2014

Evaluation Of Environmental Factors And Dengue Fever In Sri Lanka Using Geospatial Tools

Meghan Tipre
University of Alabama at Birmingham

EVALUATION OF ENVIRONMENTAL FACTORS AND DENGUE FEVER IN SRI LANKA USING GEOSPATIAL TOOLS

by

MEGHAN TIPRE, BDS, MSPH

NALINI SATHIAKUMAR, COMMITTEE CHAIR
LESLIE McCLURE
AKHLAQUE HAQUE
JEFFERY LUVALL
BEN ZAITCHIK

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Public Health

BIRMINGHAM, ALABAMA

2014

EVALUATION OF ENVIRONMENTAL FACTORS AND DENGUE FEVER IN SRI- LANKA USING GEOSPATIAL TOOLS

MEGHAN TIPRE

DOCTOR OF PUBLIC HEALTH

ABSTRACT

The incidence of dengue fever has increased exponentially in Sri Lanka, from 24.4 cases per 100,000 in 2003 to 165.3 per 100,000 population in 2013. Despite concerted effort by the Sri Lankan government, dengue control remains a challenge in the country, indicating a need for novel approach for dengue prevention and control.

The aim of this research was to identify environmental risk factors that may be associated with dengue incidence rate at the Gram Niladhari Divisions level (smallest administrative unit) in Colombo city, Sri Lanka, using geospatial tools such as remote sensing and geographic information system. These factors included climate variables, land cover classes, population demographics and housing characteristics. Data on potential risk factors were obtained from several sources including remotely sensed data, meteorology department and publicly available census data. Data on dengue cases were obtained from Colombo municipal council, Department of Public Health for period between 2005 and 2011. The analysis was conducted in two parts: part I evaluated the relationship between mean

temperature and cumulative rainfall and dengue incidence, using generalized linear regression; part II evaluated the relationship between local environmental factors such as land cover, population and housing characteristics and dengue incidence rate, using spatial (geographic weighted regression) and non-spatial analysis (Poisson regression).

The results of the study found that: 1) there is a weak association between weekly temperature and rainfall and increased risk of dengue; 2) determined spatial-temporal risk of dengue incidence rate using the Getis-Ord Gi* statistic; and 3) local environmental factors such as decreased piped-water supply, increased brick-walled housing, decreased housing density and increased vegetation were significantly associated with high incidence of dengue fever cases.

Results of the study are in agreement with several other studies conducted previously however, they are novel in Sri Lanka. The application geospatial tools for data management and analysis contributes to the evidence supporting routine use of these tools in dengue research. The study paves way for future work using methods applied in this study to targeted interventions to prevent and control dengue in high risk areas in the country and in the region.

Keywords: dengue, Sri Lanka, Remote sensing, Geographic information systems

DEDICATION

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, who have always been the pillars of my strength in all my endeavors, my in-laws for their encouragement, and to my dearest husband for his patience and support in every step of the way.

ACKNOWLEDGEMENTS

I am grateful to the following individuals who were instrumental in the successful completion of this project

- Dr. Nalini Sathiakumar, my advisor and mentor, for giving me the opportunity to work with her, for her interminable support and guidance throughout my graduate studies.
- Drs. Haque, Luvall, McClure, and Zaitchik, for their support, encouragement and many valuable contributions.
- Dr. Haque for his immense help with ArcGIS software.
- Dr. Luvall for providing the satellite data and for many insightful discussions on remote sensing.
- Drs. Sreelatha Meleth and Asha Kamath for their help with data analysis.
- My friends and colleagues, Mark Leader, Tarun Arora, Ilene Brill and Sumal Nandasena for their help with SAS coding and data management.
- Dr. Pradeep Kariyawasam, Director, Department of Public Health, Colombo Municipal Council, Sri Lanka, for providing valuable data, without which this study would not have been possible.
- International Training and Research in Environmental and Occupational Health program for supporting this research (5D43TW005750-10)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AICc	Akaike Information criterion
CFR	Case Fatality Rate
CMC	Colombo municipal council
DEN	Dengue virus
DF	Dengue Fever
DHF	Dengue hemorrhagic fever
DS	Divisional Secretariats
DSS	Dengue shock syndrome
DSSS	Spatial decision support systems
GIS	Geographic information systems
GND	Gram Niladari Divisions
GWR	Geographic weighted regression
IDW	Inverse Distance Weighting
LDAS	Land Data Assimilation System
LISA	Local Indicators of Spatial Association

LULC	Land cover land use
MLC	Maximum Likelihood algorithm
NDCU	National Dengue Control Unit
NIR	Near Infra-red

1.0 INTRODUCTION

Dengue fever is one of the most important vector-borne diseases in the world, with significant economic, political, and social impact (Kyle and Harris, 2008). It has become a major public health concern, with implications for health security due to disruption, and rapid epidemic spread beyond national borders (Guzman et al., 2010; WHO, 2009). In the last 50 years, incidence of dengue fever has increased 30-fold, with the disease being endemic in more than 100 countries in Africa, the Americas, the eastern Mediterranean, South-east Asia, and the Western Pacific (WHO (2014b)). Over 2.5 billion people are at risk from dengue fever, constituting about 40% of the world's population. An estimated 100 million infections occur each year that are symptomatic, but only a fraction (10%) of the cases are reported. About 500,000 of reported cases acquire dengue hemorrhagic fever and dengue shock syndrome which are serious forms of dengue fever. The case fatality rate (CFR) of dengue fever is about 1%; however, in cases with DHF and DSS, the CFR can rise more than 5%. If DHF and DSS remain untreated the CFR can rise up to 50% (Kyle and Harris (2008); WHO (2014a)). There is no specific treatment for dengue fever or DHF, except supportive care, including fluid replacement and pain management (Murray et al., 2013; WHO, 2009).

In the absence of an effective vaccine, the dengue prevention and control mainly depends on early detection and management of dengue cases and effective vector control management strategies (Murray et al. 2013; WHO 2009, 2012). However, current surveillance systems and control strategies used in several countries have had only a transient and limited effect on disease incidence (Wilder-Smith et al., 2012). The growing threat of dengue worldwide has led many agencies and institutions, including World Health Organization (WHO), to call for a development of comprehensive early warning and surveillance system, which has predictive capability for dengue epidemics. Also the need for exploration of new and innovative tools for laboratory diagnosis and vector monitoring has been emphasized (Racloz et al., 2012). In recent years, mapping methods have been tried in terms of forecasting risk zones for vector-borne diseases, through the geographic information systems (GIS) and the use of satellite-based data (Bergquist, 2011). Dengue studies from South-east Asia and the Americas have successfully shown the use of GIS and remote sensing to analyze spatial-temporal patterns and relationship between the environmental factors and dengue incidence rate. However, the science is still in its nascent stage and the uses of such novel technologies are limited to few countries or institutions.

