mu_i &= \log p_i + \log \rho_i = \log p_i + \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} \end{aligned} \tag{5.19}$$

where $P[\mu_i]$ denotes the Poisson distribution with probability mass function:

$$p(y_i; \mu_i) = \frac{\exp^{-\mu_i} \mu_i^{y_i}}{y_i!} \qquad y_i = 0, 1, 2, \ldots \tag{5.20}$$

and where ρ_i denotes dengue incidence rate, x_{ji}, $j = 1, \ldots, p$, are suitably chosen covariates and $\log p_i$ is an offset included to account for the different (known) population sizes in each area i.

However, there are well known possible problems with using such a Poisson GLM to model disease counts, a key issue being overdispersion. Overdispersion is the commonly encountered situation where the variance of observed counts exceeds the mean whereas the Poisson distribution implies equality of mean and variance.

Inappropriate use of the Poisson assumption when overdispersion is present may underestimate the standard errors and overstate the significance of the model parameters and consequently, give misleading inference about these parameters.

We could change the model to accommodate overdispersion by switching to a negative binomial model (Simon et al., 2003). A negative binomial GLM is also suitable for modelling count data, but does not assume the mean is equal to the variance. It can be considered as a generalisation of the Poisson GLM since it has the same mean structure but an extra parameter to model the overdispersion (Breslow, 1984). The negative binomial is a common choice in modelling disease counts in epidemiological applications, Richard and John (2007) similarly advocated use of the negative binomial distribution in criminology applications when there is evidence of overdispersion. Research by Osgood (2000) has also suggested using a negative binomial distribution in such circumstances. Some studies have also adopted the negative binomial GLM in modelling dengue counts such as Simões et al. (2013), Markon (2014) and Ahmed et al. (2015).

Accordingly, the previous Poisson model becomes:

$$\begin{aligned} y_i \sim NB[\mu_i, \theta] &= NB[p_i \rho_i, \theta] \qquad i = 1, \ldots, n \\ \log \mu_i &= \log p_i + \log \rho_i = \log p_i + \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} \end{aligned} \tag{5.21}$$

where $NB[\mu_i, \theta]$ denotes the negative binomial distribution with probability mass function:

$$p(y_i; \mu_i, \theta) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) y_i!} \frac{\mu_i^{y_i} \theta^{\theta}}{(\mu_i + \theta)^{y_i + \theta}} \tag{5.22}$$

where μ_i is the mean and θ is the scale parameter. The variance of this distribution is given by $\mu_i + \frac{\mu_i^2}{\theta}$, hence θ can be used to accommodate overdispersion.

However as seen in Chapter 4, there are possibilities of non-linear relations with explanatory variables in modelling dengue incidence rate (e.g annual trends and seasonal cycles). Therefore this study will extend the negative binomial GLM to that of a negative binomial generalised additive model (GAM). The GAM (Hastie and Tibshirani, 1990) extends the GLM by allowing the linear predictor to include unknown smooth non-parametric functions of one or more of the explanatory variables (e.g. $\beta_0 + f_1(x_{i1}) + \ldots + f_p(x_{ip})$). The GAM model can be fitted by iteratively fitting weighted additive models by localised regression or smoothing splines in a analogous way as the iteratively weighted least squares procedure relates to ordinary least squares (Simon, 2006). The main strength in GAM is the ability to deal with highly non-linear and non-monotonic relationships between the response and the set of explanatory variables (Hastie and Tibshirani, 1986). The ability of GAM to handle non-linear data structures has encouraged its use in ecological models such as those that will be developed for dengue later in this chapter (e.g. Thomas and Neil, 1991 and Cheong et al., 2013).

Algorithms to fit GAM are available in off-the-shelf statistical software, such as `R` (R Core Team, 2010). Cubic splines or Thin Plate Splines are commonly used to estimate the smooth functions. Two techniques may be used to estimate associated smoothing parameters (Craven and Wahba, 1979 and Wahba, 1990) namely generalised cross validation (GCV) or Un-Biased Risk Estimation (UBRE) which are defined as follows:

$$GCV = n\frac{D_M}{n - edf} \tag{5.23}$$

$$UBRE = \frac{D_M}{n} + 2\theta\frac{edf}{n - \theta} \tag{5.24}$$

where n is the number of observations, D_M is the deviance of the model, θ is the scale parameter and *edf* is the effective degrees of freedom of the model. The function `gam` in the `mgcv` package in `R`, handles the basic fitting of GAM models (Hastie and Tibshirani, 1986) using an easy to use model specification interface where smooth functions can be used on their own or mixed with parametric functions as shown in the following expressions.

$$s(a) + s(b) + s(c) \tag{5.25a}$$

$$a + s(b) + c \tag{5.25b}$$

Expression 5.25a has smooth functions for all three of its continuous explanatory variables a, b and c using the smoother (s), while expression 5.25b fits parametric terms for a and c and a non-parametric smooth function for b.

So, when and if required in subsequent sections of this chapter, we may extend the previously suggested negative binomial model for dengue incidence to:

$$\begin{aligned} y_i \sim NB[\mu_i, \theta] &= NB[p_i \rho_i, \theta] \qquad i = 1, \ldots, n \\ \log \mu_i &= \log p_i + \log \rho_i = \log p_i + \beta_0 + \sum_{j=1}^{q_1} \beta_j x_{ji} + \sum_{j=1}^{q_2} f_j(x_{ji}) \end{aligned} \tag{5.26}$$

where the previous p is now $q_1 + q_2$ and $f_j(x_{ji})$ are unknown smooth functions of a subset of the explanatory variables.

5.2 Covariate Selection

The exploratory analyses in Chapter 4 have suggested dengue incidence in Malaysia may be related to factors such as population, population density, sea surface temperature (referred to as Niño 4), average rainfall and temperature, and number of rainy days as well as an annual seasonal cycle in dengue counts and influences of monsoon and regional or state differences. These then are the set of variables that will be further investigated more formally in this section using the modelling frameworks introduced in the previous section. The objective is to select a specific 'best subset' of covariates from those discussed in Chapter 4. This is not a straightforward automatic process. The volume of data implies that model coefficients will often be formally statistically significant as a result of the sheer number of observations rather than because of more substantive reasons. Multicollinearity between variables will also be present. The role of ENSO may be obscured either by a local climate heterogeneity, insufficient data or randomly coincident outbreaks as discussed by Johansson et al. (2009a). It will be important to explore time lags for the climate variables as highlighted by Cuadras and Fortiana (2002), for example where heavy rainfall in a preceding month may cause an increased supply of standing water sources, mosquito breeding sites which then in the following months increase the dengue fever risk. Interactions between some variables may also be important, particularly climatic variables, due to the coupled nature of the dynamical processes involved. Non-linear relationships may also need to be investigated.

So a very large number of models will need to be compared and the final variable selection will need to balance parsimony and pragmatism with formal considerations of coefficient significance, AIC and automatic stepwise procedures. The approach adopted in this section will be to focus on the negative binomial model introduced in the previous section. This is because this framework will protect against the likely presence of overdispersion whilst being more computationally efficient than the GAM for the purposes of variable selection. Once variables have been selected in this section then the issue of overdispersion and use of smooth functions for some variables will be followed up in the subsequent section. So if y_{st} denotes the observed dengue counts for state s $(s = 1, \ldots, 12)$ and month t $(t = 1, \ldots, 108)$ (recall there are 12 states and 108 monthly observations) then considering these counts to be negative binomial distributed we will use a GLM of the general form:

$$\begin{aligned} y_{st} &\sim \text{NegBin}(\mu_{st} = p_{st}\rho_{st}, \theta) \\ \log \mu_{st} &= \log(p_{st}) + \log(\rho_{st}) = \log(p_{st}) + \alpha + \sum_{j=1}^{p} \beta_j x_{ji} \end{aligned} \tag{5.27}$$

where the expected number of dengue cases, μ_{st}, are given by the population p_{st} multiplied by the unknown relative dengue risk, ρ_{st} for a given state, s and month, t. Models involving all available explanatory variables and subsets of them were explored. These included climate covariates rainfall, number of rainy days, temperature, sea surface temperature (SST) and lagged values of these variables from the current month up to lag of 6 months. Then the population offset and population density as well as a general global trend and factors to represent monthly seasonal effects and regional, state or monsoon influences. Non-linearity in some of the relationships was explored by inclusion of low order polynomial terms in the relevant variable. Interactions between relevant variables were also tested. Finally lagged values of the logarithm of DIR were included to allow for the dynamic nature of the disease.

A very large number of models was compared to determine the final variable selection which is listed in Table 5.1. There, for the sake of clarity the general $\beta_j x_{jst}$ terms in the Model 5.27 have been broken down into three groups. First, the selected climate terms, $\beta_j x_{jst}$, which are respectively average rainfall in the same month and at a lag of 3 months, number of rainy days in the same month and at a lag of 3 months, average temperature in the same month and lag 1 month, sea surface temperature lag 6 months and interaction between temperature lag 1 month and sea surface temperature lag 6 months. Second, terms $\gamma_j z_{jst}$ $(j = 1, \ldots, n)$ which relate to population density, year (for global trend), a factor month (for seasonal cycle) and log dengue incidence rate lagged 3 months. Finally, a factor $\delta_{r'(s)}$ representing a regional effect with $r'(s)$ being an indicator function mapping each state, s into one of the four regions.

Specific inference concerning the various influences of the variables selected is followed up in detail in the next section, but some brief summary comments are useful at this point. This overall increase is superimposed on an annual seasonal cycle which sees DIR peaks in January and July. DIR is higher in the areas where there is a higher population density. This was particularly marked in those areas in the South West of the country where the main urban areas of Malaysia are located (Kuala Lumpur and Selangor states). However, in line with the views of Muhammad Azami et al. (2011), it is not the case that dengue in Malaysia is mainly restricted to urban areas - where there are similar seroprevalence rates between urban and rural areas dengue is present in both.

The variable selection indicated that geographical differences can be adequately captured without significant loss of detail by grouping the twelve states into the four broad regions of North East, South East, North West and South West. That said the state factor did have some residual explanatory power as did the monsoon factor and these factors will be further investigated in the subsequent section. The effects of the climate variables broadly reflect those found in other dengue studies, but the interaction between ENSO and average temperature is an interesting and unexpected finding.

5.3 Model Comparison and Development

Having determined a base set of selected variables in the previous section, this section is concerned with comparing and developing models for dengue incidence in Malaysia using these variables along with the modelling frameworks described in Section 5.1.

We start by formally considering overdispersion. Variable selection in the previous section was carried out using a negative binomial formulation as opposed to a Poisson on the basis that overdispersion was likely and therefore potentially needed to be allowed for in variable selection. Having selected the variables it is now appropriate to formally establish that overdispersion is indeed present and that it is necessary to continue with the negative binomial formulation rather than be able to adopt the simpler Poisson case. In order to do that we fitted a Poisson GLM using the full set of covariates identified in Table 5.1. Table 5.2 shows a summary of statistics of model fit in terms of likelihood L, deviance, D, Aikaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) between a Poisson GLM using these variables and a negative binomial GLM. The value of the negative log likelihood, L of the Poisson GLM is 7 times higher than negative binomial GLM, whilst deviance scores are 64 times higher and AIC and BIC values also 7 times higher. These results clearly indicate very strong evidence to adopt a negative binomial formulation as opposed to Poisson in modelling DIR in Malaysia.

Having established the need for a negative binomial formulation as opposed to a Poisson, Table 5.3 provides detailed estimation results for the negative binomial model using the covariates in Table 5.1. Note that for conciseness the numerous factor effects for month and region are not reported in this table. The baseline in this model (included in the intercept) is the North East region and the month of

Table 5.1: *Final Selected Variables from negative binomial GLM runs.*

Coefficient	Covariates
α	Intercept
β_1	Rainfall
β_2	Rainfall Lag 3 Months
β_3	No.Rainy Days
β_4	No.Rainy Days Lag 3 Months
β_5	Temperature
β_6	Temperature Lag 1 Month
β_7	Sea Surface Temperature Lag 6 Months
β_{67}	Temperature Lag 1 Month and Sea Surface Temperature Lag 6 Months
γ_1	Population Density
γ_2	Year
γ_3	Month factor
γ_4	Log Dengue Incidence Rate Lag 3 Months
$\delta_{r'(s)}$	Region Mapping To Each State

Table 5.2: *Likelihood statistic (L), degrees of freedom $(n-p)$, Deviance (D), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for models with similar subsets of covariates fitted using Poisson and negative binomial GLMs.*

Statistics Test	Poisson	Negative Binomial
L	-49117.26	-7463.944
$n-p$	1282	1281
D	89632.34	1381.785
AIC	98266.53	14959.89
BIC	98349.13	15042.49

January the regional and monthly effects (not reported) then adjust this intercept but the other effects remain the same relative to the intercept.

Having justified and established a baseline negative binomial GLM for dengue incidence in Malaysia, the next step in model development is to investigate whether this model can be improved by moving to the negative binomial GAM framework introduced in Section 5.1. For example, the exploratory analyses in Chapter 4 indicated the seasonal cycle in dengue to be far from simple and possibly region specific, suggesting that it may be preferable to represent this by a non-parametric region specific smooth function, rather than by a monthly factor (Aziz et al., 2012). Accordingly the previous negative binomial GLM was extended to a negative binomial GAM with the dengue incidence rate ρ_{st} modelled as in Equation 5.28, with other aspects of the model remaining as for the negative binomial GLM i.e. the observed dengue counts, y_{st}, for state s $(s = 1, \dots, 12)$ and month t $(t = 1, \dots, 108)$ are assumed to follow a negative binomial distribution with mean value $\mu_{st} = p_{st}\rho_{st}$ and scale parameter θ, where p_{st} is the known population offset.

$$\begin{aligned}\log \rho_{st} &= \alpha + \textstyle\sum_j^7 \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + \gamma_2 z_{2st} + \gamma_3 z_{3st} \\ &\qquad + f_{r'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \end{aligned} \tag{5.28}$$

The GAM Model 5.28 is essentially the same as the negative binomial GLM except that the seasonal cycle is now represented by a smooth function of calendar month

Table 5.3: *Estimates of coefficients, standard errors and p-values of negative binomial GLM model.*

Variables	Coefficient	Estimate	Std. err	Z-value	Pr(>Z)
Intercept	α	8.479e+01	2.742e+01	3.092	< 0.01
Rainfall	β_1	-3.407e-04	1.330e-04	-2.561	< 0.05
Rainfall Lag 3 Months	β_2	2.364e-04	1.099e-04	2.151	< 0.05
Rainy Days	β_3	1.383e-02	4.514e-03	3.064	< 0.01
Rainy Days Lag 3 Month	β_4	1.263e-03	4.094e-03	0.308	> 0.1
Temperature	β_5	2.302e-02	3.670e-02	0.627	> 0.1
Temp. Lag 1 Month	β_6	-3.225e+00	1.014e+00	-3.180	< 0.01
Nino4 Lag 6 Months	β_7	-2.891e+00	9.552e-01	-3.026	< 0.01
Temp. Lag 1 and SST Lag 6 Months	β_{67}	1.114e-01	3.530e-02	3.157	< 0.01
Population Density	γ_1	1.282e-04	1.412e-05	9.076	< 0.001
Year	γ_2	5.002e-02	8.077e-03	6.192	< 0.001
DIR Lag 3 Months	γ_4	7.800e-03	2.069e-03	3.770	< 0.001
Region North West	δ_2	-2.217e-02	6.059e-02	-0.366	> 0.1
Region South East	δ_3	-6.922e-02	6.783e-02	-1.021	> 0.1
Region South West	δ_4	4.535e-01	6.846e-02	6.624	< 0.001

with the interaction between this and region.

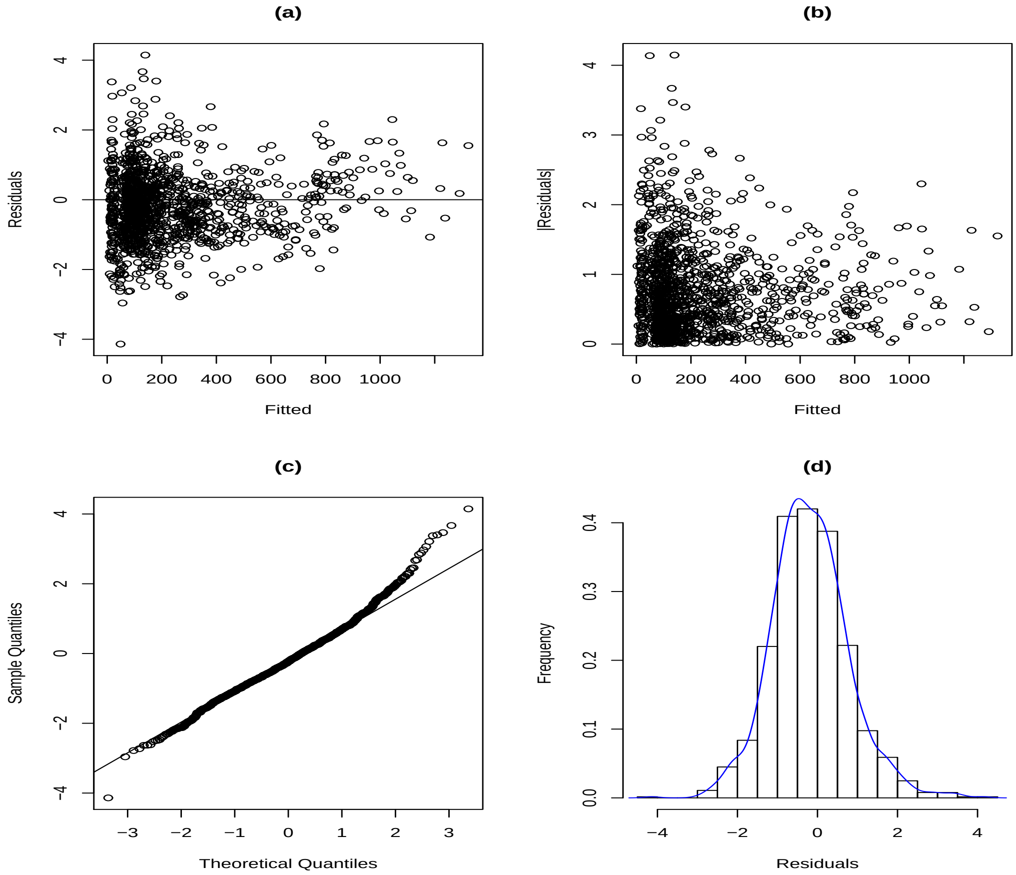

Figure 5.1: *(a) Residuals vs. fitted plot, (b) Absolute value of Residuals vs. fitted plot, (c) Residuals vs. Theoretical Quantiles plot and (d) Frequency vs. Residuals plot.*

Figure 5.1 shows four plots of residuals from Model 5.28. Figures 5.1(a) and (b) show little evidence of non-constant variance. Figure 5.1(c) shows the upper tail deviates somewhat from the straight line but Figure 5.1(d) shows the expected bell shape, given that the deviance residuals should be normally distributed. Overall these residuals are broadly acceptable. Figure 5.2 shows the smooth function for the seasonal cycle for each region (as fitted within the model using the UBRE criterion). The cycles are somewhat different in each of the regions, but each shows clear cyclic behaviour with two peak points broadly falling in July and January each year.