1.1 EPIDEMIOLOGY OF DENGUE FEVER

Dengue fever is a flu-like illness that affects infants, children and adults (WHO 2014a). It often gets misdiagnosed, similar to other tropical diseases, or even goes unrecognized. Dengue fever is characterized by high fever ($40°C/104°F$), severe headache, pain behind the eyes, muscles and joint pain, nausea, vomiting, swollen glands, or rashes. Symptoms usually appear after 4-10 days of incubation period following a bite from an infected mosquito. Dengue fever seldom causes death, however, severe forms of dengue fever including DHF and DSS, can result in fatal complications due to plasma leaking, fluid accumulation, respiratory illness, severe bleeding or organ impairment (WHO 2014a).

1.1.1 Dengue virus

Dengue fever is caused by four distinct, but closely related serotypes of a Flavivirus virus: DEN-1, DEN-2, DEN-3, and DEN-4. A recovery from the infection from one of the serotypes provides life-long immunity against that particular serotype, but only provides temporary, and at times partial immunity to other three serotypes. Thus, a person can be infected with dengue four different times. Epidemiological evidence also suggest that a person's risk of developing DHF increases with subsequent infections (WHO 2014a). It is also suggested infants born to mothers, immune to dengue virus are at high risk to develop more severe form of dengue during a primary infection (Rodenhuis-Zybert et al., 2010).

1.1.2 Dengue vectors

Dengue virus is transmitted to humans through the bite of infected female *Aedes* mosquitoes, principally *Ae. Aegypti* (WHO, 2009). This mosquito is a tropical and subtropical species, predominantly found between latitudes 35°N and 35°S. It is a small, black and white in color, and highly domesticated mosquito that prefers artificial containers commonly found in and around homes, for breeding. Other breeding places include flower vases, old automobile tires, buckets that collect rainwater, earthenware jars, metal drums, concrete cisterns, and trash in general. Other species of *Aedes* mosquito, including *Ae. Albopictus, Ae. polynesiensis* and several species of *Ae. scutellaris* complex, have also been known to cause dengue outbreaks. Each of these species has a specific ecology, behavior, and geographical distribution. *Ae. Albopictus,* a secondary vector for dengue fever, has spread to Europe and the Americas. This mosquito has been known to adapt to the environment and the eggs can survive is cooler climates.

Ae. Aegypti is one of the most efficient vectors for arboviruses, because it is highly anthropophilic, frequently bites several times before completing oogenesis, and thrives in close proximity to humans. Typically, after being bitten, the virus undergoes an incubation period of 3 to 14 days (average of 4 to 7 days), after which the person may experience acute onset of fever accompanied by a variety of symptoms associated with dengue fever. This results in the virus circulating in the peripheral blood, a state of viremia, that can last from 5 to 12 days. During this time, if an uninfected *Aedes* mosquito bites an infected person, the mosquito becomes infected.

Once the virus enters the mosquito's system in the blood meal, the virus spreads through the body of the mosquito over a period of 8-12 days. The mosquito once infected remains infected for its entire life. The life span of a mosquito is about three to four weeks, during which time, an infected mosquito can infect several healthy persons. Only the female mosquitoes of the species bites humans for blood meals. During the bite, the mosquito injects saliva to prevent blood from clotting during the feeding. The dengue virus enters the human blood stream through this saliva.

The complete lifecycle of an *Ae. Aegypti* including the stages of an egg, larvae, pupae and adult stage takes about 4 to 6 six weeks. Several environmental factors can accelerate or delay the growth of the mosquito to its adult stage or the time taken for the virus to mature in the body of the mosquito. These factors are discussed below.

1.2 ENVIRONMENTAL FACTORS AFFECTING TRANSMISSION OF DENGUE VIRUS

The dynamics of transmission of dengue is complex, and is affected by the interactions between several factors, including environmental and climate factors affecting the abundance of the adult vector per human population; the host (human)-vector (mosquito) interactions such as human biting rate of the mosquito, proportion of bites that would produce infections, average duration of the infection in humans; and herd immunity (to the dengue viruses) in the human population (Focks et al., 1995; Pham et al., 2011). Dengue disease is, therefore, strongly influenced by several environmental and ecological factors, both at the macro level and micro levels (Quintero et al., 2014).

1.2.1 Role of temperature and rainfall on dengue transmission

The survival of the dengue mosquitoes and the dengue virus requires optimal environmental conditions, including temperature, precipitation, and to a lesser extend, sea level, elevation, wind, and daylight duration (Alto and Bettinardi, 2013). Elevated temperatures, relative humidity, and rainfall in endemic areas have been significantly associated with increased number of dengue cases. Higher temperatures accelerate the development rate of the virus by reducing incubation period, and relative humidity increases mosquito biting rate, as well as decrease the adult mosquito mortality (Naish et al., 2014; Pham et al., 2011). Studies have also found that, elevated temperatures at various lag time periods ranging from 0 lag to up to 20 weeks, increased dengue incidence (Hii et al., 2009, 2012). Rainfall, however has contrasting effects on dengue transmission. Heavy rains may flush away eggs and larvae and pupae from containers in a short term. But, collection of rainwater following a heavy rainfall provides suitable breeding grounds for the mosquitoes (Sarfraz et al., 2012). Dry weather or drought in endemic areas necessitated the storage of water in containers, which then may become breeding sites for *Ae. Aegypti.*

A number of studies have analyzed the effects of temperature and rainfall on dengue incidence using different models, including generalized linear models, time series, Bayesian, and non-linear models (Naish et al., 2014). Dengue transmission in endemic settings is characterized by non-linear dynamics, with strong seasonality, multi-annual oscillations and non-stationary temporal variations. Seasonal and multi-annual cycles in dengue incidence vary over time and space. Besides seasonality of dengue transmission, periodic epidemics and more irregular intervals of outbreaks are commonly observed. Overall, these studies consistently showed a positive association between the climate variables and dengue, but the relationships varies in different settings.

1.2.2 Role of local environmental factors on dengue transmission

Along with climate variables, local environmental factors, such as the type of housing, housing density, and peri-urban and peri-domestic areas can provide favorable conditions for the *Ae. Aegypti* breeding (Arunachalam et al., 2010; Braga et al., 2010; Ooi et al., 2006; Van Benthem et al., 2005; Vanwambeke et al., 2006; **?**). A study conducted in Thailand found that sparse vegetation in an urban area, low altitude, good transportation routes and rapid and unplanned urban development favor breeding of *Ae. Aegypti* (Van Benthem et al., 2005). A study in the US, found that older homes may provide breeding grounds for *Ae. Aegypti* (Walker et al., 2011). Another study conducted in South-western India found that housing patterns such as closely packed terraced houses were associated with increased incidence of dengue cases (Fulmali et al., 2008; Schmidt et al., 2011). Other studies have found that houses with gardens containing tires, plants with temporary water pools, rainwater gutters, as well as construction sites are favorite breeding grounds of dengue mosquitoes (Ashford et al., 2003; Heukelbach et al., 2001; Kholedi et al., 2012).

1.2.3 Role of socio-demographic factors on dengue transmission

Changes in social and demographic factors, such as increase in population density, unplanned urbanization, etc. can increase dengue transmission. Sudden increase in urbanization result in crowded neighborhoods and increased construction sites which may become breeding sites for dengue vectors (Vanwambeke et al., 2007. Other factors, such as civic services, including poor sanitation and waste disposal, economic status of the populations, human behaviors and education, also play important roles in spread of dengue. Other studies have found that lower socio-economic urban communities with poor sanitation and waste-disposal, and limited tap-water supply are positively associated with abundance of dengue vector species (Bhandari et al., 2008; Bowman et al., 2014). Due to lack of personal level data for these socio-demographic factors, researchers have used surrogate measures to quantify these factors. For example, Khormi and Kumar (2011) used satellite data to identify inhabited areas in Jeddah County, Saudi Arabia, to estimate population density. They also used house-sizes and street widths to determine the neighborhood quality. They found that, areas with higher population density and poor neighborhood quality were associated with higher numbers of dengue cases. Other demographic factors including higher education levels in the population, which may influence health related habits and increase awareness about prevention and control of breeding sites in homes, were negatively associated with dengue incidence in Vietnam (Schmidt et al., 2011). Similarly, implementation and monitoring of prevention , health education and vector-control programs by the local health authorities negatively impact dengue transmissions.

In sum, the studies above clearly demonstrate the complexity of several socio-environmental factors interacting with each other, to impact the distribution and abundance of dengue vectors and its transmission. It is also clear that the dynamics between the risk factors and dengue incidence vary from country to country. Thus, in order to develop models for predicting dengue risks for a country or even a smaller region within a country, country-specific or region specific risk factors of dengue are needed. Several environmental determinants of dengue fever can be spatially mapped, measured, and quantified with newer technologies specifically remote sensing and GIS. These spatially mapped variables can be incorporated in risk prediction models used for early warning systems as well for targeting surveillance efforts and control measures.

1.3 ROLE OF GIS AND REMOTE SENSING IN DENGUE RESEARCH

Remote sensing and GIS are geospatial tools that have been increasingly used for spatial mapping of vector-borne diseases; and for identifying and evaluating spatial relationships between environmental risk factors and the spread of these diseases. Remote sensing is the science and art of obtaining information about an object, area, or phenomenon, through the analysis of data acquired by a device that is not in contact with the subject under investigation. Remote sensing sensors, such as earth-observing satellites, acquire data of various earth surface features that emit and reflect electromagnetic energy, and the data is analyzed to provide information about the resources under investigation. The data are then displayed in form of digital images, commonly known as satellite imagery or aerial imagery. In order to make use of the data from the satellites, the images are analyzed and processed to extract underlying information of the earth's surface that was digitally photographed. This information could include land cover class such as lakes, mountains, tree cover, information on land cover use such as roads, residential houses, industrial parks, and other characteristics such as surface temperature, soil moisture, vegetation indices, etc. Remotely sensed data, both historical and real-time, can generate spatial maps of environmental characteristics such as land-cover, land-use, elevations, surface temperatures, and rainfall. Such maps can be used to visualize spatial-temporal changes over time in a geographic area or population using a GIS software.

GIS is a technique designed to capture, manage, analyze, and display all forms of data that can be linked by common geographical coordinate system. GIS can be used for visualization of spread of dengue cases and its vector as well as for mapping dengue vector habitats , abundance and density. It also enables mapping changes in the spread of the disease over time, and identifying the spatial-temporal relationships between environmental risk factors and dengue incidence.

Furthermore, GIS can be used to identify drivers of dengue in the initial stages of analysis, by visually evaluating the relationship between the variables, and linking them geographically. GIS has also been applied in studies to conduct exploratory analyses to predict outbreak zones using cluster detection algorithms such as Kulldorf's spatial and space time scan statistics, hotspot analyses, or advance mathematical modeling such as geographically weighted regression (GWR). Such models can be integrated with disease surveillance system which can target surveillance and control measures in areas with higher risk of dengue occurrence as opposed to random surveillance. Finally, GIS-based platforms have been used to develop spatial decision support systems (DSSS), designed to enhance decisions at various stages of planning process, to produce the most effective results, in terms of resources allocation and disease control. Such systems allow for the incorporation of a wide range of data, from entomological surveillance, dengue case surveillance, vector and disease control intervention monitoring, and stock control (Eisen and Eisen, 2011). Using GIS and other reporting tools, these DSS provide wide range of outputs tables, graphs, and maps. This information can easily be interpreted by stakeholders and policy makers to make informed decisions regarding implementation, monitoring, and evaluation of prevention programs.

Combined with mathematical analyses toolkits, remote sensing and GIS can become powerful tools to analyzing and predicting spatio-temporal patters of vector-borne diseases like dengue. Such analysis can inform risk assessment and prevention.

1.4 DENGUE IN SRI LANKA

Dengue is a major health problem in several Asian countries, where the disease has become the leading cause of hospitalizations and death among children. As a result of rapidly growing population, unplanned urbanization, poor water storage and unsatisfactory sanitation conditions, countries such as Sri Lanka, India, and Bangladesh, have observed sudden increase in the number dengue outbreaks and epidemics in the last two decades (Raheel et al., 2011; WHO-SEARO and WHO-WPRO, 2008). Existing surveillance systems including vector surveillance has had little effect in curbing dengue epidemics in these countries.

Before the 1989, epidemiology of dengue fever in Sri Lanka was characterized by low incidence of dengue fever and dengue hemorrhagic fever cases. The four dengue virus serotypes including DENV-1, DENV-2, DENV-3, and DENV-4 have been circulating in Sri Lanka for over 30 years (Messer et al., 2002). The main vectors of dengue transmission in Sri Lanka are *Ae. Aegypti* and *Ae. Albopictus*. After 1989, there has been a dramatic increase in dengue incidence in Sri Lanka, that have led several major epidemics in the last decade. Scientists trying to understand this sudden emergence of dengue have attributed it to increasing urban population, poor water supply and waste disposal practices, traditional water storage facilities, and changing lifestyles. Some genetic studies have attributed the emergence to genetics changes in DENV-3 serotypes (Kanakaratne et al., 2009). Also, climate plays an important role in intensity and severity of epidemics in Sri Lanka. The disease is endemic all year round, with peak incidence during the middle of the tear (May-June-July) following the South West monsoon rains, and towards the end of the year (October to January) following the North East monsoon rains (Sirisena and Noordeen, 2014).

1.4.1 Previous studies on dengue in Sri Lanka.

Several studies have discussed dengue in Sri Lanka, however, very few have evaluated role of environmental factors in the increase of the disease incidence. A study conducted in Sri Lanka found a potential impact of global climate change on the disease transmission by increasing the salinity tolerance of mosquito vectors in coastal regions (Ramasamy et al., 2011). *Ae. Aegypti* and *Ae. albopictus* larvae were found in brackish water with salinity up to 15 parts per thousand in discarded plastic and glass containers, abandoned fishing boats and unused wells in coastal peri-urban areas in Jaffna and Batticola districts in norther parts of Sri Lanka (Surendran et al., 2007b). The results indicate that Aedes mosquitoes which usually are found in freshwater collections can also be found in brackish waters. Results from a study conducted as part of a multi-country study found a positive temporal association between rainfall and the number of laboratory confirmed dengue cases reported (Arunachalam et al., 2010).The overall results of the study found that the number of pupae in household containers showed a strong positive association with the presence of shrubbery above the container, as well as the lack of use of the container for the previous 7 days or more, and the complete or partial absence of a container cover. This study also reported a negative association between piped-water supply in households and increased incidence of dengue cases but the results were not significant.