Table 5.4 compares the fits of the previous negative binomial GLM and the GAM.

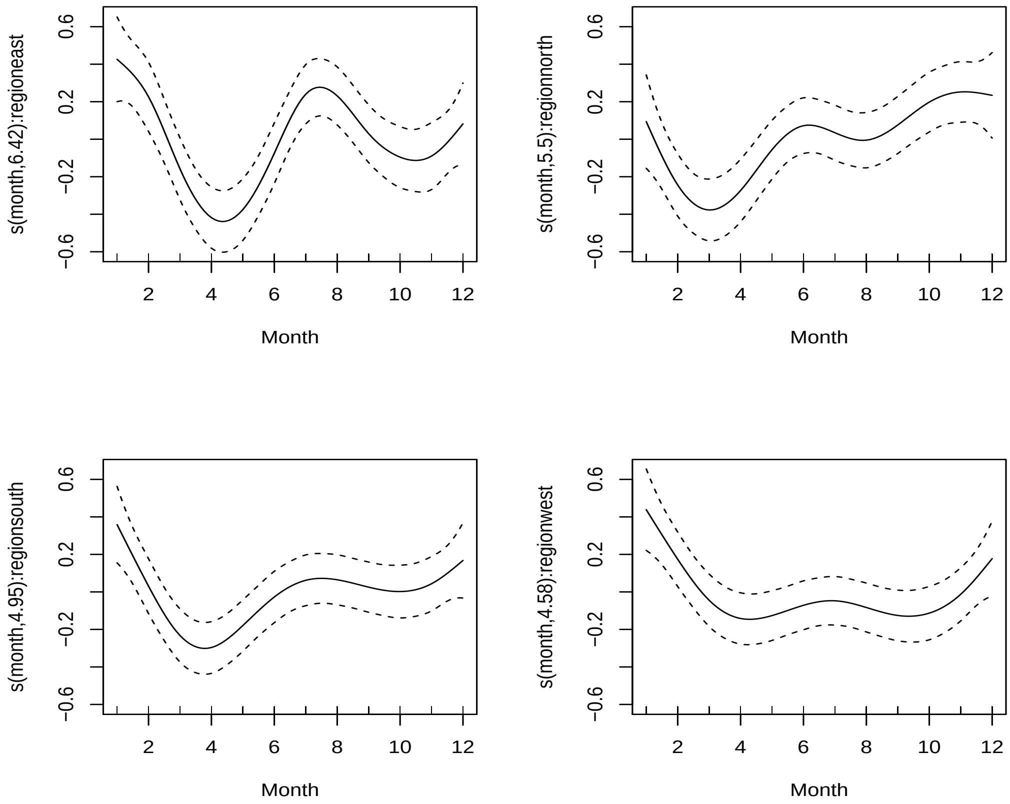

Figure 5.2: *Smooth annual cycle functions of four regions in Malaysia; upper row from left to right are North East and North West region, and lower row from left to right are South East and South West region. In each panel, the solid line is the estimate, and the dashed line is the confidence interval.*

Table 5.4: *Comparing the negative binomial GLM and GAM: ANOVA results.*

Model	Resid. Df	LogLik	Diff Resid. Df	Deviance	Pr(>ChiSq)
GLM	1274.0	-7463.944			
GAM	1253.6	-7398.473	20.439	130.94	2.2e-16

The likelihood ratio test confirms that there is a significant difference in fit with the GAM fitting better.

It is interesting to explore the contribution of climatic variables in the GAM model with that of the non-climatic variables. According two sub-models were defined as follows:

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_6 x_{6st} \beta_7 x_{7st} \tag{5.29a}$$

$$\log \rho_{st} = \alpha + \gamma_1 z_{1st} + \gamma_2 z_{2st} + \gamma_3 z_{3st} + f_{r'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.29b}$$

The climate Model 5.29a only includes the climate covariates; x_{jst} with $j = 1, \ldots, 7$ which are rainfall in the same month and lag 3 months, number of rainy days in the same month and lag 3 months, temperatures in the same month and lag 1 month and ONI lag 6 months together with interaction of climate covariates, $\beta_{67} x_{6st} x_{7st}$ (ONI lag 6 months and temperature lag 1 month). On the other hand, the non-climate Model 5.29b comprises just the non-climate variables i.e. z_{1st} as population density, z_{2st} as year (considered 2001 to 2009), DIR lag 3 months z_{3st}, region specific smooth functions of month $f_{r'(s)}\left(z_{4st}\right)$ and the region factor $\delta_{r'(s)}$.

Table 5.5 compares the fit of the climate model, the non-climate model and full combined model. The results show that by including climate covariates alone the model fit explains 0.4% of the deviance, whereas the non-climate covariates alone explain 8.7% of the deviance, the full model combining the two sets of covariates results in an additional 5% of the deviance being explained. This implies that although the climate effects are significant in the dengue model their explanatory power is relatively weak in the absence of the other non-climate influences.

The parameter estimates, standard errors and p-values for the parametric terms in the full model are presented in Table 5.6. Note that, for conciseness the factor effects for region are not reported in this table but the baseline in this model (included in the intercept) is the North East region the regional effects (not reported) then adjust this intercept but the other effects remain the same relative to the intercept.

Table 5.5: *Deviance, pseudo-R_D^2, number of parameter(p), degrees of freedom $(n-p)$, AIC and BIC for models with different subsets of covariates fitted using the negative binomial GAM.*

Model	Deviance	R_D^2	p	$n-p$	AIC	BIC
Climate model	1441.435	0.0046	9	1287	16518.72	16565.19
Non-Climate model	1322.690	0.0866	7	1289	14922.42	15060.84
Combined model	1250.843	0.1362	15	1281	14869.82	15057.94

As may be seen from Table 5.6 the mean rainfall 3 months previously (β_2) has a positive relationship with DIR, but mean rainfall in the same month (β_1) has a negative relationship with DIR. This could possibly be because more rainfall earlier in the year could encourage mosquito development, while heavy rainfall in the same month could wash out mosquito breeding places and lower dengue transmission (Hii et al., 2012).

The number of rainy days both 3 months previously (β_4) and in the same month (β_3) and temperature in the same month (β_5) all have a positive relationship with DIR, while temperature at lag 1 month (β_6) and sea surface temperature (SST) 6 months previously (β_7) have a negative relationship. However, the latter must be seen in conjunction with the interaction between sea surface temperature (SST) 6 months previously and lag 1 month temperature (β_{67}) which has a positive relationship with DIR. The population density (γ_1), global trend (γ_2) and DIR at lag 3 months (γ_3) all have a positive relationship with DIR. Also recall that this model also includes the factor reflecting region and the smooth function of month (by region) and these terms allow the baseline of the model to vary depending on which region and calendar month is of interest.

Further refinements to the GAM Model 5.28 were then extensively explored. This included replacing the parametric climate terms with smooth functions, however this did not improve the fit of the model for any of these variables. Also replacing the global trend term with a region specific smooth function which did improve

Table 5.6: *Parameter estimates, standard errors and p-values for negative binomial GAM Model 5.28.*

Variables	Coefficient	Estimate	Std. err	Z-value	Pr(>Z)
Intercept	α	4.110e+01	2.476e+01	1.660	< 0.1
Rainfall	β_1	-3.201e-04	1.133e-04	-2.826	< 0.01
Rainfall Lag 3 Months	β_2	2.149e-04	9.507e-05	2.261	< 0.05
Rainy Days	β_3	1.846e-02	4.607e-03	4.008	< 0.001
Rainy Days Lag 3 Months	β_4	8.081e-03	4.195e-03	1.927	< 0.1
Temperature	β_5	1.184e-01	3.472e-02	3.410	< 0.001
Temp. Lag 1 Month	β_6	-1.661e+00	9.147e-01	-1.815	< 0.1
Nino4 Lag 6 Months	β_7	-1.530e+00	8.577e-01	-1.784	< 0.1
Temp. Lag 1 and SST Lag 6 Months	β_{67}	5.927e-02	3.176e-02	1.866	< 0.1
Population Density	γ_1	1.256e-04	1.213e-05	10.361	< 0.001
Year	γ_2	5.505e-02	7.003e-03	7.861	< 0.001
DIR Lag 3 Months	γ_3	8.198e-03	1.792e-03	4.575	< 0.001
Region North West	δ_2	-1.163e-01	5.445e-02	-2.136	< 0.05
Region South East	δ_3	-1.305e-01	5.931e-02	-2.201	< 0.05
Region South West	δ_4	3.772e-01	6.169e-02	6.114	< 0.001

the fit. Then experimenting with replacing the region specific seasonal cycles with similar terms but split not by the four regions but rather by a two level monsoon factor delineating states primarily affected by one or other of the two monsoons. Also replacing region specific cycles with similar terms but split by a 12 level states factor delineating all the separate states rather than simply four regions. These latter changes did produce some interesting results, so we now proceed to consider and compare three refined GAM models as follows:

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{r'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.30a}$$

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{m'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.30b}$$

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{s}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.30c}$$

where $m'(s)$ denotes a function mapping states to monsoon type. Table 5.7 presents summary statistics of the fits of Models 5.30a (A), 5.30b (B) and 5.30c (C). Model A presents the lowest AIC, but more comprehensive analyses of the overall fit of these three models will be reserved until the next chapter; here we simply concentrate on reporting the results of parameter estimates and associated smooth functions.

Figure 5.3 shows the smooth functions of month by region and smooth function of year by region for Model A. The first four plots shows the peak months of each region (similar to that in the earlier Figure 5.2) while the other four plots show the trend of DIR over years which highlights the epidemics of dengue in 2002 and 2008. Figure 5.4 shows smooth functions for Model B i.e. a smooth function of month by monsoon area (the Northeast monsoon area is a combination of the North East and South East regions and the Southwest monsoon area is a combination of the North West and South West regions). Both plots have a similar shape and show

Table 5.7: *The deviance (D), log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Un-Biased Risk Estimator (UBRE) for Model A, Model B and Model C conditioned by region, monsoon and state respectively using negative binomial GAM.*

Model	Deviance	LogLik	AIC	BIC	UBRE
Model A (by region)	1374.425	-6859.32	13833.42	14132.49	0.2174
Model B (by monsoon)	1153.703	-6879.90	13850.52	14087.34	0.0160
Model C (by state)	1249.592	-6838.15	13856.68	14237.55	0.1683

the two monsoon areas having similar DIR peaks in January and July. Figure 5.4 also shows the smooth functions for global trend in each of the four regions which are broadly similar to their equivalent in Figure 5.3 revealing some differences in the global trend of DIR between the regions particularly in the North West region where an epidemics of DIR started in 2004 some two years later than in the other three regions.

Finally, Figure 5.5 shows the smooth functions of month and year for Model C i.e. for month for each of the 12 states and for year for each of the four regions. Tables 5.8, 5.9 and 5.10 give the parameter estimates, standard errors and p-values for the parametric terms in each of Models A, B and C respectively.

In order to facilitate comparison between the estimated coefficients in the different models, Figure 5.6 graphically presents coefficient values and associated 95% confidence intervals for the most significant parameters in Models A, B and C. It is notable that the direction of effects for all coefficients in the three models is broadly similar. Average rainfall in the current month has a negative relationship on DIR (biting behaviour of mosquitoes?) whereas that 3 months previously has a positive relationship (mosquito breeding?). Number of rainy days (a surrogate for rainfall intensity) has a negative relationship with DIR at 3 month lag ('washing' of mosquito larvae?) and a positive relationship with DIR in the current month (more intensive biting in dry periods?). Temperature in the current month shows

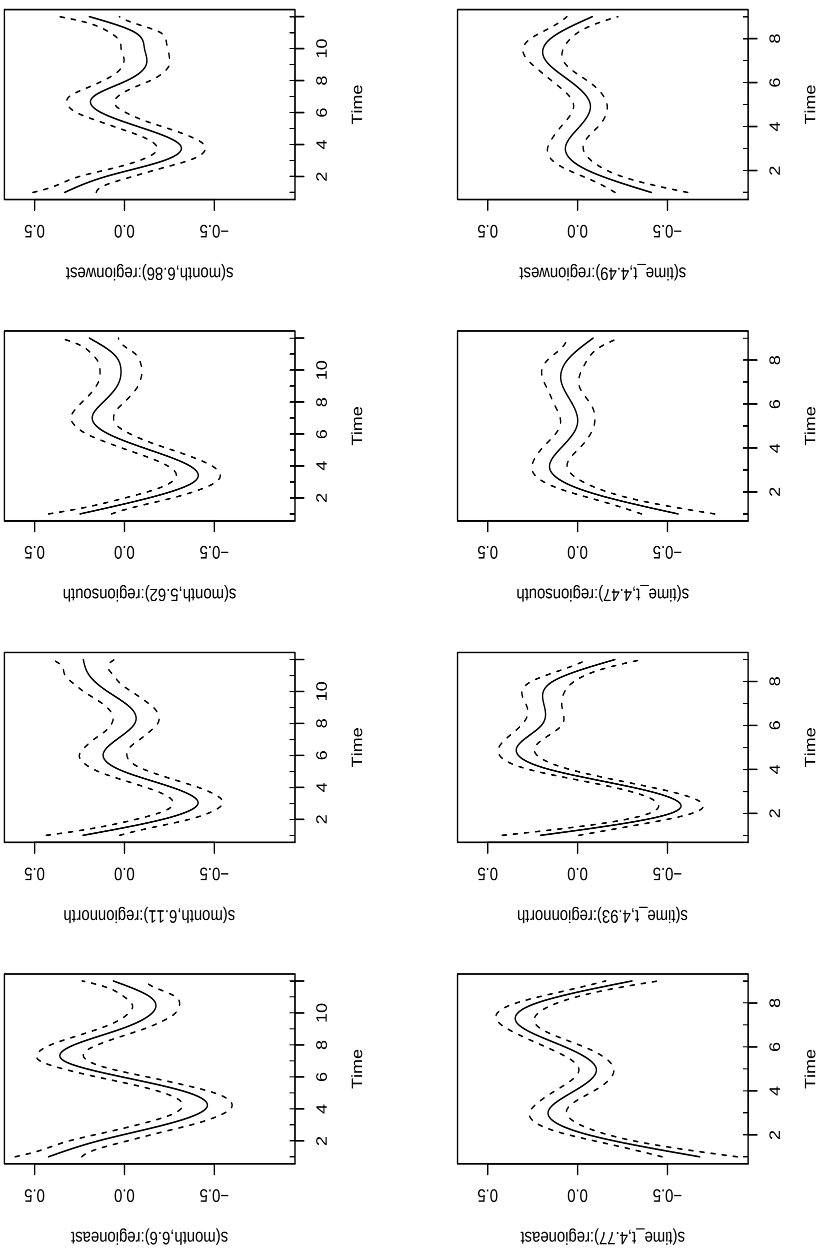

Figure 5.3: *Smooth functions in Model A. Upper row from left to right are North East, North West, South East and South West region for smooth function of month (by region). Lower row from left to right are North East, North West, South East and South West region for smooth function of year (by region). In each panel, the solid line is the estimate, and the dashed line is the confidence interval.*

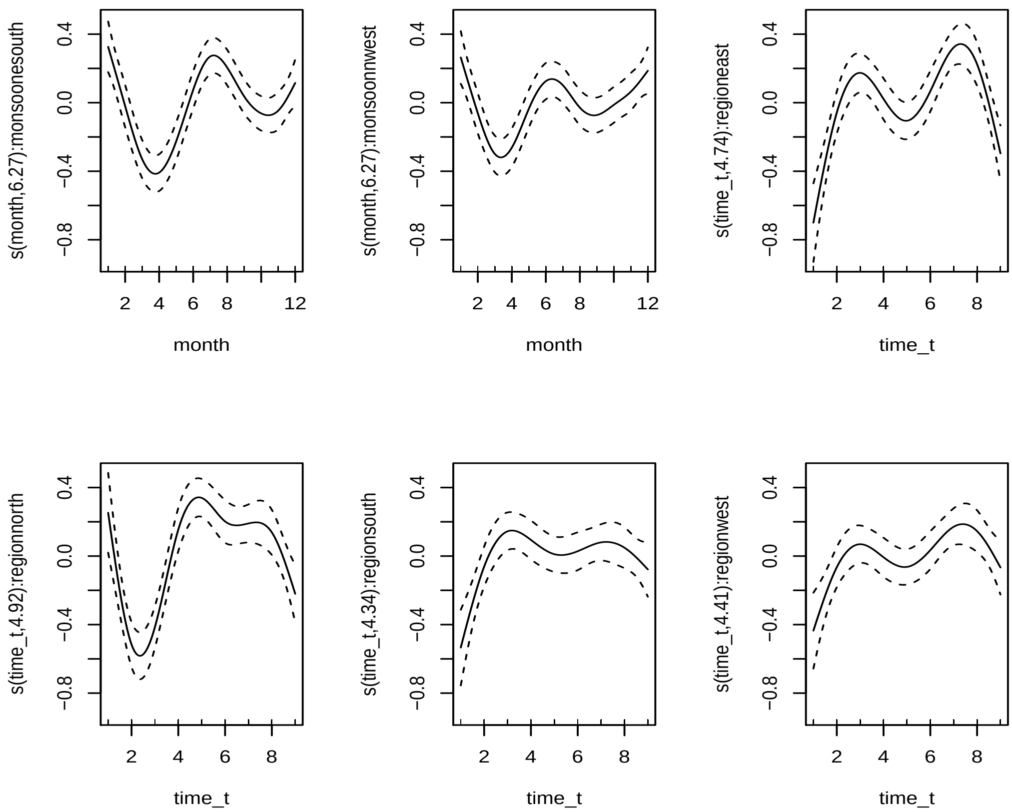

Figure 5.4: *Smooth functions in Model B. Upper row from left to right are monsoon Southwest and Northeast for smooth function of month (by monsoon) followed by North East for smooth function of year (by region). Lower row from left to right are North West, South East and Sourelationship with th West for smooth function of year (by region). In each panel, the solid line is the estimate, and the dashed line is the confidence interval.*