An important report summary by WHO, assessing the epidemiology of dengue in Sri Lanka in relation to intervention measures, found a decrease in dengue incidence in 2011 as compared to previous years, following interventions by the ‘Presidential Task Force for Dengue Control’. However, the report also suggested that the possible risk factors for resurgence are i) changing rainfall pattern, ii) abundance of various vector breeding habitats, mainly disposed plastic containers, plastic sheets, roof gutters. etc.

iii) existence of natural habitats such as tree holes, large land mass of Sri Lanka being covered with vegetation, and iv) human behaviors with reference to waste water and solid disposal management.

1.4.2 Prevention and control of dengue in Sri Lanka

The government in Sri Lanka has taken major steps towards prevention and control of dengue. A National Dengue Control Unit (NDCU) was established in 2005 by the Ministry of Health in Sri Lanka, following the major epidemic in 2004. Following the Asia-Pacific Dengue Strategic Plan (2008-2016) established by the WHO, the Ministry of Health, established a National Strategic Plan for Dengue Prevention and Control for 2011 to 2015. The plan outlined six strategies, including vector surveillance and integrated vector management, disease surveillance, case management, social mobilization; and inter-sectoral coordination, outbreak response, and research. The vector surveillance included sentinel site surveys and environmental surveys; integrated vector management emphasized on social mobilization and environmental sanitation for sustainable vector control. The disease surveillance included passive surveillance or the routine reporting of suspected or confirmed cases, special surveillance to obtain information on clinical presentation, severity and outcome of dengue and dengue hemorrhagic cases, establishment of sentinel site surveillance and early warning system. The plan also recommended promotion of E-base surveillance that used email based reporting in parallel to existing passive surveillance. The strategic plan also identified five key areas of research needed in the country to support control and prevention of dengue. These five areas included information on burden of dengue in Sri Lanka, innovative methods for vector control and Bionomics, dengue diagnostics, clinical management and policy and behavioral studies.

Despite several efforts by the Sri Lankan government, Dengue incidence in Sri Lanka continues to rise (Figure 1). As highlighted in the National Strategic Plan, innovative and effective tools for vector control and disease surveillance are needed to supplement current efforts.

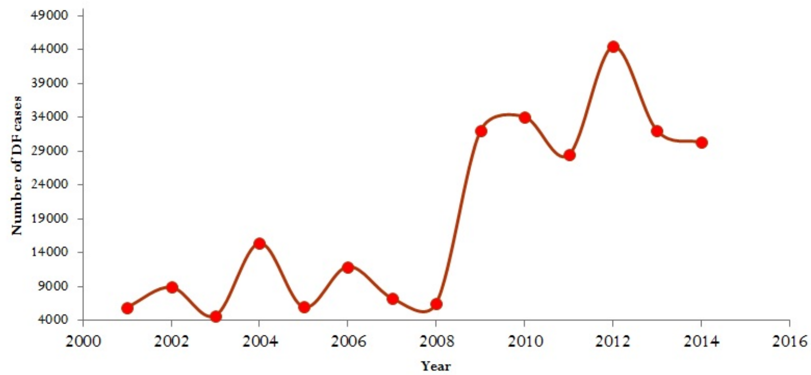


Figure 1: Dengue Incidence in Sri Lanka between 2010 to 2014 (Epidemiology Unit, 2014)

1.5 AIMS AND OBJECTIVES

The overall goal of the study is to explore the possibility of developing a dengue risk prediction model, that integrates data on environmental and socio-economic variables obtained from various sources within a GIS framework. Sources include, satellite imagery and remote sensing data, meteorological and census statistics . The specific aims of the study are:

1. *To conduct a review of previous studies pertaining to GIS and RS applications in dengue research in South Asia (Paper 1)*

   The literature will summarize the evidence supporting the role of GIS and remote sensing in dengue surveillance. Further, the review will be identify the research gaps in South Asia, particularly in Sri Lanka, in light of the increasing incidence of dengue in the region. The review will also demonstrate how the current study will build on the existing evidence base, and addresses some of the research gaps in Sri Lanka.

2. *Using RS and GIS, characterize the local environmental characteristics Colombo municipal council (CMC) area, Sri Lanka (Paper 2)*

   This work will analyze the remotely sensed satellite data generated on LULC classes for the study area. These data will be imported to ArcGIS and SAS, to evaluate their association with dengue incidence in Aim 3. The tasks under Aim 2 are as follows:

   - A high spatial resolution satellite imagery (Quickbird) from April 16th, 2007, will be used to obtain information on LULC classes. Remotely sensed data will then be analyzed using supervised classification (maximum likelihood) and object-based classification. A comparison of the accuracy index of these from method will be done; a priori, we will use the object-based classifier as it is used in most studies and will facilitate cross study comparisons.
   - Spatial maps would be generated to demonstrate the distribution of the environmental variables for the CMC area.

3. *To identify the environmental determinants of dengue in the CMC area, Sri Lanka (three papers)*

   In this part of the study, we evaluate the association between several environmental including temperature and rainfall, land cover types estimated in aim 2, population characteristics and household characteristics and dengue incidence rate in CMC, Sri Lanka. Data on various environmental factors are available at different temporal and spatial resolution (Figure 2). Data on dengue cases are available between 2005 and 2011, for each day for of the 55 GNDs in the CMC area. Data on climate variable is available daily but for the entire CMC. Other local environmental characteristics such as LULC classes obtained from image classification, population and household characteristics are available for each of the 55 GNDs but only at one point in time. Thus, the analysis for aim 3 was divided in two parts. In part I, we will evalaute the relationship between the climate variables and daily counts of dengue cases for the entire CMC area. In part II, we will evaluate the relationship between local environmental factors and dengue incidence rate by GND using non-spatial and spatial analysis. The tasks under aim 3 are:

- **Step 1**. We will assess the relationship between weekly mean temperature and cumulative rainfall, and total weekly counts of dengue cases using generalized linear regression models. We will also evaluate the effect of lagged climate data on weekly counts of dengue data (part I).
- **Step 2.** Using generalized linear regression models, we will evaluate the relationship between local environmental factors and dengue incident rates per GNDs. The environmental characteristics include variables such as wall and roof materials used for households, toilet facility, piped-water supply, housing density; and population density (part II).
- **Step 3**. We will use Getis-Ord-Gi* statistics in ArcGIS, to demonstrate dengue fever risk, spatio-temporally on a monthly basis across seven years and annual basis across 12 months to generate spatio-temporal risk maps (part II).
- **Step 4.** We will evaluate the spatial relationship between the environmental variables and risk of dengue incidence rate, using ordinary least squares (OLS) and GWR (part II).
- **Step 5**. We will develop a risk index based on the factors associated with dengue incidence rate, and develop a risk map.

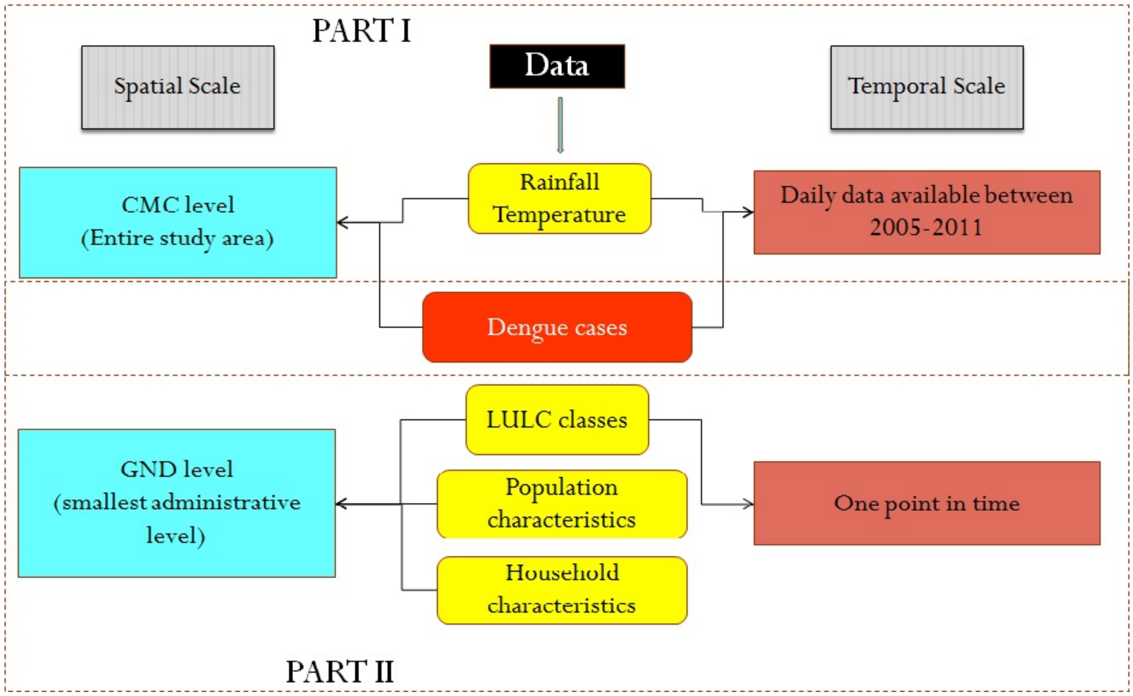


Figure 2: Diagrammatic representation of temporal and spatial scale of data

1.5.1 Conceptual framework for data analysis

To address the three specific aims described above, a conceptual framework was developed. The analysis of data at different spatial and temporal resolutions required a tier-down approach. The following questions, listed below, demonstrate the thought process behind planning the analysis of different resolution data.

1. Which environmental risk factors are associated with dengue incidence in CMC area in Sri Lanka?
2. What are the time periods when the number of dengue cases occur in CMC area? Do temperature and rainfall affect these time periods?
3. Which local environmental and socio-demographic factors are the determinants of the occurrence of higher number of dengue cases in a given location when temperature and rainfall remain constant for the entire area?

The framework is illustrated in Figure 3

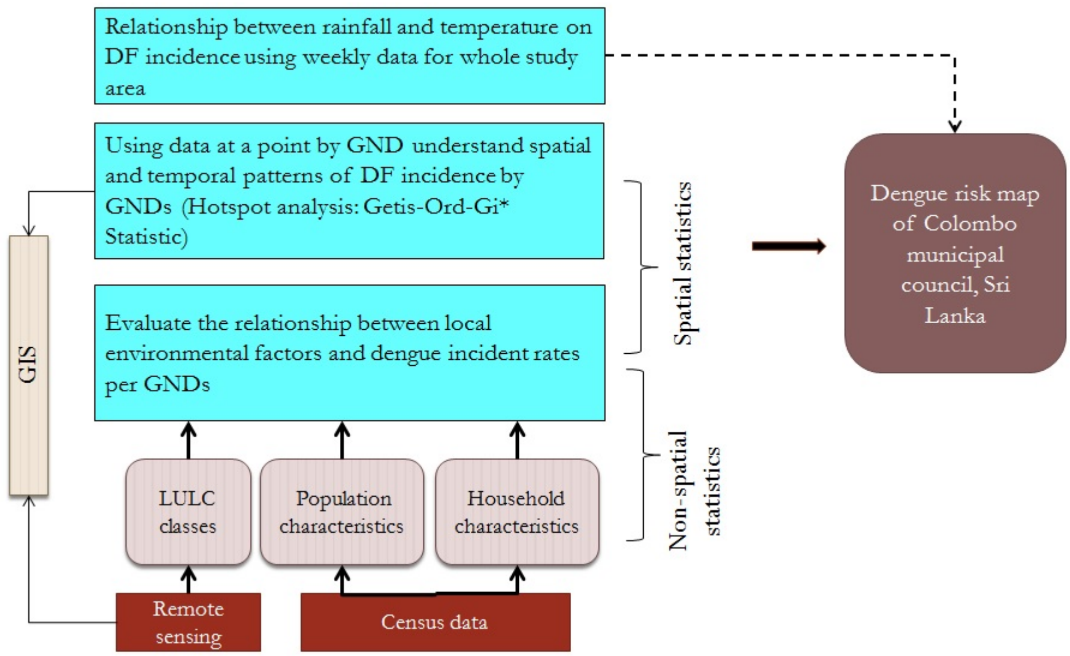


Figure 3: Conceptual Framework for Data Analysis

1.5.2 Ethical approval

The ethical approval for this study was obtained from the Institutional Review Board at University of Alabama at Birmingham and from the University of Kelaniya, Sri Lanka.

2.0 LITERATURE REVIEW

2.1 OVERVIEW

This paper reviews the studies conducted in South Asian countries (India, Sri Lanka, and Bangladesh), that have included application of remote sensing and/or GIS tools to monitor, control, or prevent dengue fever. The goal of this review was to summarize the evidence-base supporting the role of GIS and RS in dengue research in South Asia and to identify research gaps.

2.2 METHODS

We conducted a review of the available literature, pertaining to the use of remote sensing and GIS in the context of dengue research in South Asia. Through a comprehensive literature review, major databases including Pubmed, Blackwell synergy, Google Scholar, Web of Science, CINAHL with full text, and Medline were searched. The key words used in the search included "dengue fever, remote sensing, GIS, dengue risk factors, *Ae. Aegypti,* spatio-temporal models, surveillance and early warning systems." Studies were selected if the focus of the study was dengue fever or *Ae. Aegypti*, and any component of GIS and/or remote sensing analysis. During the initial search, studies were selected based on a review of titles and abstracts. Selected studies were retrieved, and a synopsis was tabulated (see Table 1). All the papers included in the review were evaluated to assess whether (i) remote sensed data was used as data source for risk factors and; (ii) the context of application of GIS techniques such as mapping diseases, and evaluating relationships between the environmental risk factors and dengue incidence rate.

2.3 RESULTS

A total of nine studies were identified from the Indian sub-continent; three were conducted in Sri Lanka, India and Bangladesh each. Remote sensing was used only three studies; two in Sri Lanka and one in Bangladesh, to obtain information on land use/cover classification. GIS was primarily used for mapping the distribution of dengue cases and their risk factors. Spatio-temporal mapping of dengue cases was conducted in two studies. Five of the nine studies used GIS to identify environmental determinants of dengue incidence. Exploratory spatial analyses using cluster algorithms such as inverse distance weighting, hot-spot analysis and Local Indicators of Spatial Association (LISA) measure were used to develop predictive risk maps, identifying areas with high risk of disease occurrence.