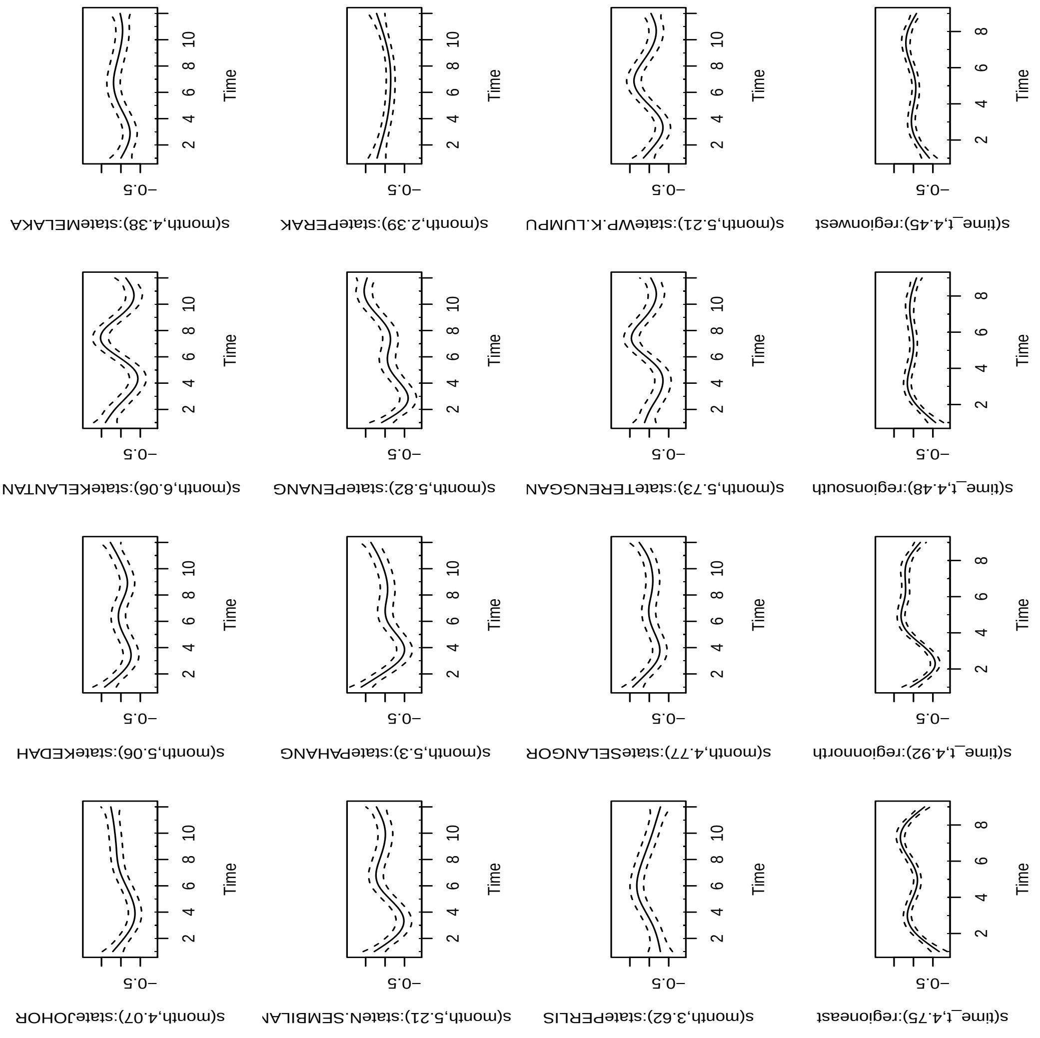

Figure 5.5: *Smooth functions in Model C. First three rows are smooth functions of month for each of the twelve states. Last row is smooth function of year for the four regions. In each panel, the solid line is the estimate, and the dashed line is the confidence interval.*

Table 5.8: *Estimates of the dengue count Model A (by region) parameters.*

Variables	Coefficient	Estimate	Std. err	Z-value	Pr(>Z)
Intercept	α	-9.534e+00	2.232e+01	-0.427	> 0.1
Rainfall	β_1	-1.003e-04	9.124e-05	-1.100	> 0.1
Rainfall Lag 3 Months	β_2	1.913e-04	8.303e-05	2.304	< 0.05
Rainy Days	β_3	1.457e-02	3.923e-03	3.713	< 0.001
Rainy Days Lag 3 Months	β_4	-7.097e-03	3.573e-03	-1.987	< 0.05
Temperature	β_5	9.811e-02	3.590e-02	2.733	< 0.01
Temp. Lag 1 Month	β_6	3.015e-01	8.152e-01	0.370	> 0.1
Nino4 Lag 6 Months	β_7	3.186e-01	7.750e-01	0.411	> 0.1
Temp. Lag 1 and SST Lag 6 Months	β_{67}	-1.097e-02	2.828e-02	-0.388	> 0.1
Population Density	γ_1	7.932e-05	1.070e-05	7.413	< 0.001
DIR Lag 3 Months	γ_3	2.050e-02	1.578e-03	12.989	< 0.001
Region North West	δ_2	-1.089e-01	4.414e-02	-2.468	< 0.05
Region South East	δ_3	1.263e-02	3.770e-02	0.335	> 0.1
Region South West	δ_4	2.871e-01	4.723e-02	6.079	< 0.001

Table 5.9: *Estimates of the dengue count Model B (by monsoon) parameters.*

Variables	Coefficient	Estimate	Std. err	Z-value	Pr(>Z)
Intercept	α	-1.199e+01	2.471e+01	-0.485	> 0.1
Rainfall	β_1	-9.138e-05	9.808e-05	-0.932	> 0.1
Rainfall Lag 3 Months	β_2	1.968e-04	9.081e-05	2.168	< 0.05
Rainy Days	β_3	1.503e-02	4.021e-03	3.737	< 0.001
Rainy Days Lag 3 Months	β_4	-4.486e-03	3.678e-03	-1.219	> 0.1
Temperature	β_5	1.067e-01	3.915e-02	2.724	< 0.01
Temp. Lag 1 Month	β_6	3.745e-01	9.027e-01	0.415	> 0.1
Nino4 Lag 6 Months	β_7	4.097e-01	8.583e-01	0.477	> 0.1
Temp. Lag 1 and SST Lag 6 Months	β_{67}	-1.407e-02	3.132e-02	-0.449	> 0.1
Population Density	γ_1	8.365e-05	1.187e-05	7.047	< 0.001
DIR Lag 3 Months	γ_3	1.977e-02	1.742e-03	11.353	< 0.001
Region North West	δ_2	-1.073e-01	4.863e-02	-2.207	< 0.05
Region South East	δ_3	8.060e-03	4.190e-02	0.192	> 0.1
Region South West	δ_4	2.839e-01	5.230e-02	5.428	< 0.001

Table 5.10: *Estimates of the dengue count Model C (by state) parameters.*

Variables	Coefficient	Estimate	Std. err	Z-value	Pr(>Z)
Intercept	α	-9.110e+00	2.309e+01	-0.394	> 0.1
Rainfall	β_1	-1.435e-04	9.341e-05	-1.537	> 0.1
Rainfall Lag 3 Months	β_2	1.568e-04	8.561e-05	1.832	< 0.1
Rainy Days	β_3	1.109e-02	3.932e-03	2.820	< 0.01
Rainy Days Lag 3 Month	β_4	-1.418e-03	3.503e-03	-0.405	> 0.1
Temperature	β_5	8.729e-02	3.641e-02	2.397	< 0.05
Temp. Lag 1 Month	β_6	2.890e-01	8.438e-01	0.343	> 0.1
Nino4 Lag 6 Months	β_7	3.014e-01	8.019e-01	0.376	> 0.1
Temp. Lag 1 and SST Lag 6 Months	β_{67}	-1.013e-02	2.927e-02	-0.346	> 0.1
Population Density	γ_1	7.461e-05	1.073e-05	6.955	< 0.001
DIR Lag 3 Months	γ_3	2.196e-02	1.612e-03	13.621	< 0.001
Region North West	δ_2	-1.067e-01	4.438e-02	-2.403	< 0.05
Region South East	δ_3	2.178e-02	3.774e-02	0.577	> 0.1
Region South West	δ_4	2.806e-01	4.753e-02	5.904	< 0.001

a significant positive relationship to DIR (biting behaviour of mosquitoes?). As might be expected, DIR at lag 3 months previously has a significant positive relationship with DIR (epidemic nature of the disease) as does population density (infectious nature of the disease). Note that certain model coefficients are not included in Figure 5.6 because formally the estimated values, treated individually, are not significantly different from zero. This is the case for temperature one month previously, sea surface temperature (SST) 6 months previously and the interaction term between sea surface temperature 6 months previously and lag 1 month temperature. However, the lack of individual significance for each of these terms does not necessarily imply a lack of significance for their combined effect (multicollinearity) and so these terms are retained in all three models on the basis that each has emerged as important in some of the previous exploratory and model selection analyses that have been reported in this chapter.

5.4 Summary

This chapter has introduced a range of modelling frameworks for dengue counts, used the negative binomial GLM from this framework to select a subset of 'best' covariates from those explored in Chapter 4 and demonstrated that an equivalent Poisson model is inappropriate because of overdispersion. The chapter has then gone on to further develop the negative binomial GLM by extending it to a range of three negative binomial generalised additive models (GAMs) reporting associated results and comparisons. In the process, it has revealed a considerable amount of useful information about dengue incidence in Malaysia.

The explanatory variables selected were mean rainfall at lag zero and at lag 3 months, mean temperature at lag zero and lag 1 month, number of rainy day at lag zero and lag 3 months, sea surface temperature (SST) lag 6 months, dengue incidence rate (DIR) lag 3 months and interaction between temperature lag 1 month and sea surface temperature lag 6 months. Other covariates which are statistically

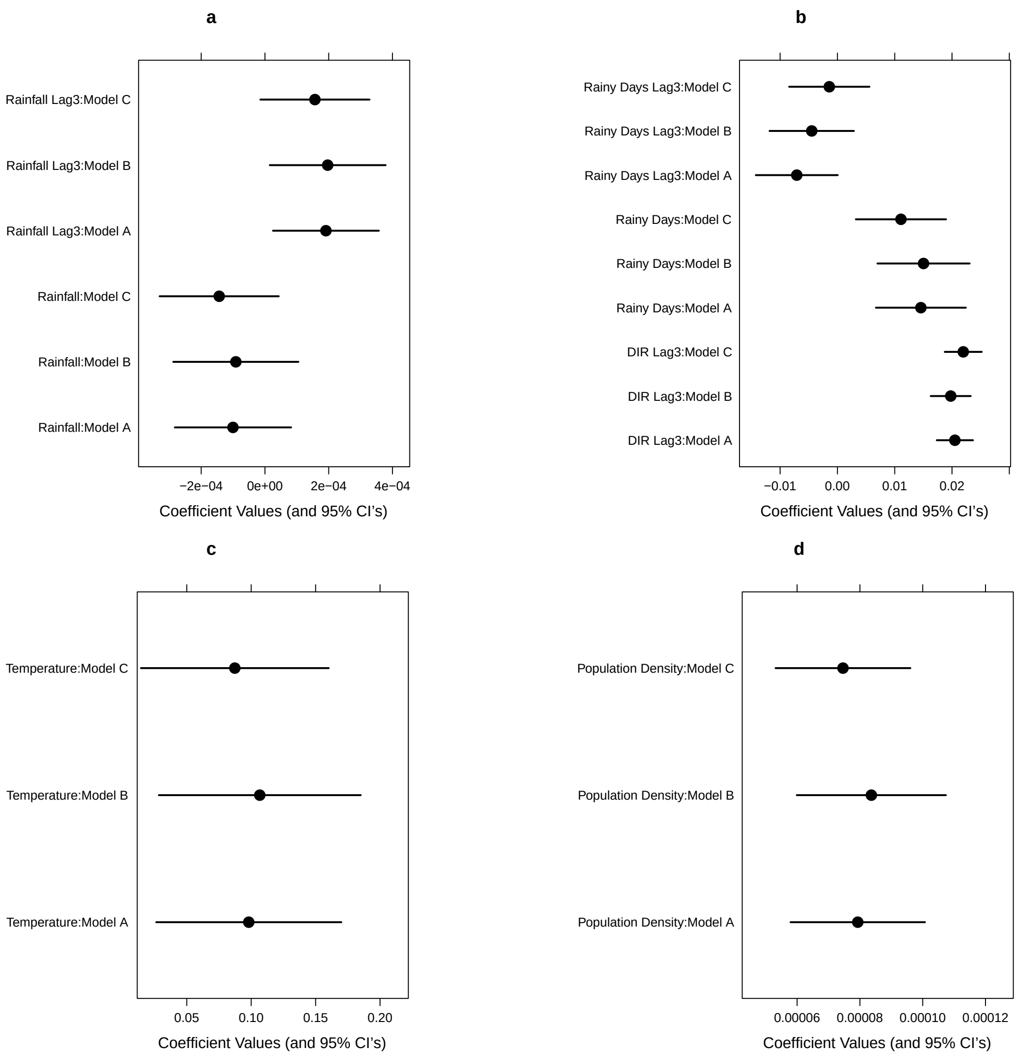

Figure 5.6: *Comparison of most significant coefficients (a) Rainfall and Rainfall Lag 3, (b) Rainy Days, Rainy Day Lag 3 and DIR Lag 3, (c) Temperature and (d) Population Density for Model A, Model B and Model C with estimates error bars represent* 95% *confidence intervals around the respective mean values.*

significant are population, population density, year, month, monsoon area, state and region. It was established that climate information alone does not account for a large proportion of the overall variation in DIR of Malaysia, however, spatio-temporal climate information does significantly account for some of this variability.

The three final models selected for the dengue incidence rate were:

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{r'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.31a}$$

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{m'(s)}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.31b}$$

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f_{r'(s)}\left(z_{2st}\right) + \gamma_3 z_{3st} + f_{s}\left(z_{4st}\right) + \delta_{r'(s)} \tag{5.31c}$$

These three models now need to be investigated further to ascertain their predictive power and hence the scope for using them in developing an early warning system for future dengue epidemics in Malaysia. In order to do that, predictions from these models for future 'out of sample' data need to be fully assessed. In the next chapter this will be investigated by fitting each of the models to 'training data' from 2001 to 2007 and then comparing and contrasting predictions of DIR on out-of-sample 'test data' for 2008 to 2009.

Chapter 6

Model Testing

Although there has been some recent progress in vaccine development for dengue and other innovations focussed on dengue vector control such as the gene-based sterile insect technique by using the RIDL technology and *Wolbachia*-infected *Aedes aegypti* (Lee et al., 2015); it remains the case that for the foreseeable future good predictive models for dengue outbreaks are of key importance in practical dengue prevention in Malaysia. Amongst the many studies on dengue in Malaysia, very few have focussed on predicting future dengue incidence for the purposes of 'early warning' of outbreaks. The only recent relevant work is by Chen and Chang (2013) who report on a predictive tool for dengue outbreaks several weeks in advance of the occurrence based on a moving approximate entropy algorithm applied to the DIR time series. Mohamad Mohsin et al. (2013) investigated associated performance reporting a reasonable balance between the detection rate and the false alarm rate of this model, but the scope for applying such techniques on a wide geographical scale is limited. It is against this background that it is important to establish the extent to which the models developed in Chapter 5 can provide future predictions of dengue incidence that are of practical use at a national scale in Malaysia.

Following the extensive comparative analyses in Chapter 5, three potential statistical models for monthly dengue incidence in Peninsular Malaysia (Models A, B and C) were formulated for further investigation. These models were developed on the full 108 month data set described in Chapter 4. The three models included the same selected climate variables (and associated lags), population density, and a factor allowing for the four geographical regions (and global time trend therein) of Peninsular Malaysia which were delineated in Chapter 5. The difference between the three models was in how the annual seasonal dengue cycle was handled in each i.e. whether this varied according to geographical region, or by monsoon area, or by individual states. Estimates of coefficient values (and standard errors), and of associated smooth functions (and confidence envelopes) were fully reported in Chapter 5, however no detailed consideration was given there to comparing and contrasting the overall fit of the three models, nor to attempting to evaluate formally their 'predictive skill' with regard to dengue incidence.

According, this chapter focusses on these issues in more depth. The chapter proceeds by first contrasting the fit of the three models to the full 108 month data set through formal significance tests, time series plots, analysis of root mean square error and consideration of confidence intervals for fitted values including both parameter uncertainty and uncertainty arising from the negative binomial random element of the models. It then moves on to look at how well the best fitting of the three models performs when predicting 'out of sample' dengue incidence. For that purpose the original data set is divided into two - the first part for model fitting and the second part for testing out of sample predictive validity. Such analyses allow identification of viable prediction lead time and of areas of the country where predictions are weakest. Subsequently, predictions in the weakest areas are investigated in more depth using geographical subsets of the data. Such analyses of out of sample predictions are an important consideration given that a key aim of this study is to investigate the potential to develop early warning systems for future dengue epidemics in Malaysia. However, it should be emphasised that while out of sample analyses in this chapter are valuable pointers, they constitute only a partial evaluation of 'true predictive skill' in the sense that values of the explanatory variables driving the model are not simultaneously being predicted which, obviously, they would also have to be in the context of implementing a practical early warning system for dengue in Malaysia. We return to further discussion of the latter issue in the Chapter 7, but in this chapter it should be understood that when ‘predictions’ are referred to then these are out of sample fitted values assuming the explanatory variable values are known - they are not ‘true predictions’ in the broader sense.

6.1 Model Testing — Comparison of overall fit

This section discusses overall model fit and associated analyses of fitted values for the full 108 month data set from 2001-2009 for Model A, Model B and Model C as specified in Equations 5.30a, 5.30b and 5.30c in Chapter 5. To be clear, these three models differ only in the way that the annual dengue seasonal cycle is represented - Model A has a seasonal cycle represented by a smooth function of month by region, while Model B replaces that with a smooth function of month by monsoon area and Model C replaces it with a smooth function of month by state. So the three models are essentially nested Model A is nested within Model C and Model B is nested within Model A.

Summary statistics of overall fit of these three models to the full 108 month data set were reported in Table 5.7, but not formally compared there. Key aspects of those summaries are reproduced in Table 6.1 along with information on the effective degrees of freedom associated with each of the models.

Using the information in Table 6.1 we can carry out straightforward likelihood ratio tests of differences in fit between the three models. Comparing Model A and Model B gives a likelihood ratio statistic of $2 \times (6879 - 6859)$ to be referred to $\chi^2(12)$ under the null hypothesis of no difference in overall fit between the two models which gives a p-value of $< .0001$ and so indicates a highly significant difference in fit in favour of Model A. A similar comparison between Model A and Model C gives a likelihood ratio statistic of $2 \times (6859 - 6838)$ to be referred to $\chi^2(45)$ which gives a p-value of > 0.123 and so indicates no significant difference in fit between

Table 6.1: *Log-likelihood (LogLik), Akaike Information Criterion (AIC) and effective degrees of freedom (EDF) for Model A, Model B and Model C for overall fit to the full 108 month data set*

Model	LogLik	AIC	EDF
Model A (by region)	-6859	13833	57
Model B (by monsoon)	-6879	13851	45
Model C (by state)	-6838	13857	90

Models A and C.

The broad conclusion that the overall fit of Model A dominates that of Model B and is similar to that of Model C is reinforced by visual inspection of Figure 6.1 which shows a time series plot of fitted monthly DIR versus observed values for Peninsular Malaysia for all three models from 2001-2009. Note that here the monthly DIR values for Peninsular Malaysia in this plot are appropriately averaged from the corresponding state level fits derived from the model, however the root mean square error (RMSE) values reported are derived from the individual versus state monthly predictions. The plot shows little substantive difference between the pattern of fitted versus observed value for the whole of Malaysia from any of the three models - none of them is clearly better than any other. The RMSE values of Model A and Model C are about the same and both somewhat better than those for Model B.

If we then look at analyses of residuals from the overall fit of the three models to the full 108 month data set as shown in Figures 6.2, 6.3 and 6.4 we see a similar comparative picture. The Q-Q plots of standardised deviance residuals for Models A and C are broadly similar - there is some deviation from the 45% line, but this is not extreme indicating that both models fit acceptably and there is little discernible difference between them. However, the plot for Model B exhibits marked deviation from the 45% line suggesting that this model fits less well than A or C.