2.3.1 Studies in Sri Lanka

Three studies were published from Sri Lanka (Table 1) . Pathirana et al. (2009) conducted a study in an urban area in the western province of Sri Lanka, to evaluate the relationship between climate factors, including daily rainfall and temperature, on incident dengue cases in children below 18 years of age, from 2000 to 2004. Using maps and visual comparison, a positive correlation was found between areas with less rainfall and higher incidence of dengue cases. However, regression analyses found inverse association between weekly rainfall and dengue outbreak in Colombo, which reduced when other factors such as family environment, poverty, lack of proper preventive care facilities, difficulty to diagnose disease in time, unsafe drinking water, lack of proper sanitation etc., were included in the model. A predictive model developed using an interpolation method in GIS, described as inverse distance weighting, found very high risk areas of dengue incidence in north-western part of the western province.

Jayasooriyaa et al. (2009) compared relationship between Breteau indices (BI) of *Ae. Aegypti* and dengue case incidence (Jayasooriyaa et al., 2009). The study was conducted in a urban and semi-urban Kadugannawa Medical Officer of Health area (MOH), in Kandy district in Sri Lanka. The Kandy district is located in north-central part of Sri Lanka and comprises of 95 Gram Niladari Divisions (GNDs) (smallest administrative units in Sri Lanka). Data on incident cases of dengue fever and dengue hemorrhagic fever reported between 2004 and 2007 were collected from each GND. Similarly data on BI of *Ae. Aegypti* and *Ae. Albopictus* collected from a representative sample of 100 households from each GND between 2004 and 2007 were gathered. Comparison between incident dengue cases and BI for *Ae. Aegypti* and *Ae. Albopictus* found that GNDs with high BI for *Ae. Aegypti* positively correlated with high incident dengue cases ($r^2 = 0.55$); while a weak negative correlation was present between high BI for *Ae. Albopictus* and incident dengue cases ($r^2 = -0.04$). Based on these results, a risk map was drawn for the Kadugannawa MOH area. Areas with high BI of *Ae. Aegypti* were designated high risk areas for dengue, GNDs adjacent to the high risk GNDs were considered medium risk, and GNDs with high *Ae. Albopictus* BI were considered as low risk areas. Following the assessment of risk, an educational intervention was implemented in high risk areas in the month July, 2008, considered a peak time for dengue cases. Subsequent monitoring of cases August 2008-December 2008 indicated a decline in the incidence of dengue cases. A recent study conducted by Kannathasan et al. (2013) identified risk factors associated with dengue incidence to create a risk map for dengue. The study was conducted in the Jaffna municipal area, a major cosmopolitan area in the Jaffna district of Sri Lanka. Remotely sensed data from Quickbird satellite imagery was used obtain information on land use/cover classes and create maps. Variables including land cover use, population density, locations of public places like schools, hospitals, economic status, and housing types were mapped in GIS using GND as a base unit. Spatial multi-criteria analysis was used to assign standardized values between 0 and 1 for each level of the variables.

Finally a composite score was assigned to each GND with a value between 0 and 1. Values approaching 1 were considered representing high risk areas and values approaching 0 as areas having the least risk of dengue transmission.

2.3.2 Studies in India

Three studies that fulfilled the inclusion criteria were identified from India (Table 2). A study conducted in Jalor, a rural district of Rajasthan in India, to evaluate association between socio-cultural factors and dengue incidence (Bohra and Andrianasolo, 2001). Investigators selected a random sample of dengue cases' households ($n = 37$) and households with no dengue cases ($n = 40$), and collected information on more than 60 socio-economic and cultural variables. A stepwise regression analysis found that eight variables significantly associated with dengue incidence($r^2 = 0.96$). A combined social risk score was developed using the significant variables, and weights were assigned based on published literature. The risk was assigned to each of the 77 households included in the study. The locations of the households were mapped to GIS and overlayed on the administrative map of Jalor. A risk map with five levels of social risks was created. Results of the risk found that majority of the area had low social risk for dengue incidence (61%).

An investigation following an outbreak of dengue fever was conducted in a rural village in Southern India (Nisha et al., 2005). A house-to-house survey was conducted to identify cases and to collect demographic information. Location of the houses and the streets were mapped in GIS. Information on mosquito breeding sites were collected and mapped in ArcGIS. Spatial analysis of the data demonstrated a centrifugal spread of cases from the most affected street until it involved the entire village. Further, space-time permutation model done using SatScan software showed that cases occurred in clusters and that was unlikely to explain the spatial clustering by chance.

Another study conducted in New Delhi, the capital city of India, identified six risk factors, that significantly associated with dengue incidence (Bhandari et al., 2008). In this study, geocoded incident dengue cases were mapped in GIS ($n = 127$). Of the 127 suspected cases, thirty-seven were confirmed cases of dengue. Detailed information including socio-demographic factors, climatic factors and socio-cultural practices, such as storage of water containers in households, mosquito protection, sanitation, and health care were collected from all the thirty-seven cases. Significant risk factors were assigned risk scores between $1 - 3$ to develop four social risk levels, ranging from very high to low risk. A spatial risk map was developed for New Delhi by inputting data from the 37 cases and using inverse distance weighting tool in ArcGIS. All the 127 cases were mapped using GIS and overlayed on the risk map. Almost 88 of the total cases were reported from high risk areas covering 11.18 km^2, followed 10% in high risk and 2% in medium risk.

2.3.3 Studies in Bangladesh

Three studies reported from Bangladesh were included in this review (Table 3). A household entomologic assessment, as well as an attitude and behavior survey, was conducted in Dhaka during the peak epidemic period for dengue/dengue hemorrhagic fever in 2000 (Ali et al., 2003). GIS was used to map the mosquito breeding sites and households with reported dengue cases. Visual inspection of the density maps for vector population and dengue cases indicated high and low vector density areas. Multivariate regression models for analyzing spatial association between vector populations, hospital locations, and dengue clusters found that the clusters were located closer to the four major hospitals in Dhaka suggesting that proximity to hospitals may be a determinant of diagnosis of cases. Spatial association were noted between dengue clusters and vector populations in the area. The study also reported higher number of dengue cases in household with *Ae. Albopictus* larvae suggesting a change in disease epidemiology, since *Ae. Aegypti* has been the most common vector in South Asia.

Banu et al. (2012) reported a study conducted in urban and rural areas in Bangladesh. The study evaluated dengue cases reported over a period of 10 years from

64 districts. A space-time statistical analysis was applied to detect high risk clusters of DF using SaTScan software at the district level. Poisson regression analysis were conducted to identify space-time clusters after adjustment for the uneven geographical density of district population. Results of the analysis found that the epidemic pattern fluctuated from 2000 to 2009. The temporal trend suggested a decrease in dengue outbreaks after 2002. The cluster analyses showed that the space-time distribution of dengue fever cases were clustered during three periods: 2000–2002, 2003–2005 and 2006–2009. Clusters were most likely to be found in Dhaka district followed by Khulna and Chittagong districts.