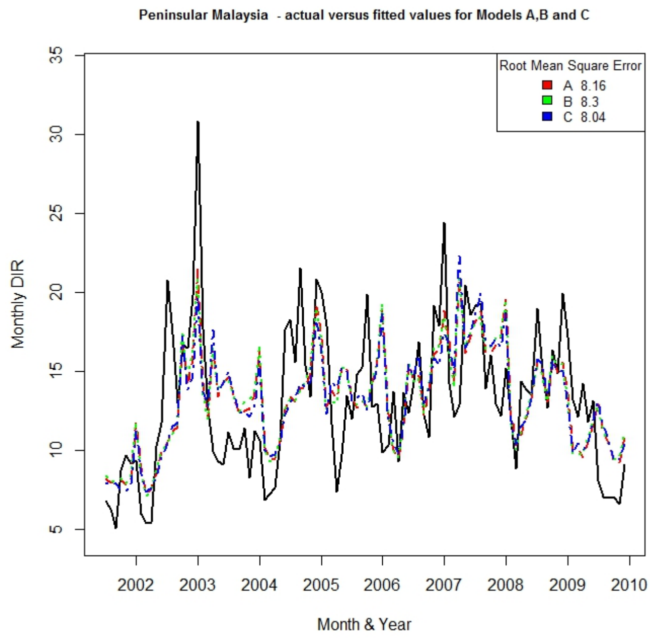

Figure 6.1: *Monthly DIR fitted values for 2001-2009 for Peninsular Malaysia for Models A, B and C and associated root mean square error (RMSE) of constituent fitted values for each month and for each state.*

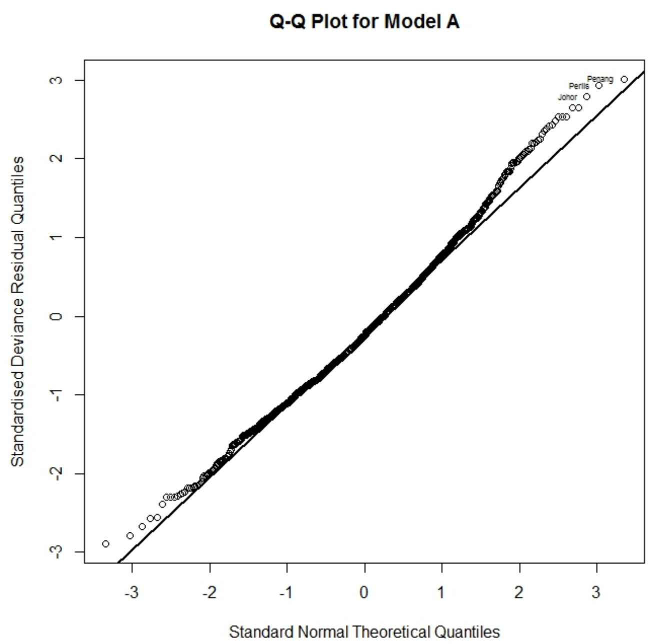

Figure 6.2: *Q-Q plot of standardised deviance residuals for Model A for each state and month from 2001 to 2009.*

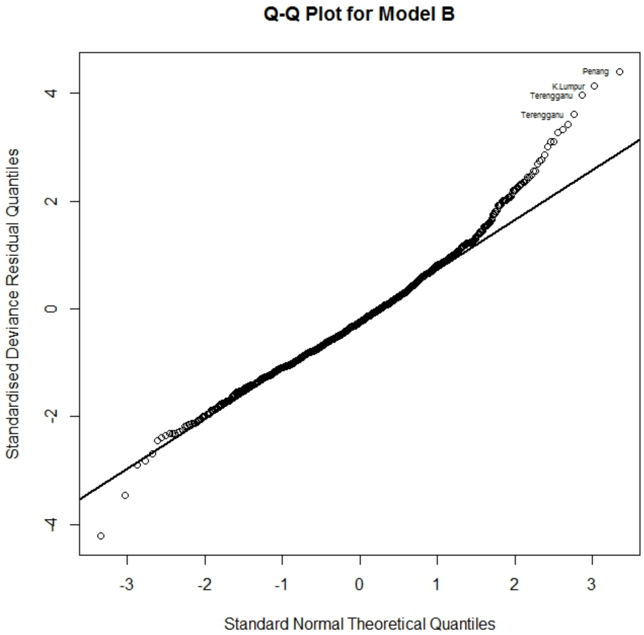

Figure 6.3: *Q-Q plot of standardised deviance residuals for Model B for each state and month from 2001 to 2009.*

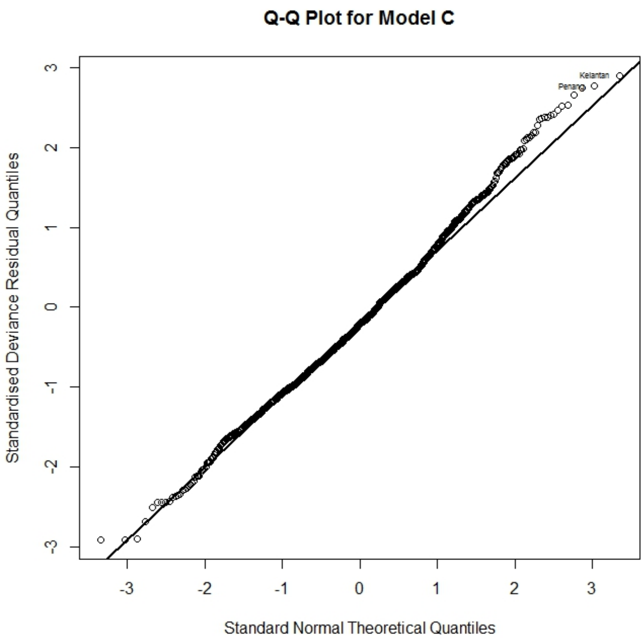

Figure 6.4: *Q-Q plot of standardised deviance residuals for Model C for each state and month from 2001 to 2009.*

The above analyses of model fit for the full 108 month data set strongly suggest that Model A should be the preferred model - there is only a marginal and statistically insignificant difference in fit between it and that of Model C, it is more parsimonious that Model C, and it is markedly better than Model B. For those reasons, we concentrate in the remainder of this section on Model A and investigate further the specification of this model. Fitted versus observed DIR values from this model for the whole of Peninsular Malaysia (at individual state level) are presented in Figure 6.5 along with the root mean square error. Admittedly these fits are well spread for higher monthly DIR values (as might be expected the more extreme values are more difficult to reproduce from the model), but the differences between fitted and observed do appear to be reasonably symmetrically positive and negative and the overall root mean square error of around 8 cases per 100,000 population is acceptable.

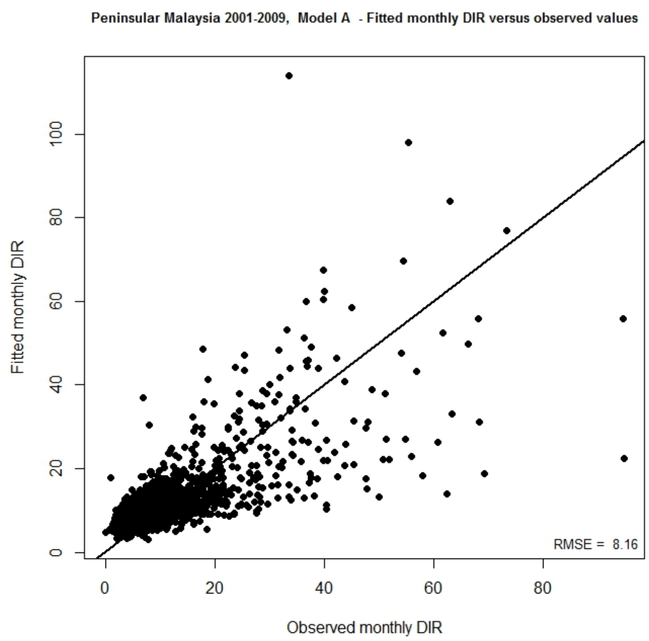

Figure 6.5: *Fitted versus observed values of monthly DIR from Model A for each state and each month for 2001-2009 and associated RMSE.*

An additional way to look at the overall fit and specification of this preferred Model A is to consider the estimated uncertainty in the fitted values and how the envelopes generated from that compare with the observed DIR values. Here there are two sources of uncertainty - that arising from uncertainty in the parameter estimates in the model (which determine fitted mean values from the model), but also that from the additional negative binomial component of the model that might be expected in the responses about those means. Adopting the common parlance from the normal theory linear regression literature we might respectively refer to these as 'confidence intervals' and 'prediction intervals' for the fitted values.

In the case of Model A which is a complex semi-parametric model involving a negative binomial response, a log link for the mean, and a mixture of smooth functions and parametric terms in the mean specification, the theoretical determination of such confidence and prediction errors is not straightforward. It can, however, be addressed through appropriate simulation experiments. To look at parameter uncertainty we take the fitted means for each state and month on the log scale and simulate 1000 Gaussian values around each of those with zero mean and standard deviation equal to their estimated standard deviations (the latter are readily available from the model fitting). We then exponentiate these simulations so obtaining an empirical distribution for model means incorporating parameter uncertainty, from which we may extract quantiles as required. To look at the additional negative binomial uncertainty that might be expected around these predicted means, we then generate an additional 1000 simulations for each of these from a negative binomial distribution with that mean and the associated dispersion parameter estimated for Model A from the fit (recall the mean and variance of the negative binomial are μ and $\frac{\mu^2}{\theta}$ and note that the estimate for θ for Model A here is 4.63).

The associated results for Peninsular Malaysian as a whole are presented in Figures 6.6 and 6.7 (note the simulations involved here are performed at the level of the individual model predictions for each state and then averaged to Peninsular Malaysia as a whole). The key point here is that with the exception of an extreme DIR peak in mid 2002, the 95% Model A prediction envelopes do encompass the observed values. Yes the prediction intervals are wide in some cases, but overall the specification of Model A (random and systematic components) does seem to be appropriate.

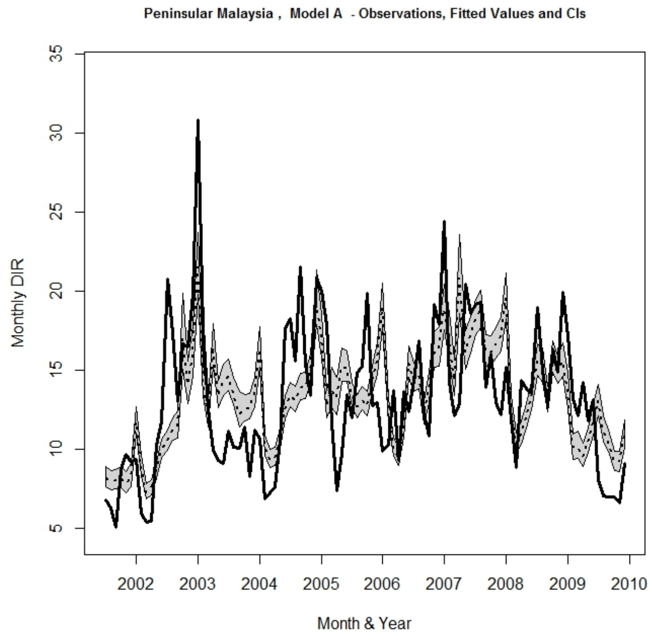

Figure 6.6: *Monthly DIR fitted values for 2001-2009 for Peninsular Malaysia for Model A and associated simulated 95% CIs for mean values for each month and for each state.*

That said, it is not necessarily the case that Model A performs equally well across the whole of Peninsular Malaysia. This model is regional specific and therefore it is reasonable to look at analyses of fitted values broken down to averages within each of the constituent four regions, rather than Peninsular Malaysia as a whole. In most of these regions the fits are similar and reasonably acceptable in terms of RMSE, as typified by those in the North East and shown in terms of fitted versus observed values in Figure 6.8. The South West region is, however, somewhat different - the equivalent plot is given in Figure 6.9 and gives an RMSE which is approximately double that in the other regions. So there may be potential issues here in terms of how well Model A might be able to perform in terms of practical predictions of monthly DIR in the South West Region despite the fact that fitted values in this region do seem to remain within simulated 95% negative binomial prediction bounds as indicated in Figure 6.10. This issue, amongst others, is pursued further in looking at 'out of sample' predictions from Model A in the subsequent section.

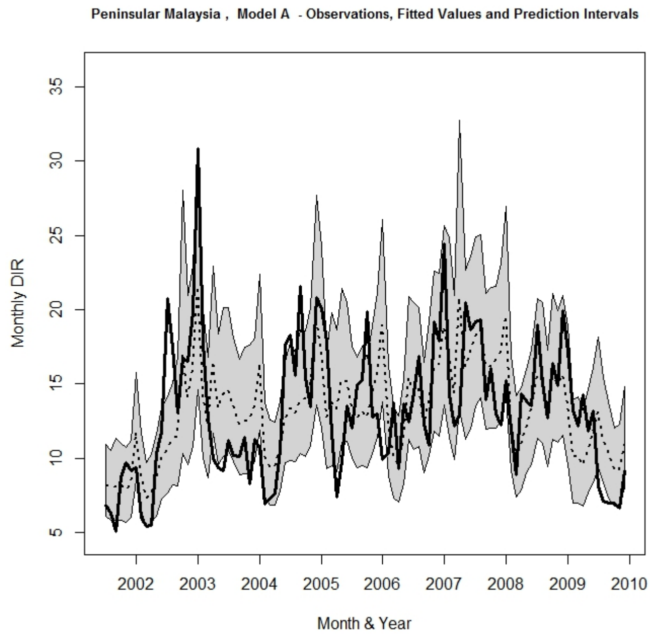

Figure 6.7: *Monthly DIR fitted values for 2001-2009 for Peninsular Malaysia for Model A and associated simulated 95% prediction intervals for each month and for each state.*

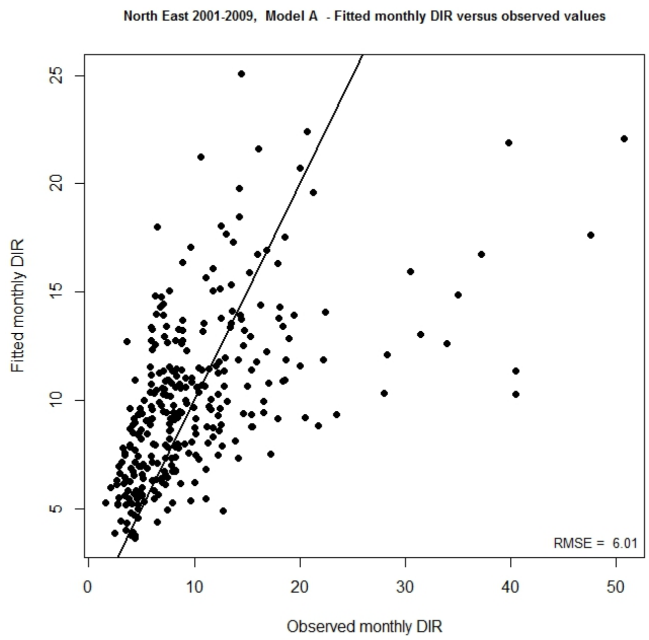

Figure 6.8: *Fitted versus observed values of monthly DIR from Model A for each state and each month for North East region for 2001-2009 and associated RMSE.*

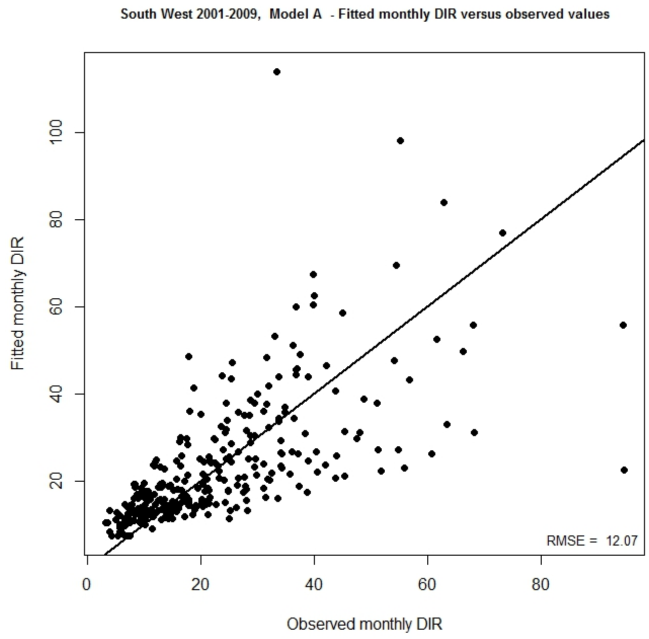

Figure 6.9: *Fitted versus observed values of monthly DIR from Model A for each state and each month for South West region for 2001-2009 and associated RMSE.*

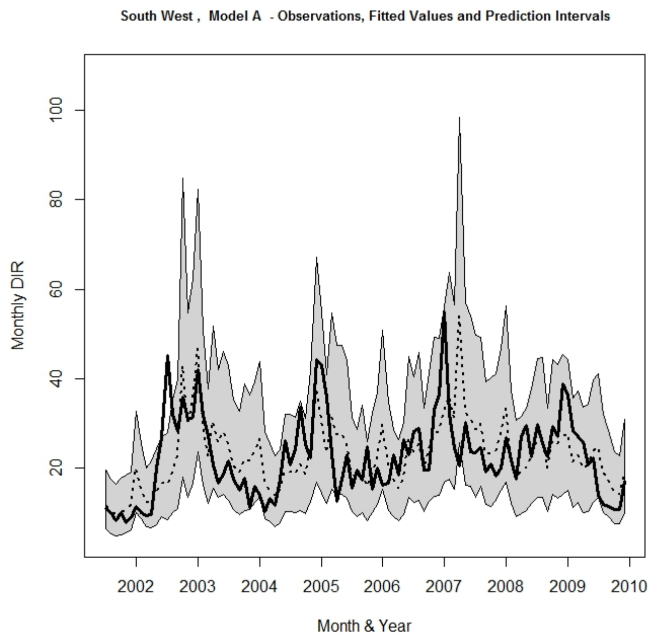

Figure 6.10: *Monthly DIR fitted values for 2001-2009 for South West region for Model A and associated simulated 95% prediction intervals for each month and for each state.*

6.2 Model Testing — Out of sample predictions

The previous section comprehensively analysed the fit of Models A, B and C to the full 108 month data set covering 2001-2009 and concluded that Model A was the preferred model, although possible issues were also raised about the performance of this model in the South West region of Malaysia in particular. This section therefore focusses on the performance of Model A in more detail. More specifically we look at out of sample fits from this model, by splitting the 108 month data set into two periods - the first period being used to fit the model and the second used to evaluate predictions from that model. As said earlier in this chapter, this is not a complete evaluation of the predictive validity of the model (because the values of the explanatory variables are not simultaneously being predicted), nevertheless it does provide a strong indication of the ability of the model to predict future DIR and over what lead times. In practical terms future predictions of monthly DIR beyond two years (24 months) are of little interest. Accordingly the analyses in this section use the data set up to December 2007 to fit the model (so six and a half years of data given that six months are lost due to the lagged variables in the model) and then we consider out of sample predictions for the two year subsequent period (Jan 2008 - Dec 2009).

We start by looking at such out of sample monthly DIR predictions for the whole of Malaysia (averaged over all states) for Model A as shown in Figure 6.11 which also indicates the simulated prediction envelope (this is based upon the simulation scheme described in the previous section of this chapter and incorporates both parameter uncertainty from fitting the model to data from 2001-2007 and also that from the negative binomial response using the associated estimated dispersion parameter, which for Model A during 2001-2007 is 4.72).

Clearly, DIR predictions from the model (and the associated prediction intervals) degrade the further into the future we consider. Beyond a 12 month lead time these begin to break down becoming somewhat uninformative during that period (and even misleading towards the end of the period). We can look at this more

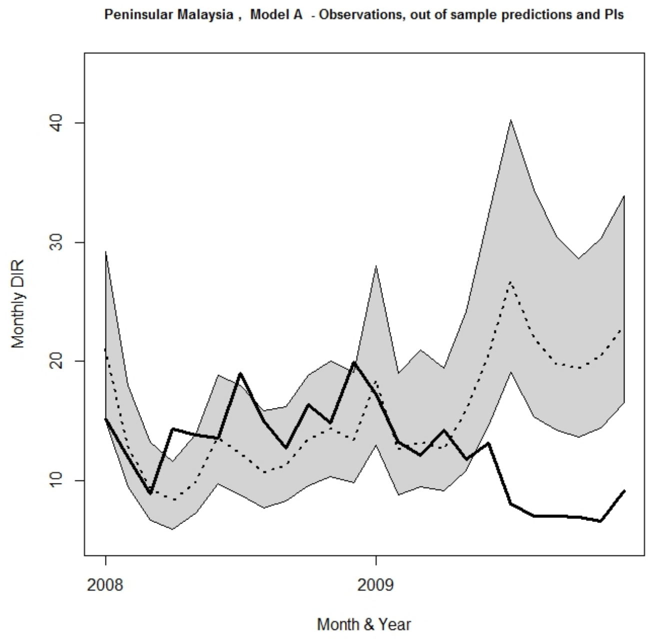

Figure 6.11: *Monthly DIR out of sample predicted values for 2008-2009 for Peninsular Malaysia for Model A and associated simulated 95% prediction intervals for each month and for each state.*

specifically by considering out of sample predictions versus observed values for each month and state for the whole of Peninsular Malaysia for lead times of 3, 6, 12 and 24 months along with associated RMSE as shown in Figure 6.12. This indicates that the 3 and 6 month lead time predictions may be acceptable, beyond that the RMSE becomes much larger.

Given this overall picture, it is sensible to look at the same analyses within each of the four regions of Peninsular Malaysia. The results which emerge from that are broadly similar for the North East, North West and South East areas as typified by those shown for the North East region in Figures 6.13 and 6.14. The suggestion is that model DIR predictions for the future six months may be acceptable - beyond that they begin to break down.

However, the situation in the South West region is somewhat different as shown in Figures 6.15 and 6.16.

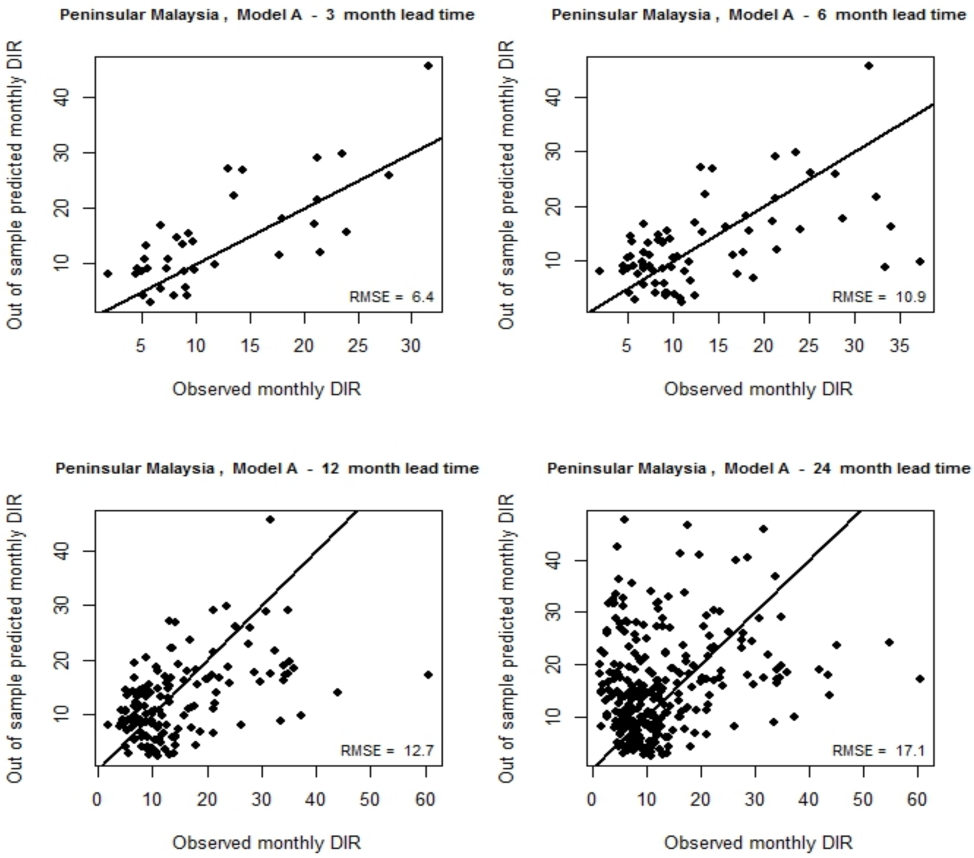

Figure 6.12: *Out of sample predicted values for 2008-2009 for each month and for each state versus observed values for Peninsular Malaysia for Model A and associated RSME values for lead times of 3, 6, 12, and 24 months.*

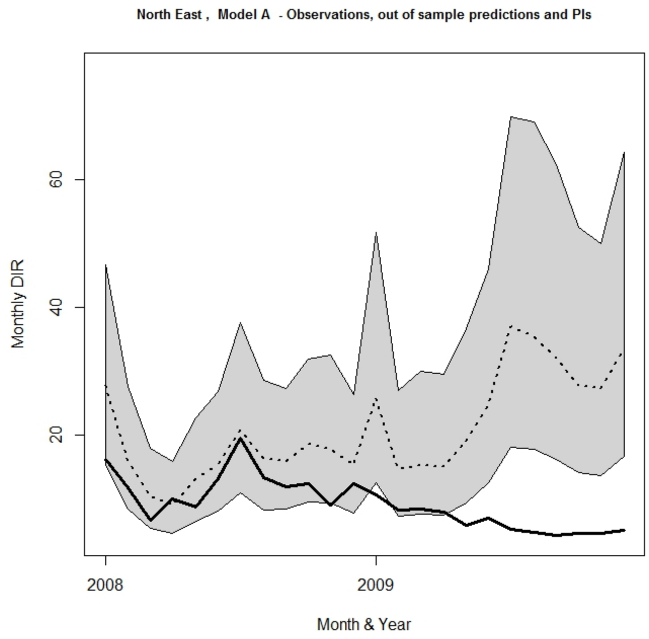

Figure 6.13: *Monthly DIR out of sample predicted values for 2008-2009 for North East region for Model A and associated simulated 95% prediction intervals for each month and for each state.*

Here, although the prediction intervals do broadly incorporate the observed values, we see the RMSE of out of sample predictions from the preferred Model A dramatically increase beyond a 3 month lead time in contrast to that in the other three regions where both 3 and 6 month lead times have an acceptable RMSE.

6.3 Model Testing — Out of sample predictions (Kuala Lumpur/Selangor)

Given the conclusions of previous sections in this chapter, we focus in this section on the out of sample predictions from the preferred Model A for the South West region of Peninsular Malaysia as compared with those in the other regions. Looking at the RMSE in each of the three states in this region it is clear that it is the states of Kuala Lumpur and Selangor which exhibit distinct differences from the rest of

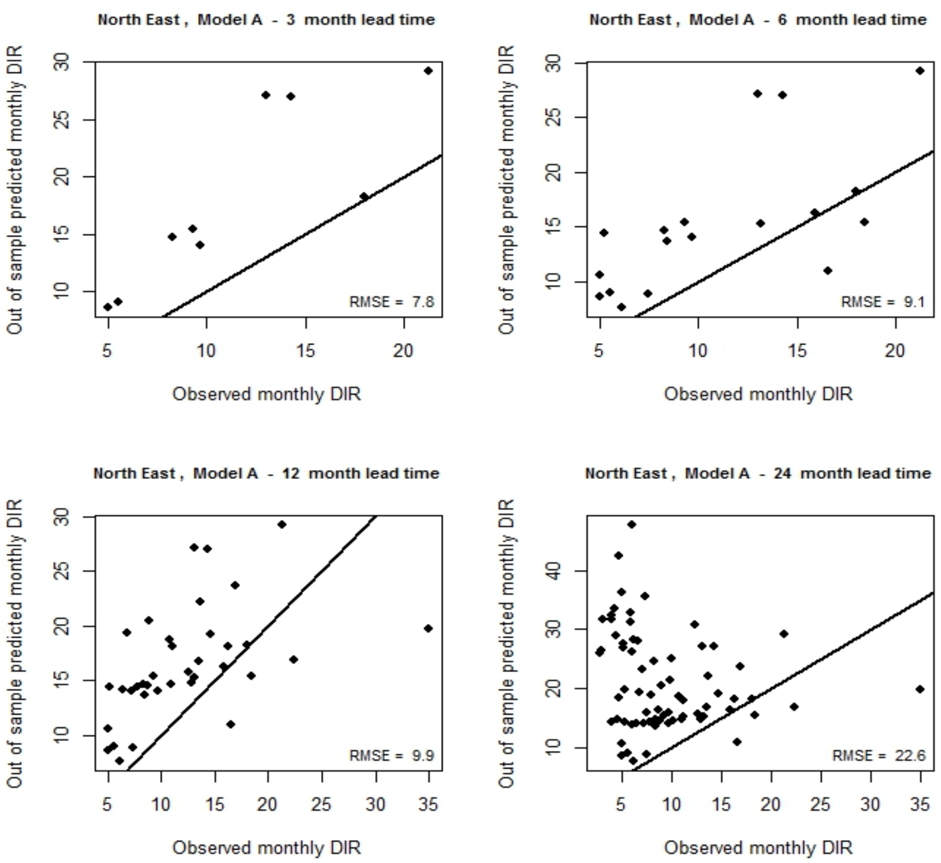

Figure 6.14: *Out of sample predicted values for 2008-2009 for each month and for each state versus observed values for North East Malaysia for Model A and associated RSME values for lead times of 3, 6, 12, and 24 months.*

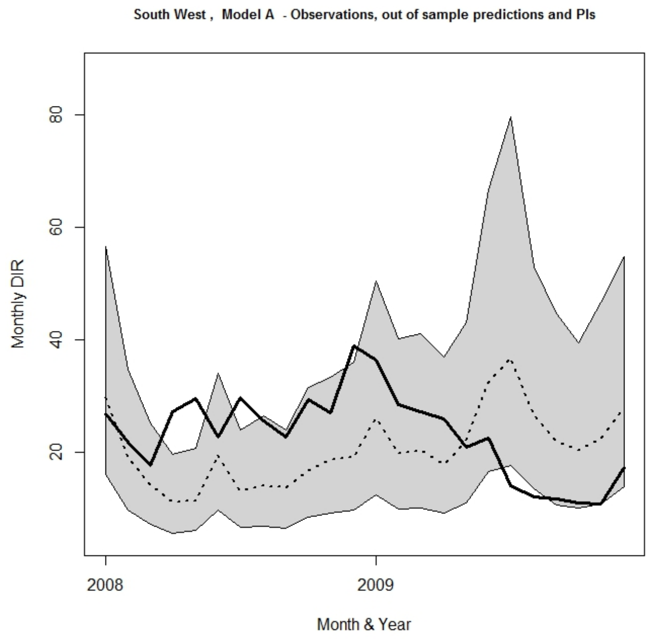

Figure 6.15: *Monthly DIR out of sample predicted values for 2008-2009 for South West region for Model A and associated simulated 95% prediction intervals for each month and for each state.*

Malaysia. Hence in this section we consider splitting the data set into two, one which just contains Kuala Lumpur and Selangor and the other which contains the rest of Peninsular Malaysia. We fit Model A separately to each of these using data up to December 2007 to fit each model (so, as before, six and a half years of data given that six months are lost due to the lagged variables in the model) and then we consider out of sample predictions for the two year subsequent period (Jan 2008 - Dec 2009). Note that the form of Model A for the fitting to Peninsular Malaysia without Kuala Lumpur and Selangor remains as previously; however, neither the region factor nor separate regional seasonal cycles are necessary when Model A is fitted to the data set which contains just Kuala Lumpur and Selangor because there is only one region involved so the linear predictor (using the notation introduced previously) in that case is just:

$$\log \rho_{st} = \alpha + \sum_{j}^{7} \beta_j x_{jst} + \beta_{67} x_{6st} x_{7st} + \gamma_1 z_{1st} + f\left(z_{2st}\right) + \gamma_3 z_{3st} + f\left(z_{4st}\right) \quad (6.1)$$

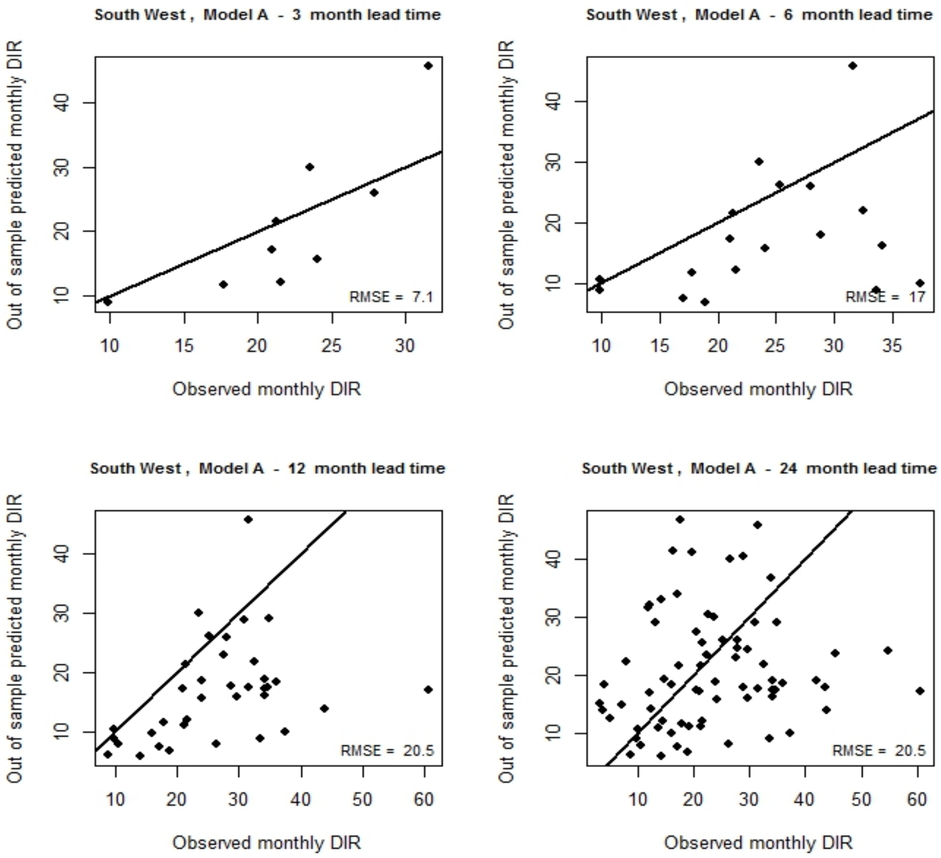

Figure 6.16: *Out of sample predicted values for 2008-2009 for each month and for each state versus observed values for South West Malaysia for Model A and associated RSME values for lead times of 3, 6, 12, and 24 months.*

First we look at out of sample predictions for the case of the whole of Peninsular Malaysia without Kuala Lumpur and Selangor as shown in Figures 6.17 and 6.18.

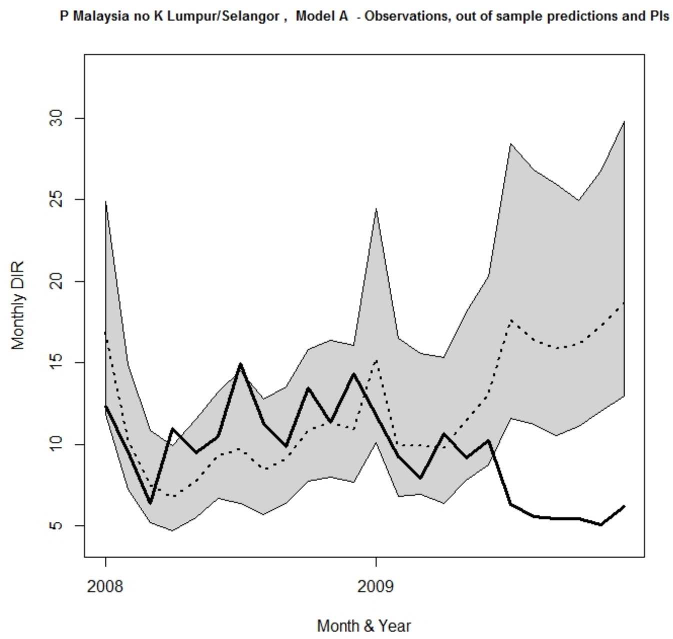

Figure 6.17: *Monthly DIR out of sample predicted values for 2008-2009 for Peninsular Malaysia without the states of Kuala Lumpur and Selangor for Model A and associated simulated 95% prediction intervals for each month and for each state.*

Both of these figures are broadly similar to the equivalent plots from Model A for the whole of Peninsular Malaysia including Kuala Lumpur and Selangor shown in Figures 6.11 and 6.12 in the previous section. As there, the DIR predictions from the model (and the associated prediction intervals) degrade the further into the future we consider. Looking at the predictions and RMSE for lead times of 3, 6, 12 and 24 months in Figure 6.18 indicates that the 3 and 6 month lead time predictions may be acceptable and indeed the RMSE are improved in those cases from those seen in 6.12. So excluding Kuala Lumpur and Selangor has improved the fit of Model A and the associated out of sample predictions for the other states of Peninsular Malaysia.