Ali et al. (2014) conducted a study in Dhaka, for evaluating factors affecting dengue incidence. Data on dengue cases were obtained between 2005 and 2010 ($n = 3169$) from 11 major health service providers. Data included information on date of admission, location of patient's residence, demographic and clinical data, and date of discharge and outcome (dead/alive). Population data were obtained from the 2001 census; data on mosquitoes were collected from entomological surveys. Frequency of the dengue cases between 2005-2010 was mapped by census tract. Visual assessment of the direction diffusion pattern of the dengue virus over space and time showed that the dengue occurrences follow a diagonal South-South Easterly to North-North Westerly pattern with little change over the years. The temporal trend over time found decrease in dengue incidence over time. Investigator evaluated association between several factors including seasonal rainfall and demographics using epidemiological analysis and found that age 18 to 35 years and being male were significantly associated with dengue incidence. The spatial analysis included employing autocorrelation techniques and cluster pattern identification. Results of the Moran's I global spatial autocorrelation statistics indicated a clustered pattern for dengue cases. A further investigation of the cluster analysis using automated complex Monte Carlo randomization procedure revealed significant cluster patterns.

2.4 DISCUSSION

Dengue has emerged as a significant health problem in Sri Lanka, India, and Bangladesh, and the burden continues to grow each year.The current review identified nine papers that applied remote sensing and/or GIS for ecological modeling of dengue fever. All the studies included in this review were published since 2000, and seven in the last five years. This suggests that the application of GIS and remote sensing in dengue research is limited in these countries and needs to be scaled up. A review of the available studies found that GIS was primarily used for spatial mapping of dengue cases, highlight high risk areas of diseases incidence, and in some instances, to identify the determinants or risk factors of dengue fever. Most studies used GIS for assessing spatial distribution of dengue in the study area. Of nine, five studies evaluated the association between various environmental factors such as temperature, rainfall, seasonality, housing patterns/density, frequency of water storage, frequency of cleaning garbage, presence of breeding sites around and within the households of dengue cases, awareness of mosquito protection, and dengue incidence, to find the determinants of dengue incidence. Four studies used inverse distance weighting, Local Indicators of Spatial Association (LISA) measure, or nearest neighbor techniques in GIS, to generate risk maps for dengue incidence. Remote sensing was used in three studies to obtain information on land cover use.

Overall, the studies exemplify the use of geospatial tools such as remote sensing and GIS in dengue risk mapping. Remote sensing can be used as data source for several risk factors including information on population density, vegetation density, and housing density (Khormi and Kumar, 2012; Vanwambeke et al., 2007; Kannathasan et al., 2013). GIS provides a platform for understanding of spatial relationships between the environmental risk factors affecting the dengue vector abundance and consequently dengue incidence. Results of the studies were useful in implementing targeted interventions in high risk areas as well as for planning future dengue surveillance at a local level. A large body of literature has been published from countries in South -East asia and the Americas where dengue fever is a significant health problem. The quality of papers included in this review are comparable with the previous literature. However, the number of studies are very few indicating limited capacity in terms of manpower and training.

2.5 CONCLUSION

Remote sensing and GIS can be included in the routine vector and disease surveillance for dengue fever. These tools can aid in targeted surveillance and interventions in high risk areas for dengue fever, particularly in resource limited settings. The review underscores the need for additional studies to build the evidence-base in South Asia that may inform policy both at the national and at the local level to incorporate the geo-spatial tools in management of dengue fever. Training and capacity building in the use of these tools would be a fruitful investment of resources in these countries.

Table 1: Studies included in the review from Sri Lanka

Author/ study area/time period	Data Sources	Methods and results			
		Mapping of distribution of dengue	Spatiot-emporal mapping of dengue	Identification of drivers of dengue	Exploratory application to predict outbreak zones
Pathirana et al. (2009) /Urban/ 2000-04	DF cases <18Y *Ae. Aegypti* data; Census data; LULC (Landsat 7ETM); rainfall and temperature	$\uparrow$ DF with $\downarrow$ rainfall	None	Weekly rainfall not significant in adjusted model	IDW used to predict DF risk in NW parts
Jayasooriyaa et al. (2009)/ Kandy district (95 GND)/ 2004-2007	DF/DHF cases Vector data from sampling households in each GND	$\uparrow$ DF with $\uparrow$ rainfall	None	$\uparrow$*Ae. Aegypti*	None
Kannathasan et al. (2013)/ Jaffna District/ 2007-2011	DF cases LULC (Quickbird) Census	DF/DHF	None	LULC Population density House type	Spatial multi-criteria analysis to create a spatial risk map.; validated with DF data

DF = dengue fever; DHF= Dengue hemorrhagic fever; LULC = Land use land cover classes; IDW = Inverse distance weighting

Table 2: Studies included in the review from India

Author/ study area/time period	Data Sources	Methods and results			
		Mapping of distribution of dengue	Spatio-temporal mapping of dengue	Identification of drivers of dengue	Exploratory application to predict outbreak zones
Bohra and Andrianasolo (2001)/Rajasthan/2001	Households (N=77); dengue affected (n=37) and unaffected samples (n=40); SES and cultural factors from interviews; climate, population, vector, LULC	DF cases	None	Multivariable regression; 8 variables related to water storage, preventive measures and housing patterns	Nearest neighbor technique used to develop social risk map
Nisha et al. (2005)/Tamil Nadu, rural/ 2001	DF cases (n=301 in 116 houses) from house-house survey (n= 989 persons in 186 houses) Demographic data (age, gender, occupation etc.) Census data, Entomological data (mosquito indices)	DF and non-DF cases, clustering	Space-time mapping(1mnth) Spread of DF centrifugal, starting from southern part of the village	None	None
Bhandari et al. (2008)/Delhi, Urban/ 2008	Suspected DF cases (n=127)→ confirmed DF cases (n=37). Census data, climate, LULC from Google Earth, socio-cultural practices (n=127) Geocoded solid waste collection points	DF/DHF	None	Identified six factors using regression in GIS Weights (1-3) were assigned to each of the six variable to develop social risk score	Spatial risk map based on the social risk scores associated with confirmed cases (n=37)

DF = dengue fever; DHF= Dengue hemorrhagic fever; LULC = Land use land cover classes; IDW = Inverse distance weighting

Table 3: Studies in included in the review from Bangladesh

Author/ study area/time period	Data Sources	Methods and results			
		Mapping of distribution of dengue	Spatio-temporal mapping of dengue	Identification of drivers of dengue	Exploratory application to predict outbreak zones
Ali et al. (2003)/Dhaka, Urban/	Surveys →entomological data, KAP (n= 9463 houses) Random sample (n=100) each from 90 wards Population data	DF clusters, Krigging to identify clusters of ↑density vectors	None	↑Presence of Ae. Albopictus with ↑ DF cases (RR = 1.45; CI= 1.14-1.79). No relation betn. presence of *Ae Aegypti* and DF Clustering of cases linked to clustering of *A*edes after adjusting for distance to the major hospitals	
Banu et al. (2011)/Urban and rural/2000-2009 (10 years data)	DF from 64 districts Population data from census Digitized maps for districts	↑risk clusters of DF across years by district	Patterns observed for DF between 2000-09; ↓trend after 2002		
Ali et al. (2014)/ Dhaka/Urban 2005–2010	DF cases (n=3169) Demographic data Census data	DF cases	using stdev. ellipses for 10y ↓DF between 2005-09 except '08	Seasonal rainfall, Age group 18-35 years and being male	Hotspot analysis and spatial auto-correlation