Turning to out of sample predictions for the case of data from just Kuala Lumpur and Selangor, the predictions for different lead times and the associated RMSE are

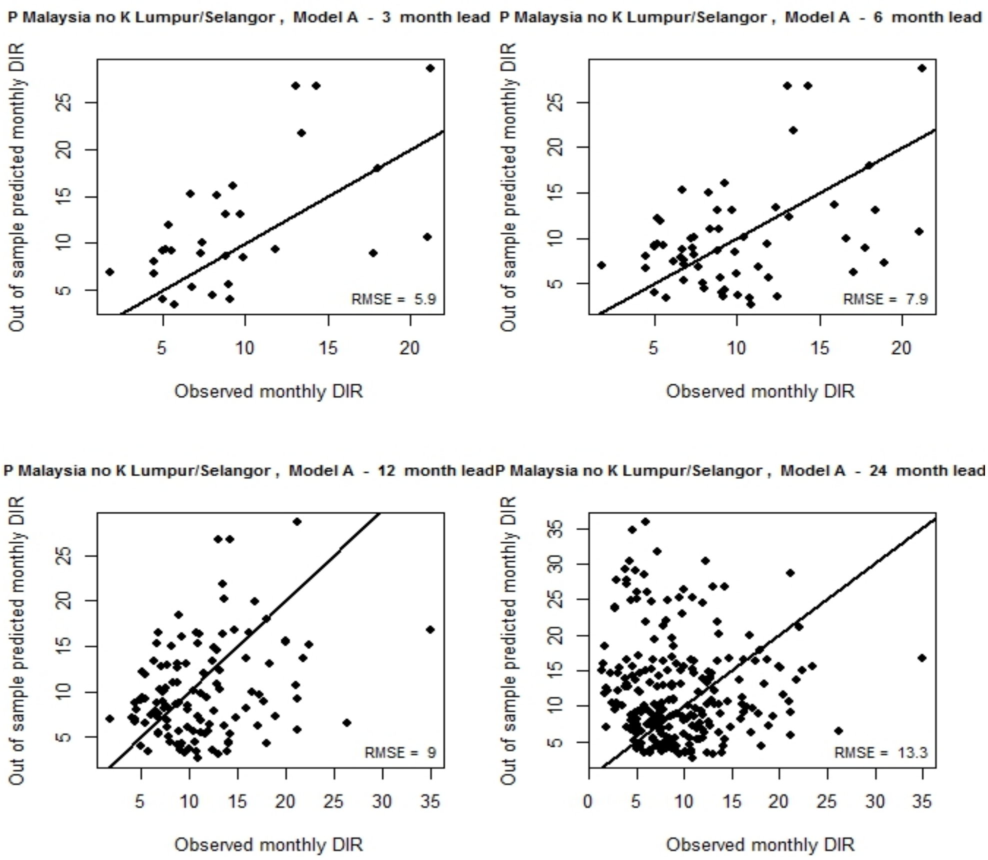

Figure 6.18: *Out of sample predicted values for 2008-2009 for each month and for each state versus observed values for Peninsular Malaysia without the states of Kuala Lumpur and Selangor for Model A and associated RSME values for lead times of 3, 6, 12, and 24 months.*

shown in Figure 6.19.

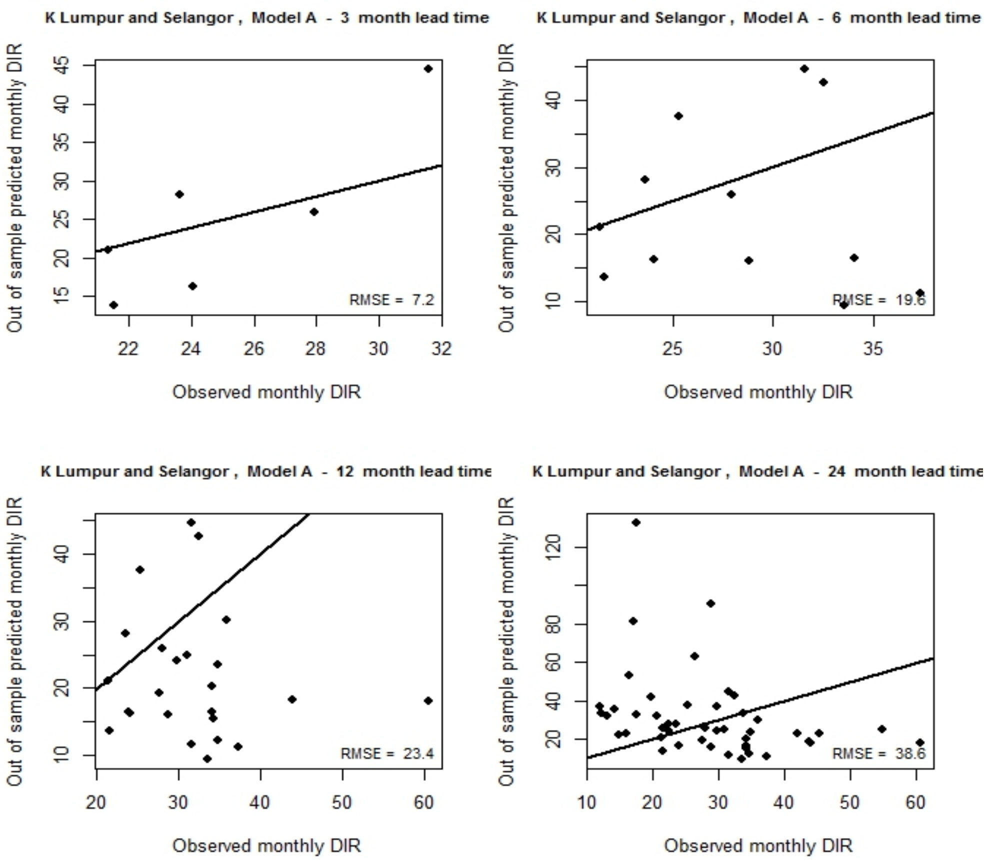

Figure 6.19: *Out of sample predicted values for 2008-2009 for each month and for each state versus observed values for Peninsular Malaysia without the states of Kuala Lumpur and Selangor for Model A and associated RSME values for lead times of 3, 6, 12, and 24 months.*

Even though now Model A is just fitted to these two states in producing these plots, the out of sample predictions for Kuala Lumpur and Selangor still perform badly. The RMSE dramatically increases beyond a 3 month lead time in contrast to the equivalent results obtained for the other ten states. If one looks at estimates for the parametric terms for this model (i.e. Model A fitted to data from 2001-2007 for just Kuala Lumpur and Selangor) it is immediately apparent that there are significant differences in these two states as opposed to the rest of Peninsular Malaysia. The only significant parametric covariates are population density and DIR lagged by 3 months all the climate covariates are insignificant. The non-parametric term for yearly trend is significant but that for seasonal cycle is not. From this and the RMSE results discussed earlier in this section and earlier sections, it is evident that the kind of models we have developed in Chapter 5 simply do not work well for Kuala Lumpur and Selangor. Clearly, there are other more complex factors involved here and the climate signal seems to have very little influence. This supports the conclusions of Hafiz et al. (2012) concerning the special circumstances pertaining to dengue incidence in these highly urbanised areas. Aziz et al. (2012) reported that mean monthly rainfall in Kuala Lumpur did not seem to influence the pattern of dengue cases. Cheong et al. (2013) also tested the effects of minimum temperature, bi-weekly accumulated rainfall and wind speed on dengue cases in Selangor and Kuala Lumpur from 2008 to 2010 indicated that temperature and rainfall have complex influences on dengue transmission in high population density areas such as Kuala Lumpur (clustering versus dispersion). Such results are also supported by Hafiz et al. (2012) who produced risk maps indicating patchy high risk for dengue in Kuala Lumpur and parts of the surrounding districts, Gombak and Petaling (Selangor state). Another consideration is that of higher misreporting of dengue cases in these highly urbanised areas which do experience higher numbers of dengue cases than elsewhere in Peninsular Malaysia. Clinicians should report all dengue cases to the Health Office, but there are some cases that may be reported in the wrong month as noted in studies by Earnest et al. (2012a). Up to year 2000, most dengue cases were reported accurately in the working and school-age groups (Aziz et al., 2014) but the increase in urbanisation is recognised as a confusing factor in Malaysia (Seng et al., 2005). At the same time, the framework for the e-Dengue teleconsultation system built by Setyono et al. (2011) allows patients to express early concern about the disease and perform teleconsultation via the Internet or mobile phone rather than proceed through more traditional routes so adding to uncertainties in notified cases.

6.4 Summary

The results from this chapter show a mixed picture. On the one hand, there does seem to be some predictive potential for lead times of up to six months in the models we have developed outside of Kuala Lumpur and Selangor. On the basis of the parsimony principle and the comparative analyses presented in this chapter, Model A is preferred over either Model B or Model C. Model C seems to add little to predictive accuracy, despite being state based rather than region based. Model A is therefore proposed as a potentially useful models for developing early warning systems. These results are broadly in line with views expressed by Wan Fairos et al. (2010).

However, on the other hand, the analyses in this chapter strongly indicate that none of these models work well for the highly urbanised states of Kuala Lumpur and Selangor where the climate signal seems to have little importance and there are clearly more complex influences involved. The latter issue will be further discussed in the next and concluding Chapter 7.

Chapter 7

Summary and Conclusions

In this final chapter, the main findings of the study are summarised. The limitations of the study and various important remaining issues, including possible directions for future work, are then discussed.

7.1 Main findings

The main contribution of this study is in identifying the extent to which current nationally available data in Malaysia forms a basis for potentially developing statistical models to predict future (e.g. three to six month) spatio-temporal variations in dengue incidence risk for Malaysia. The main direction of the study was to develop a 'best' model for dengue incidence at a national scale based upon routinely available data which combined climatic and non-climatic factors and then to evaluate to what extent predictions from such a model are able to reflect what actually did occur. To the best of the author's knowledge, this is the first study that has looked at these issues in Malaysia at a national level including climate information in modelling spatio-temporal variations for dengue incidence and considering a long period (January 2001 to December 2009) on a monthly basis with due regard to regional, monsoon area and state specific issues.

The study involved two data sets; one containing annual data of dengue cases and related crude demographic explanatory variables from 1991 to 2009 (only used to analyse global trends in DIR) and, more substantially, a second in which the modelling focussed on monthly data of dengue cases and related explanatory variables from January 2001 to December 2009 (108 months) and associated possible climate and demographic covariates (the key data set used for the modelling of DIR in Chapters 5 and 6). In the modelling, the twelve states of Peninsular Malaysia were divided into four regions based on geographical location; North East, South East, North West and South West. The North East region refers to the East of the Malaysia including the states of Kelantan, Terengganu and Pahang, the South East region consists of the states of Johor, Melaka and Negeri Sembilan located in the South of Malaysia. The North West refers to the North part of Malaysia containing Kedah, Penang and Perlis meanwhile the South West region includes the capital of Malaysia, Kuala Lumpur as the Centre of the Country, along with Selangor and Perak. As established in the exploratory analyses in Chapter 4, the highest monthly dengue incidence rates from 2001 to 2009 were recorded in Penang (2001), Kuala Lumpur (2002 to 2007) and Selangor (2008 and 2009). Kuala Lumpur and Selangor (South West region) are confirmed to be the states with the most significantly high DIR patterns after 2001. The South West region shows a significant difference in DIR patterns compared to the other three regions for the 108 months and this is discussed further subsequently in this section. Meanwhile, the overall results in other areas of Peninsular Malaysia indicate that significant (if weak) relationships exist between DIR and climatic variables (and their lags) and that these may be able to be exploited in developing predictive early warning systems in association with relevant weather/climate forecasts.

Exploratory data analyses in Chapter 4 showed that there is some evidence of an increasing annual trend in DIR in all states of Malaysia. There is also a strong in-year seasonal cycle in DIR and differences in this cycle may need to be allowed for in different broad geographical regions of Malaysia and possibly in different states. High population density is positively related to monthly DIR as is DIR in immediately preceding months. Relationships between monthly DIR and climate variables and lagged values of these are generally weaker, although significant in some cases. In summary the analyses in Chapter 4 concluded that DIR in Malaysia is potentially associated with country wide trend, regional seasonal cycle, population, population density, dengue incidence in preceding months, lagged average temperature, average rainfall, number of rainy days and ENSO.

In Chapter 5 a negative binomial GLM was used to select a subset of 'best' covariates from those explored in Chapter 4. The explanatory variables selected were mean rainfall current and lag 3 months, mean temperature current and lag 1 month, number of rainy day current and lag 3 months, sea surface temperature (SST) lag 6 months, dengue incidence rate (DIR) lag 3 months and interaction between temperature lag 1 month and sea surface temperature lag 6 months. Population, population density, year, month, monsoon area, state and region. It was demonstrated that a equivalent Poisson formulation of the final model selected was inappropriate because of overdispersion. The negative binomial GLM was then extended to a range of negative binomial generalised additive models (GAMs) and associated results and comparisons were reported. Using these models it was established that climate information alone does not account for a large proportion of the overall variation in DIR of Malaysia, however, spatio-temporal climate information does significantly account for some of this variability. The influence of monsoon area and regional differences were important. It was found that for the most part geographical differences can be adequately captured without significant loss of detail by grouping the twelve states into the four broad regions mentioned earlier, however there is some evidence of more localised state effects particular in the South West of the country where the main urban areas of Malaysia are located (Kuala Lumpur and Selangor). The smooth functions for seasonal cycle differ in detail between regions but all see DIR peaks in July and January each year. Global trend in DIR in Malaysia also differs in detail in different regions, but in general there is significant upward trend. Chapter 5 ends by identifying three negative binomial GAM models that may be useful in predicting monthly DIR at a national level in Malaysia. The key differences between these three models was in whether the smooth function in the model representing the seasonal cycle was region specific (Model A), or monsoon area specific (Model B), or state specific (Model C).

The three Models A, B and C then became the focus for further investigation in Chapter 6. Here the fit of the three models to the full 108 month data set were analysed in detail including the use of simulation experiments to take both parameter and negative binomial model uncertainty into account. Model A was found to be the preferred model. This model was then fitted to first 78 months of the data set up until December 2007, and then 'out of sample' predictions for the subsequent 2 years from January 2008 to December 2009 were analysed and compared again using simulation experiments to allow for both parameter and negative binomial model uncertainty. Different lead times for predictions of 3, 6, 12 and 24 months were considered. The results indicated that model A did provide acceptable out of sample predictions for lead times of up to six months in areas other than the highly urbanised areas of Kuala Lumpur and Selangor in the South West of the country. Subsequent analyses split the data set into that pertaining to Kuala Lumpur and Selangor and that relating to the rest of Peninsular Malaysia and repeated the 'out of sample' analyses with Model A being fitted separately to each of those data sets. This improved results for the states other than Kuala Lumpur and Selangor, but did not help in those latter states where predictions remained poor. The overall conclusions from Chapter 6 were that there does seem to be some predictive potential for up to six months lead time from Model A in areas outside of Kuala Lumpur and Selangor. However, on the other hand, this preferred model evidently does not work well for Kuala Lumpur and Selangor where there are clearly more complex influences involved. There are a large number of patchy densely populated urban centres in Kuala Lumpur and Selangor and the lack of data collation relative to these intra-state localised conditions make DIR predictions very difficult compared to the rest of Malaysia.

7.2 Limitations of the study

One clear set of limitations of this study is the gaps in the dengue data that are available on a national scale and for a long enough time period in Malaysia. First, there are questions concerning the quality of data and of how well the national surveillance systems are operating - how reliable and consistent they are both over time and in different geographical areas. It is likely that the data used are subject to under-reporting and mis-reporting problems. Second, the lack of dengue data based on different age groups, given that there are documented relationships between age and dengue incidence. Third, the lack of serotype data, given that the pattern of serotype circulation is critical in understanding dengue epidemics. Finally, the absence of long-term dengue data at localised district level within each state. The state is just too low a level of spatial resolution to resolve the complex interacting factors which determine variations in DIR. Put simply, Malaysia does not have publicly available information systems in place to provide for full-scale analysis of the impact of climate and weather on dengue transmission. Both the Ministry of Health (MOH) and the Malaysian Meteorological Department (MMD) do not have accessible databases of sufficient detail, reliability, geographical coverage and longevity.

There are also limitations relating to the kinds of statistical models that have been used in the study. The lack of localised data has meant there has been little alternative to adopting quite high level ecological models. We believe that the GAMs used are the best that could have done with the available data and do demonstrate some potential for use in developing early warning systems for dengue, at least outside of Kuala Lumpur and Selangor. However, it has to be acknowledged that they remain essentially descriptive models rather than process models and as such can only ever provide limited information in the face of the complexity and dynamics of the vectors and hosts involved in dengue transmission.

7.3 Scope for future work

As indicated in the previous section, the scope for further work of the nature described in this study is severely limited by the availability of better data. The databases of certain government departments in Malaysia are not available for public access and research access is dogged by long delays and high levels of bureaucracy. A priority for researchers to better understand the influence of climate and other factors on dengue transmission in Malaysia is the establishment by government agencies of easily accessible linked databases of past and current information on dengue, climate and socio-economic conditions.

That said, there is some immediate further work that could be conducted:

- Clearly, more investigation needs to be carried out in Kuala Lumpur and Selangor to identify reasons for the models' failure to predict in these urbanised areas. It may be possible to 'downscale' the climate information in these specific areas and obtain sub-state district dengue data for a long enough period to throw more light on the issues involved.

- Further work could also be done on identifying different climate zones in Malaysia and using these in the models as a replacement to the rather crude divisions of state, region and monsoon area used in this study. One issue there is obtaining a better understanding of whether the severity of monsoons has any impact on DIR in the different regions and, if so, whether and how this might be linked to the interaction between sea surface temperature and atmospheric temperature as a determinant of the severity of the monsoons that Malaysia is subjected to.

- Another area is investigating to what extent DIR predictions from the models developed in this study remain valid when the observed climate data in the models is replaced with relevant seasonal forecast data. What lead times can be achieved and what levels of confidence can be placed in the associated predictions? The lags in the climatic covariates in the models raise the possibility of developing forecasts of monthly DIR up to 6 months in advance of the month in question if suitable seasonal forecasts can be obtained from the relevant government agencies. This ties in well with the results obtained in Chapter 6 which indicated acceptable out of sample predictions from Model A for lead time of up to six months in areas outside of Kuala Lumpur and Selangor. If six month seasonal forecasts of the climate variables were available then the only model variable that would prevent rolling six month ahead dengue forecasts for all areas outside of Kuala Lumpur and Selangor would be the DIR lagged by 3 months which would be unavailable for months 4, 5 and 6 of the forecast. However, it is possible that forecast values of DIR in the first three months could be used as a surrogate for that variable in deriving the forecasts for months 4, 5 and 6. Clearly this is all dependent on the availability of regular timely seasonal climate forecasts and the there would need to be further study to evaluate the accuracy of such a forecasting approach.

- Finally, given that one finding of this study is that climate information alone does not account for a large proportion of the overall variation in DIR of Malaysia, there is further scope for investigating more detailed socio-economic factors and population movement, rather than the simple demographics of population and population density used in this study.

7.4 Summary

This study has highlighted the potential for incorporating climate information into a spatio-temporal dengue epidemic early warning system for Malaysia. The covariates used and their interaction in the modelling framework developed is a new development in dengue modelling in Malaysia and provides a potential groundwork for future models to be developed. Despite the limitations of the model and the difficulties involved in developing the best predicting dengue incidence model, it is hoped that this spatio-temporal dengue prediction model is a step towards the development of a useful decision making tool for the Malaysian health services. The potential models developed could be extended to the district level of each state so that they are able to provide more localised predictions. Hopefully, the framework developed will be used as a starting point to investigate further if climate information can valuable be incorporated in an early warning system for dengue in Malaysia and that the results produced in this study will assist researchers interested in dengue from other fields (public health, clinicians, geographers, environmental scientists etc.).

References

Abu Bakar, A., Kefli, Z., Abdullah, S., and Sahani, M. (2011). Predictive models for dengue outbreak using multiple rulebase classifiers. *International Conference on Electrical Engineering and Informatics (ICEEI), 17-19 July 2011*, page doi:10.1109/ICEEI.2011.6021830.

Abu Bakar, A. L. (1996). *Melaka dan Arus Gerakan Kebangsaan Malaysia*. Penerbit Universiti Malaya, Kuala Lumpur.

Abu Bakar, S. and Shafee, N. (2002). Outlook of dengue in Malaysia: a century later. *The Malaysian Journal of Pathology*, 24(1):23–27.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E. (2003). The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *Journal Hydrometeorology*, 4(6):1147–1167.

Adriana, F. G., Aline, A. N., and Oswaldo, G. C. (2012). Temporal analysis of the relationship between dengue and meteorological variables in the city of Rio de Janeiro, Brazil, 2001-2009. *Cad. Saude Publica, Rio de Janeiro*, 28(11):2189–2197.

Ahmed, S. A., Siddiqi, J. S., Quaiser, S., and Kamal, S. (2015). Using PCA, Poisson and Negative Binomial model to study the climatic factor and dengue fever outbreak in Lahore. *Journal of Basic and Applied Sciences*, 11:8–16.

Aiken, S. R., Frost, D. B., and Leigh, C. H. (1980). Dengue hemorrhagic fever and rainfall in Peninsular Malaysia: some suggested relationships. *Social Science and Medicine*, 14d(3):307–316.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceeding of the Second International Symposium on Information Theory*, pages 199–213.

Althouse, B. M., Ng, Y. Y., and Cummings, D. A. T. (2011). Prediction of dengue incidence using search query surveillance. *PLoS Neglected Tropical Diseases*, 5(8):doi:10.1371/journal.pntd.0001258.

Ang, K. T., Rohani, I., and Look, C. H. (2010). Role of primary care providers in dengue prevention and control in the community. *Medical Journal Malaysia*, 65(1):58–62.

Anyamba, A., Chretien, J., Small, J., Tucker, C. J., and Linthicum, K. J. (2006). Developing global climate anomalies suggest potential disease risks for 2006/2007. *International Journal of Health Geographics*, 60(5).

Ashford, D. A., Savage, H. M., Hajjeh, R. A., Mcready, J., Bartholomew, D. M., Spiegel, R. A., Vorndam, V., Clark, G. G., and Gubler, D. G. (2003). Outbreak of dengue fever in Palau, Western Pacific: Risk factors for infection. *American Journal Tropical Medicine Hygiene*, 69(2):135–140.

Aura, D. H. M. and Alfonso, J. R. M. (2010). Potential influence of climate variability on dengue incidence registered in a western pediatric Hospital of Venezuela. *Tropical Biomedicine*, 27(2):280–286.

Aziz, A. A. A., Ibrahim, H., and Abdullah, R. (2011). A model of knowledge mapping in visualizing the hotspot of dengue. *IEEE Conference on Open Systems(ICOS2011)*, pages 444–449.

Aziz, S., Aidil, R. M., Nisfarizal, M. N., Ngui, R., Lim, Y. A. L., Wan Yusoff, W. S., and Wan Ruslan, R. (2014). Spatial density of Aedes distribution in urban areas: A case study of breteau index in Kuala Lumpur, Malaysia. *Journal of Vector Borne Disease*, 51:91–96.

Aziz, S., Ngui, R., Lim, Y. A. L., Solehah, I., Nur Farhana, J., Azizan, A. S., and Wan Yusoff, W. S. (2012). Spatial pattern of 2009 dengue distribution in Kuala Lumpur using GIS application. *Tropical Biomedicine*, 29(1):113–120.

Azmawati, M. N., Aniza, I., and Ali, M. (2013). Evaluation of Communication for Behavioral Impact (COMBI) Program in Dengue Prevention : A Qualitative and Quantitative Study in Selangor, Malaysia. *Iranian Journal Public Health*, 42(5):538–539.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society*, 33(1):38–44.

Breslow, N. E. (1996). Generalized linear models checking assumptions and strengthening conclusions. *Congresso Nazionale Societa Italiana di Biometria*, pages 1–18.

Cabrera, M. (2013). *Spatio-temporal modelling of dengue fever in Zulia state, Venezuela*. PhD thesis, University of Bath, Department of Health, Bath, United Kingdom.

Carbajo, A. E., Schweigmann, N., Curto, S. I., de Garin, A., and Bejaran, R. (2001). Dengue transmission risk maps of Argentina. *Tropical Medicine and International Health*, 6(3):170–183.

Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S*. Chapman and Hall, London.

Chan, Y. C., Ho, B. C., and Chan, K. L. (1971). *Aedes aegypti*(L.) and *Aedes albopictus*(Skuse) in Singapore City. *Bulletin World Health Organization*, 44:651–657.

Chee, H. Y. and Abu Bakar, S. (2003). Phylogenetic investigation of dengue virus type 2 isolated in Malaysia. *Dengue Bulletin*, 27:100–107.

Chen, C. C. and Chang, H. C. (2013). Predicting dengue outbreaks using approximate entropy algorithm and pattern recognition. *Journal of Infection*, pages 1–7.

Chen, C. D., Nazni, W. A., Lee, H. L. Seleena, B., Mohd Masri, S., Chiang, Y. F., and Sofian Azirun, M. (2006). Mixed breeding of *Aedes aegypti*(L.) and *Aedes albopictus* skuse in four dengue endemic areas in Kuala Lumpur and Selangor, Malaysia. *Tropical Biomedicine*, 23(2):224–227.

Cheong, Y. L., Burkart, K., Leitao, P. J., and Lakes, T. (2013). Assessing Weather Effects on Dengue Disease in Malaysia. *International Journal of Environmental Research and Public Health*, 10(12):6319–6334.

Chew, M. H., Rahman, M. M., Jelip, J., Hassan, M. R., and Isahak, I. (2012). All serotypes of dengue viruses circulating in Kuala Lumpur, Malaysia. *Current Research Journal of Biological Sciences*, 4(2):229–234.

Choy, E. A., Mohd Khair, E., Atan, A., Sahani, M., and Mohd Ali, Z. (2011). Climate change and dengue: Case study in Seremban District, Negeri Sembilan, Malaysia. *Jurnal e-Bangi*, 6(1):38–48.

Chuang, P. H., Chuang, J. H., and Lin, I. F. (2010). A dynamic estimation of the daily cumulative cases during infectious disease surveillance: Application to dengue fever. *BMC Infectious Diseases*, 10(136):doi: 10.1186/1471–2334–10–136.

Cleveland, W. S. and Devlin, S. J. (1998). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). *Local regression models. Chapter 8 of Statistical Models in S.* Wadsworth and Brooks-Cole.

Connor, S. J., Omumbo, J., Green, C., DaSilva, J., Mantilla, G., Delacollette, C., Hales, S., Rogers, D., and Thomson, M. (2010). Health and Climate Needs. *Procedia Environmental Sciences*, 1(3):27–36.

Costello, A., Abbas, M., Allen, A., Ball, S., Bell, S., Bellamy, R., Friel, S., Groce, N., Johnson, A., Kett, M., Lee, M., Levy, C., Maslin, M., McCoy, D., McGuire, B., Montgomery, H., Napier, D., Pagel, C., Patel, J., Antonio, J., Oliveira, P., Redclift, N., Rees, H., Rogger, D., Scott, J., Stephenson, J., Twigg, J., Wolff, J., and Patterson, C. (2009). Managing the health effects of climate change. *www.thelancet.com*, 373:1693–1733.

Craven and Wahba (1979). Smoothing noisy data with spline functions. *Numerical Mathematics*, 31:377–403.

Cuadras, C. M. and Fortiana, J. (2002). *Distributions with given marginals and statistical modelling*. Springer.

Degallier, N., Favier, C., Menkes, C., Lengaigne, M., Ramalho, W. M., Souza, R., Servain, J., and Boulanger, J. P. (2010). Toward an early warning system for dengue prevention: Modeling climate impact on dengue transmission. *Climatic Change*, 98:581–592.

Dhang, C. C., Benjamin, S., Saranum, M. M., Fook, C. Y., Lim, L. H., Ahmad, N. W., and Azirun, M. S. (2005). Dengue vector surveillance in urban residential and settlement areas in Selangor, Malaysia. *Tropical Biomedicine*, 22(1):39–43.

Donald, S. S., Rosemary, L., Chiu, W. N., Eduardo, A. U., Yara, H., and Lucy, L. ad Laurent, C. (2012). Burden of dengue in Malaysia. Report from a Collaboration between Universities and the Ministry of Health Malaysia. *Brandeis University*, 42:1–78.

Earnest, A., Tan, S. B., Wilder-Smith, A., and Machin, D. (2012a). Comparing Statistical Models to Predict Dengue Fever Notifications. *Computational and Mathematical Methods in Medicine*, 2012:doi.org/10.1155/2012/758674.

Earnest, A., Tana, S. B., and Wilder Smith, A. (2012b). Meteorological factors and El Niño Southern Oscillation are independently associated with dengue infections. *Epidemiology and Infection*, 140(7):1244–1251.

Ebi, K. L., Hartman, J., Chan, N., McConnell, J., Schlesinger, M., and Weyant, J. (2005). Climate suitability for stable malaria transmission in Zimbabwe under different climate change scenarios. *Climate Change*, 73:375–393.

Economy, U. (2010). MDG1: Eradicate extreme poverty and hunger. *Malaysia: The Millennium Development Goals at 2010*, pages 10–30.

Emilie, A., Laurent, G., Beat, S., Francois, C., and Louis, L. (2011). Urbanisation and infectious diseases in a globalised world. *www.thelancet.com/infection*, 11.

Favier, C., Degallier, N., Rosa-Freitas, M. G., Boulanger, J. P., Costa Lima, J. R., Luitgards Moura, J. F., Menke's, C. E., Mondet, B., Oliveira, C., Weimann, E. T. S., and Tsouris, P. (2006). Early determination of the reproductive number for vector-borne diseases: The case of dengue in Brazil. *Tropical Medicine and International Health*, 11(3):332–340.

Fredolin, T. T. and Liew, J. (2004). Mechanisms of Malaysian rainfall anomalies. *Journal of Climate*, 17:3616–3622.

Gage, K. L., Burkot, T. R., Eisen, R. J., and Hayes, E. B. (2008). Climate and vectorborne disease. *American Journal of Preventive Medicine*, 35(5):436–450.

Gagnon, A. S., Bush, A. B. G., and Smoyer Tomic, K. E. (2001). Dengue epidemics and the El Niño Southern Oscillation. *Climate Research*, 19(1):35–43.

George, R. (1992). Current status of the knowledge of Dengue/DHF/DSS in Malaysia: Clinical Aspects. *Microbiology and Infectious Diseases*, pages 1–5.

Gharbil, M., Quenel, P., Gustave, J., Cassadou, S., Ruche, G., Girdary, L., and Marrama, L. (2011). Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. *BMC Infectious Diseases*, 11(166).

Ghebreyesus, T. A., Tadesse, Z., Jima, D., Bekele, E., Mihretie, A., Yihdego, Y. Y., Dinku, T., Connor, S. J., and Rogers, D. P. (2009). Using climate information in the health sector. *Field Actions Science Reports*, 2:63–67.

Glantz, M. H. (2001). *Currents of change: Impacts of El Niño and La Niña on climate and society*. Cambridge University Press.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, 11:480–496.

Gubler, D. J. (2002). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. *Trends in Microbiology*, 10(2):100–103.

Gubler, D. J. and Rosen, L. (1976). A simple technique for demonstrating transmission of dengue virus by mosquitoes without the use of vertebrate hosts. *The American Journal of Tropical Medicine and Hygiene*, 25(1):146–150.

Guzman, M. G. and Kouri, G. (2002). Dengue: An update. *The Lancet, Infectious Diseases*, 2(1):33–42.

Hafiz, H., Shamarina, S., and Nor, R. H. (2012). Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospatial Health*, 7(1):21–25.

Hairi, F., Ong, C. H. S., Suhaimi, A., Tsung, T. W., Anis Ahmad, M. A., Sundaraj, C., and M., S. M. (2003). A knowledge, attitude and practices (KAP) study on dengue among selected rural communities in the Kuala Kangsar district. *Asia Pacific Journal Public Health*, 15(1):37–43.

Hales, S., Wet, N. D., Maindonald, J., and Woodward, A. (2002). Potential effect of population and climate changes on global distribution of dengue fever: An empirical model. *The Lancet*, pages 1–5.

Hashizume, M., Dewan, A. M., Sunahara, T., Rahman, M. Z., and Yamamoto, T. (2012). Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: a time-series study. *BMC Infectious Diseases*, 12(98):1–9.

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3):297–318.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall.

Heil, A. and Goldammer, J. G. (2001). Smoke-haze pollution: A review of the 1997 episode in Southeast Asia. *Regional Environmental Change*, 2(1):24–37.

Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., and Rocklov, J. (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis*, 6(11):doi: 10.1371/journal.pntd.0001908.

Hirschman, C. (1987). The meaning and measurement of ethnicity in Malaysia: An analysis of census classifications. *The Journal of Asian Studies*, 46(3):555–582.

Hopp, J. M. and Foley, J. A. (2003). Worldwide fluctuations in dengue fever case related to climate variability. *Climate Research*, 25(1):85–94.

Hu, W., Clements, A., William, G., Tong, S., and Mengersen, K. (2012). Spatial patterns and socioecological drivers of dengue fever transmission in Queensland, Australia. *Environmental Health Perspectives*, 120(2):260–266.

Hurtado-Diaz, M., Riojas-Rodriguez, H., Rothenberg, S. J., Gomez-Dantes, H., and Cifuentes, E. (2007). Short Communication: Impact of climate variability on the incidence of dengue in Mexico. *Tropical Medicine and International Health*, 12(11):1327–1337.

Ibrahim, A., Mat Zin, N. A., and Sahari Ashaari, N. (2011). Simulation model for predicting dengue fever outbreak. *World Academy of Science, Engineering and Technology*, 5(11).

Ibrahim, F., Wan Abas, W. A. B., Chong Guan, C., and Sulaiman, S. (2007). A new approach to classify risk in dengue infection using bioelectrical impedance analysis. *Dengue Bulletin*, 31:58–74.

IPCC (2014). Fifth Assessment Report - Climate Change 2014 : Impacts, Adaptation, and Vulnerability. *Intergovernmental Panel Climate Change*.

Jaafar, S., Mohd Noh, K., Abdul Mutalib, K., Othman, N. H., and Healy, J. (2013). Malaysia Health System Review. *Health Systems in Transition*, 3(1):82–92.

Johansson, M. A., Cummings, D. A. T., and Glass, G. E. (2009a). Multiyear climate variability and dengue El Niño Southern Oscillation, weather and dengue incidence in Puerto Rico, Mexico and Thailand: A longitudinal data analysis. *PLoS Medicine*, 6(11):doi: 10.1371/journal.pmed.1000168.

Johansson, M. A., Dominici, F., and Glass, G. E. (2009b). Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Neglected Tropical Diseases*, 3(2):doi: 10.1371/journal.pntd.0000382.

Juneng, K. and Tanggang, F. T. (2008). Level and source of predictability of seasonal rainfall anomalies in Malaysia using canonical correlation analysis. *International Journal of Climatology*, 28:1255–1267.

Juni, M. H., Hayati, K. S., Cheng, C. M., Pyang, G. S., Nur Hayati, A. S., and Zaidatul Syeila, Z. A. (2015). Risk behaviour associated with dengue fever among rural population in Malaysia. *International Journal of Public Health and Clinical Sciences*, 2(1):114–127.

Jury, M. R. (2008). Climate influence on dengue epidemics in Puerto Rico. *International Journal of Environmental Health Research*, 18(5):323–334.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Roy, J., and Dennis, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471.

Khormi, H. M. and Kumar, L. (2011). Modelling dengue fever risk based on socioeconomic parameters, nationality and age groups: GIS and remote sensing based case study. *Science of Total Environment*, 409(22):4713–4719.

Khormi, H. M., Kumar, L., and Elzahrany, R. A. (2011). Modeling spatio-temporal risk changes in the incidence of dengue fever in Saudi Arabia: A geographical information system case study. *Geospatial Health*, 6(1):77–84.

Khun, K., Campbell Lendrum, D., Haines, A., Cox, J., Corvalan, C., and Anker, M. (2005). Using climate to predict infectious disease epidemic. *World Health Organization.*

Kovats, R. S., Bouma, M. J., and Haines, A. (1999). El Niño and Health. *Task Force on Climate and Health, World Health Organization, Geneva*, pages 1–46.

Krisada, L. and Lily, I. (2013). Generalized linear mixed models with spatial random effects for spatio-temporal data: An application to dengue fever mapping. *Journal of Mathematics and Statistics*, 9(2):137–143.

Kumarasamy, V. (2006). Dengue fever in Malaysia: Time for review? *Medical Journal Malaysia*, 61(1):1–3.

Kusriastuti, R. and Sutomo, S. (2005). Evolution of Dengue Prevention and Control Programme in Indonesia. *Dengue Bulletin*, 29:1–7.

Lam, S. K. (1993a). Strategies for dengue control in Malaysia. *Tropical Medicine*, 35(4):303–307.

Lam, S. K. (1993b). Two decades of dengue in Malaysia. *Tropical Medicine*, 35(4):195–200.

Lee, H. L., Rohani, A., Khadri, M. S., Nazni, W. A., Rozilawati, H., Nurulhusna, A. H., Nor Afizah, A. H., Roziah, A., Rosilawati, R., and Teh, C. H. (2015). Dengue Vector Control in Malaysia-Challenges and Recent Advances. *The International Medical Journal Malaysia*, 14(1):11–16.

Loevinsohn, M. E. (1994). Climatic warming and increased malaria incidence in Rwanda. *The LANCET*, 343:714–718.

Loh, B. and Song, R. J. (2001). Modeling dengue cluster size as a function of *Aedes aegypti* population and climate in Singapore. *Dengue Bulletin*, 25:74–78.

Lowe, R. (2010). *Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil*. PhD thesis, College of Engineering, Mathematics and Physical Science, University of Exeter, Harrison Building, College of Engineering, Mathematics and Physical Science, EX4 4QF, Exeter, United Kingdom.

Lowe, R., Bailey, T. C., Stephenson, D. B., Jupp, T. E., Graham, R. J., Barcellos, C., and Carvalho, M. S. (2013). The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in Medicine*, 32(5):864–883.

Lu, L., Lin, H., Tian, L., Yang, W., Sun, J., and Liu, Q. (2009). Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health*, 9:395–399.

Ma, S., Ooi, E. E., and Goh, K. T. (2008). Socioeconomic determinants of dengue incidence in Singapore. *Dengue Bulletin*, 32.

Mammen, M. P. J., Pimgate, C., Koenraadt, C. J. M., Rothman, A. L., Aldstadt, J., Nisalak, A., Jarman, R. G., Jones, J. W., Srikiatkhachorn, A., Ypil Butac, C. A., Getis, A., Thammapalo, S., Morrison, A. C., Libraty, D. H., Green, S., and Scott, T. W. (2008). Spatial and temporal clustering of dengue virus transmission in Thai Villages. *PLoS Medicine*, 5(11):1605–1615.

Mark, A. C. (2005). The evolution of El Niño, past and future. *Earth and Planetary Science Letters*, 230:227–240.

Markon, A. O. (2014). *Dengue fever in Ribeirao Preto, Brazil, 2003-2012: Patterns of Disease and Understanding of Prevention*. PhD thesis, University of Michigan, University of Michigan, United States.

Martens, W. J. M., Niessen, L. W., Rothmans, J., Jetten, T. H., and McMichael, A. J. (1995). Potential impact of global climate change on malaria risk. *Environmental Health Prospectives*, 103(5):458–464.

Mary Ann, B. (2009). Climate change could affect mosquito-borne diseases in Asia. *www.thelancet.com*, 373:1070.

Mathuros, T., Chi Tai, F., Suratsawadee, K., Chung Ming, L., and Chwan Chuen, K. (2009). Effects of the El Niño-Southern Oscillation on dengue epidemics in Thailand, 1996-2005. *BMC Public Health*, 9:422–436.

Mazrura, S., Rozita, H., Hidayatulfathi, O., Zainudin, M. A., Mohamad Naim, M. R., Nadia Atiqah, M. N., Rafeah, M. N., Er, A., Norela, S., Nurul Ashikin, Z., and Joy, J. P. (2010). Community Vulnerability On Dengue And Its Association With Climate Variability in Malaysia: A Public Health Approach. *Malaysian Journals of Public Health Medicine*, 10(2):25–34.

McCullagh, P. and Nelder, J. A. (1989). Generalized linear models. *Ecology*, 88(11).

Mia, M. S., Begum, R. A., Er, A. C., Abidin, R. D., and Pereira, J. J. (2013). Trends of dengue infections in Malaysia, 2000-2010. *Asian Pacific Journal of Tropical Medicine*, 6(6):462–466.

Ministry, H. M. (2008). Health Facts 2008 Malaysia. *Ministry of Health Malaysia*, pages 1–12.

Ministry, H. M. (2014). Health Facts 2014 Malaysia. *Ministry of Health Malaysia*, pages 1–12.

Mohamad Ismail, N. A., Kampan, N., Abdullah Mahdy, Z., Abdul Jamil, M., and Mohd Razi, Z. R. (2011). Dengue in pregnancy. *Southeast Asian Journal Trop Med Public Health*, 37(4):681–683.

Mohamad Mohsin, M. F., Abu Bakar, A., and Hamdan, A. R. (2013). The Preliminary Design of Outbreak Detection Model Based on Inspired Immune

System. *Third World Congress on Information and Communication Technologies (WICT)*, pages 189–195.

Mohd Ismail, M. (2007). Malaysia's Health 2007. *Kementerian Kesihatan Malaysia*, pages 1–299.

Mohd Ismail, M., Ahmad, K., Kamarul Zaman, M. I., Ramlee, R., Noor Hisham, A., Maimunah, A. H., Norain, A. T., Che Mohd Zain, C. A., Abd Rahim, M., and Khadzir, S. A. (2007). Laporan Tahunan 2007. *Kementerian Kesihatan Malaysia*, pages 1–293.

Mohd Ismail, M., Ahmad Shafii, S., Kamarul Zaman, M. I., Ramlee, R., Noor Hisham, A., Maimunah, A. H., Norain, A. T., Eisah, A. R., Abd Rahim, M., and Khadzir, S. A. (2009). Laporan Tahunan 2009. *Kementerian Kesihatan Malaysia*, pages 1–345.

Monath, T. P. (1994). Dengue: the risk to developed and developing countries. *Proceedings of the National Academy of Science of the United States of America*, 91(7):2395–2400.

Mondini, A. and Chiaravalloti, N. F. (2007). Socioeconomic variables and dengue transmission. *Rev Saude Publica*, 41(6):923–930.

Muhamed, K. and Haron, S. (2011). Poverty Mapping:An Effective Approach in Determining The Poor Area - Case Study of Johor. *Department of Statistics, Malaysia*, pages 43–60.

Muhammad Azami, N. A., Salleh, S. A., Neoh, H. M., Syed Zakaria, S. Z., and Jamal, R. (2011). Dengue epidemic in Malaysia: not a predominantly urban disease anymore. *BMC Research Notes*, 216(4):1–4.

Narwani, H., Jesni, J., Nyi Nyi, N., Hamzah, A. M., Abd Haris, M., and Mohd Noor, M. (2005). A review of dengue fever incidence in Kota Bharu, Kelantan, Malaysia during the years 1998-2003. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 36(5):1179–1186.

Natchaporn, P., Noparat, M., Siripen, K., and Wisit, C. (2006). Relationship between body size and severity of dengue hemorrhagic fever among children aged 0-14 years. *Southeast Asian Journal Tropical Medical Public Health*, 37(2):283–288.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384.

Nicol, J. (1997). Bioclimatic impacts of the 1994 smoke-haze event in Southeast Asia. *Atmospheric Environment*, 31(8):1209–1219.

Nor Azimi, Y. (2000). Dengue outbreak in 1998-Terengganu experience. *Buletin Kesihatan Masyarakat Isu Khas*, pages 75–84.

Oki, T. and Musiake, K. (1994). Seasonal change of the diurnal cycle of precipitattion over Japan and Malaysia. *American Meteorological Society*, 33:1445–1463.

Ooi, E. E. and Gubler, D. J. (2008). Dengue in Southeast Asia: epidemiological characteristics and strategic challenges in disease prevention. *Cadernos de Saude Publica*, 25:115–124.

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16(1):21–43.

Osman, O., Fong, M. Y., and Devi, S. (2007). A preliminary study of dengue infection in Brunei. *Journal Infectious Diseases*, 60:205–208.

Parks, W. J., Lloyd, L. S., Nathan, M. B., Hosein, E., Odugleh, A., Clark, G. G., Gubler, D. J., Palmer, K., Martin, J. L. S., Siversen, S. R., Dawkins, Z., and Renganathan, E. (2004). International experiences in social mobilization and communication for dengue prevention and control. *Dengue Bulletin*, 28:1–7.

Patz, J. A., Martens, W. J., Focks, D. A., and Jetten, T. H. (1998). Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environmental Health Perspectives*, 106(3):147–153.

Patz, J. A. and Olson, S. H. (2006). Climate change and health: Global to local influences on disease risk. *Annals of Tropical Medicine and Parasitology*, 100(5):535–549.

Poovaneswari, S. (1993). Dengue situation in Malaysia. *Malaysian Journal Pathology*, 15(1):3–7.

Pouliot, S. H., Xiong, X., Harville, E., Paz Soldan, V., Tomashek, K., Breart, G., and Buekens, P. (2010). Maternal dengue and pregnancy outcomes: A systematic review. *Obstetrical and Gynecological Survey*, 65(2):69–144.

R Core Team, D. (2010). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Racloz, V., Ramsey, R., Tong, S., and Hu, W. (2012). Surveillance of dengue fever virus: A review of epidemiological models and early warning systems. *PLoS Neglected Tropical Diseases*, 6(5):doi: 10.1371/journal.pntd.0001648.

Richard, B. and John, M. (2007). Overdispersion and Poisson Regression. *Social and Economic Sciences*, pages 1–24.

Rogers, D. P., Boulahya, M. S., Thompson, M. A., Connor, S. J., Dinku, T., Johm, K. B., Shalaby, H. R., Ahmadu, B., and Niang, A. (2008). National climate and environmental service for development: Proceedings of the International Symposium on PWS: A Key to Service Delivery. *World Meteorological Organization*, pages 87–92.

Rosnah, H., Za'aba, B., Abdul Rasid, K., Maimunah, A. R., Yogeswary, M., Mohd Jubri, J., Noorazah, D., Khairul Rafizah, H., Siti Yuhanis, I., Rafidah, S., Norhayati, M. S., Mas'ud, H., Rasyikah, M., Rajini, M., and Tg Mawar, T. A. A. (2009). Laporan Tahunan 2009. *Jabatan Kesihatan Negeri Selangor*, pages 1–180.

Rowley, W. A. and Graham, C. L. (1968). The effect of temperature and relative humidity on the flight performance of female *Aedes aegypti*. *Journal Insect Physiol.*, 14:1251–1257.

Rozhan, S., Jamsiah, M., Rahimah, A., and Ang, K. T. (2006). The COMBI (Communication for Behavioral Impact) Program in the prevention and control of dengue-The Hulu Langat experience. *Jurnal Kesihatan Masyarakat*, 12(1):19–32.

Rozilawati, H., Zairi, J., and Adanan, C. R. (2007). Seasonal abundance of *Aedes albopictus* in selected urban and suburban areas in Penang, Malaysia. *Tropical Biomedicine*, 24(1):83–94.

Rozita, H., Hidayatulfathi, O., Nurul Azimah, J., Mazrura, S., Kamarulismail, U., Zainudin, M. A., Er, A. C., and Zailiza, S. (2013). The COMBI Approach in Managing Dengue Cases in an Urban Residential Area, Nilai, Malaysia. *International Journal of Public Health Research*, 3(2):347–352.

Rudnick, A., Tan, E. E., Lucas, J. K., and Omar, M. (1965). Mosquito-borne haemorrhagic fever in Malaya. *British Medical Journal*, 1(5445):1269–1272.

Russell, R. C., Currie, B. J., Lindsay, M. D., Mackenzie, J. S., Ritchie, S. A., and Whelan, P. I. (2009). Dengue and climate change in Australia: Predictions for the future should incorporate knowledge from the past. *The Medical Journal of Australia*, 190(5):265–268.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1988). *Akaike Information Criterion Statistics*. KTK Scientific Publisher.

Schreiber, K. V. (2001). An investigation of relationships between climate and dengue using a water budgeting technique. *International Journal of Biometeorology*, 45(2):81–89.

Schwartz, E. (2002). Study of dengue fever among Israeli travellers to Thailand. *Dengue Bulletin*, 26:162–167.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Seng, S. B., Chong, A. K., and Moore, A. (2005). Geostatistical modelling, analysis and mapping of epidemiology of dengue fever in Johor State, Malaysia. *The 17th Annual Colloquium of the Spatial Information Research Centre*, pages 1–14.

Setyono, A., Dian, N., Alam, M. J., and Soosai, B. A. M. (2011). Study and design framework for e-Dengue teleconsultation system. *Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, pages 293–297.

Shaharudin, I., Shamsul, A. S., Tahir, A., Mariam, M., Azah, D., and Nik Shamsidah, N. I. (2002). Sistem maklumat geografi (GIS) dan sektor kesihatan awam: Kajian demam denggi di Bandar Baru Bangi dan Kajang. *Jurnal Kesihatan Masyarakat Isu Khas*.

Shekhar, K. C. and Huat, O. L. (1992). Epidemiology of dengue/dengue hemorrhagic fever in Malaysia-A retrospective epidemiological study 1973-1987. Part I: Dengue hemorrhagic fever (DHF). *Asia Pacific Journal of Public Health*, 6(2):15–25.

Simões, T. C., Codeco, C. T., Nobre, A. A., and Eiras, A. E. (2013). Modeling the non-stationary climate dependent temporal dynamics of *Aedes Aegpti*. *PLoS ONE*, 8(8):1–10.

Simon, P. W., Matthew, G. K., and Fred, L. M. (2003). *Statistical and econometric methods for transportation data analysis.* Chapman and Hall/Crc.

Simon, W. (2006). *Generalized Additive Models: An introduction with R.* Chapman and Hall.

Singh, N. P., Jhamb, R., Agarwal, S. K., Gaiha, M., Dewan, R., Daga, M. K., Chakravarti, A., and Kumar, S. (2005). The 2003 outbreak of dengue fever in Delhi India. *Southeast Asian Journal Tropical Medical Public Health*, 36(5):1174–1178.

Siqueira, J. B., Martelli, C. M. T., Coelho, G. E., Rocha Simplicio, A. C., and Hatch, D. L. (2005). Dengue and dengue hemorrhagic fever, Brazil, 1981–2002. *Emerging Infectious Diseases*, 11(1):48–53.

Siqueira, J. J. B., Maciel, I. J., Barcellos, C., Souza, W. V., Carvalho, M. S., Nascimento, N. E., Oliveira, R. M., Morais-Neto, O., and Martelli, C. M. T. (2008). Spatial point analysis based on dengue surveys at household level in Central Brazil. *BMC Public Health*, 8:361–369.

Siwar, C., Ahmed, F., and Mohd Idris, N. D. (2013). Relationship between Poverty and Socio-demographic Characteristics of Poor Households : A Case Study in Kelantan, Malaysia. *2013 International Conference on Advances in Social Science, Humanities and Management (ASSHM 2013)*, 46:679–684.

Skae, F. M. T. (1902). Dengue fever in Penang. *British Medical Journal*, 2(2185):1581–1583.

Smith, C. E. G. (1957). A localized outbreak of dengue fever in Kuala Lumpur: Serological aspects. *The Journal of Hygiene*, 55(2):207–223.

Sokrin, K. and Lenore, M. (2007). Health seeking and access to care for children with suspected dengue in Cambodia: An ethnographics study. *BMC Public Health*, 7:262–271.

Soman, M. K. and Slingo, J. (1997). Sensitivity of the Asian summer monsoon to aspects of sea surface temperature anomalies in the tropical Pacific ocean. *Quarterly Journal of the Royal Meteorological Society*, 123(538):309–336.

Souza, S. S., Silva, I. G., and Silva, H. H. (2010). Association between dengue incidence, rainfall and larval density of *Aedes* aegypti, in the State of Goias. *Revista da Sociedade Brasileira Medicina Tropical*, 43(2):152–155.

Stefan, M., Eng, E. O., and Kee, T. G. (2008). Socioeconomic determinants of dengue incidence in Singapore. *Dengue Bulletin*, 32:17–28.

Strickman, D. and Kittayapong, P. (2003). Dengue and its vector in Thailand: Calculated transmission risk from total pupae counts of *Aedes aegypti* and association of wing-length measurement with aspects of the larval habitat. *The American Society of Tropical Medicine and Hygiene*, 68(2):209–217.

Suhaila, J., Deni, S. M., Wan Zin, W. Z., and Jemain, A. A. (2010a). Spatial patterns and trends of daily rainfall regime in Peninsular Malaysia during the southwest and northeast monsoons: 1975-2004. *Meteorology and Atmospheric Physics*, 110(1):doi–10.1007/s00703–010–0108–6.

Suhaila, J., Deni, S. M., Wan Zin, W. Z., and Jemain, A. A. (2010b). Trends in Peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoon seasons: 1975-2004. *Sains Malaysiana*, 39(4):533–542.

Suhaili, M. R., Hosein, E., Mokhtar, Z., Ali, N., Palmer, K., and Isa, M. M. (2004). Applying Communication-for-Behavioural-Impact (COMBI) in the prevention and control of dengue in Johor Bahru, Johore, Malaysia. *Dengue Bulletin*, 28:39–43.

Tanggang, F., Latif, M. T., and Juneng, L. (2010). The roles of climate variability and climate change on smoke haze occurrences in Southeast Asia region . *Climate change: Is Southeast Asia up to the challenge*, 32(5):36–49.

Tanggang, F. T. and Bahari, A. (2002). ENSO influences on precipitation and air temperature variability in Malaysia. *Proceedings of the Regional Symposium on Environment and Natural Resources*, 1:124–131.

Teixeira, M. G., Costa, M. C. N., Guerra, Z., and Barreto, M. L. (2006). Dengue in Brazil: Situation-2001 and trends. *Dengue Bulletin*, 26:70–76.

Teng, A. K. and Singh, S. (2001). Epidemiology and new initiatives in the prevention and control of dengue in Malaysia. *Dengue Bulletin*, 25:7–14.

Thammapalo, S., Chongsuwiwatwong, V., Mc Neil, D., and Geater, A. (2005). The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. *Southeast Asian Journal Tropical Medicine Public Health*, 36(1):191–196.

Thomas, W. Y. and Neil, D. M. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2(5):587–602.

Tren, R. and Bate, R. (2001). *Malaria and the DDT Story.* The Institute of Economic Affairs 2001.

Wahba (1990). *Spline models of observational data.* SIAM.

Wan Abd Raof, W. R. (2010). *Statistics Handbook Malaysia 2010.* Department of Statistics Malaysia, Pesuruhjaya Banci, Malaysia.

Wan Abd Raof, W. R. (2011). *Population Distribution and Basic Demographic Characteristics.* Department of Statistics Malaysia, Pesuruhjaya Banci, Malaysia.

Wan Fairos, W. Y., Wan Azaki, W. H., Mohamad Alias, L., and Bee Wah, Y. (2010). Modelling dengue fever (DF) and dengue haemorrhagic fever (DHF) outbreak using poisson and negative binomial model. *World Academy of Science, Engineering and Technology*, 62:903–908.

Wan Norafikah, O., Chen, C. D., Soh, H. N., Lee, H. L., Nazni, W. A., and Sofian Azirun, M. (2009). Surveillance of *Aedes* mosquitoes in a university campus in Kuala Lumpur, Malaysia. *Tropical Biomedicine*, 26(2):206–215.

Wan Rozita, W. M., Yap, B. W., Veronica, S., Muhammad, A. K., Lim, K. H., and Sumarni, M. G. (2006). Knowledge, attitude and practices (KAP) survey on dengue fever in an urban malay residental area in Kuala Lumpur. *Malaysian Journal of Public Health Medicine*, 6(2):62–67.

Webster, P. J. and Palmer, T. N. (1997). The past and future of El Niño. *Nature*, 390:562–564.

WHO (2012a). Communication for behavioural impact (COMBI): A toolkit for behavioural and social communication in outbreak response. *Geneva: World Health Organization (WHO).*

WHO (2012b). Dengue and severe dengue. *Fact sheet.*

Wolf, P. S., Motoi, S., Vu, D. T., Richard, G. W., Ataru, T., Lay, M. Y., Hideki, Y., Ubydul, H., Le, H. T., Dang, D. A., and Koya, A. (2011). Population density, water supply, and the risk of dengue fever in Vietnam: Cohort study and spatial analysis. *PLoS Med*, 8(8):1–10.

Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M., and Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia. *Hydrol. Earth Syst. Sci.*, 6:5471–5503.

Wu, P. C., Guo, H. R., Lung, S. C., Lin, C. Y., and Su, H. J. (2007). Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Tropica*, 103(1):50–57.

Yap, H. H., Chong, N. L., Foo, A. E. S., and Lee, C. Y. (1994). Dengue vector vontrol: Present status and future prospects. *Kaohsiung Journal Medical Science*, 10:S102–S108.

Yara, A. H., Vishal, D., Narendra, A., Tyagi, B. K., Deoki, N., and Donald, S. S. (2011). Overcoming data limitations: Design of a multicomponent study for estimating the economic burden of dengue in India. *Dengue Bulletin*, 35:1–14.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R.* Springer Science, New York.