DF = dengue fever; DHF= Dengue hemorrhagic fever; LULC = Land use land cover classes

3.0 SUPERVISED AND OBJECT-BASED CLASSIFICATION FOR QUICKBIRD IMAGERY

3.1 INTRODUCTION

Ae. Aegypti is a main vector for the spread of dengue fever world wide (Gubler, 1998). The habitats for *Ae. Aegypti* are closely associated with human settlements and built urbanized environments. One of important tools currently available for characterizing urban environments are obtaining LULC information using remotely sensed data. This information is available at various temporal and spatial resolutions depending on the source of remotely sensed data. With the help of high-resolution imagery such as IKONOS and Quickbird at 1m and 0.62 m resolution, respectively, it possible to map urban environment to an extent (Troyo et al., 2009). These mapped environments, together with GIS could be used in spatio-temporal models to explain spatial patterns of dengue incidence and spread. Several methods exists for analysis of remotely sensed data depending on the spectral, spatial and temporal resolution of the data (Lillesand et al., 2006). The choice of analysis also depends on the objectives of the application of the remotely sensed data. The goal of the current research was to identify urban environments that may be associated with dengue incidence in the study area, using high resolution imagery. We proposed to compare between two methods commonly used for analyzing high resolution data, to select the optimal method for image classification.

The process of analyzing a remotely-sensed image to extract data is known as image classification. Since 1970s, when satellite imagery was first available, scientists have used several methods to classify the images, to obtain the most accurate information LULC on the earth's surface. Traditionally, image classification was conducted using decision rules solely based on the spectral radiance of the pixels in an image data. These classification methods were referred to as 'pixel-based classification approaches. The pixel-based approach use conventional statistical techniques for classifying the pixels including parallelepiped, maximum likelihood, and minimum distance procedures (Lillesand et al., 2006). The most commonly known and used pixel-based approaches are the supervised and unsupervised classification methods. In the last decade or so, the advancement of medium and high-resolution imagery has allowed for inclusion of other image characteristics, such spatial, textural, and contextual along with spectral response, to be used in image classification. This comparatively newer method of classification has been described as 'object-based classification approach.' In this method, the basic procession units are image objects or segments and not pixels. These image objects or segments are formed by dividing the image in smaller segments that have similar spectral, spatial and/or textural characteristics. A fuzzy logic is applied to each segment or image object to assign membership to a class. The membership value usually lies between 1.0 and 0.0, where 1.0 expressed a complete assignment to a class, and 0.0 expressed absolute improbability.

Studies comparing pixel-based and object-based classification approaches have found that the object-based classification produces significantly better results in for high spatial resolution multispectral imagery, such as Quickbird and IKONOS. A major advantage of these high-spatial resolution images is that such data greatly reduce the mixed-pixel problem Lu and Weng (2009), providing a greater potential to extract much more detailed information on land-cover structures than medium or coarse spatial resolution data. However, classification of high spatial resolution imagery present other problems which include (1) spectral confusion between impervious surfaces and other land covers, due to limited spectral resolution (usually only visible and near-infrared (NIR) wavelengths) and high spectral variation within the same land cover due to the very high spatial resolution; and (2) shadows caused by tall objects and the confusion with dark impervious surface and water/wetland Chen et al. (2007); Dare (2005); Zhou et al. (2009). Figure 4 is a Quickbird false color composite consisting of NIR, Red, and Green bands as red, green and blue, illustrating the complexity of urban landscapes, and the potential confusion between impervious surfaces and other land covers, as well as within these surfaces. For example, different building roofs, roads, parking lots, and shadows appear as different colors on the images, making an automatic extraction of impervious surfaces difficult based on spectral signatures. These problems may lower the classification accuracy of pixel-based classification.

Figure 4: Quickbird false color composite consisting of NIR, Red, and Green bands as red, green and blue; Courtesy: DigitalGlobe Inc.

Users of remote sensing have successfully differentiated impervious surfaces such as buildings, roads, parking lots, etc., from other land-cover classes such as bare lands and water areas, using object-based classification methods in their studies. However, in these studies all the impervious surfaces have been grouped under single class (Moran, 2010; Lu and Weng, 2009). Only few studies have attempted to classify these land use classes separately.

Despite growing evidence supporting the preference of object-based approach for classifying high spatial resolution imagery, it is still not clear how the spatial and other characteristics such as texture, size, and shapes can be effectively used. In particular, comparison between both these approaches needs to be further investigated in densely populated urban areas, where there is high spatial and high spectral variability within the same land-cover class. In Figure 4, one can notice that a large number buildings and roads show spectral homogeneity (grey buildings and grey roads), making it difficult to separate them both spatially and texturally in the classification.

The objective of this part of the study (Aim 2) is to determine whether an object-based analysis of remotely sensed imagery will produce a LULC classification, that is statistically more accurate than a pixel-based analysis, when applied to the same imagery in a densely populated urban area. The current study is part of an epidemiological study evaluating relationship between environmental risk factors such as LULC, and local environmental factors and incidence of dengue fever in Colombo, Sri Lanka. The results of this section would determine the final method of classification for obtaining information on LULC classes for the next step in the data analysis in Aim 3.

3.1.1 Study area

The study area was located in Sri Lanka, an island country in the Indian Ocean (Figure 5). Sri Lanka is an island off the southern coast of India , with an estimated population of around 20 million in 2010 (**Census.lk, 2014**). The global positioning coordinates for the area are 6°57'46.85"N″ to $6°53'39.00"N$ and $79°54'6.93"E$ to $79°50'35.31"E$. It includes the city of Colombo covering an area of 37 square kilometers. Colombo is the commercial capital and a major urban center in Sri Lanka. The city has a residential population of approximately 555,031 people (Census.lk, 2014). As a whole, the country of Sri Lanka is divided in 25 districts in nine provinces . The districts are are further divided in Divisional Secretariats (DS) and each DS division is divided in to Gram Niladaris Divisions (GNDs). The CMC area is located in the CMC district and comprises of 55 GNDs from the Colombo and Thimbigriya DS divisions district (Figure 6). For administrative purposes however, the CMC is divided in six administrative districts which are further divided in wards. There are a total of 47 wards in the CMC area. The digitized maps of CMC with GNDs and Wards were overlapped and compared. It was found that most of the GNDs overlapped with the CMC wards except for eight wards (Figure 7). Since population data was freely available for GNDs, and not for the Wards, it was decided to use GNDs as the smallest geographical unit for dividing the CMC area. Figure 7 shows the overlapping maps of GNDs and Wards maps for 55 GNDs and their respective wards. Population demographic data for these wards was not available for the study. To overcome this problem, we linked the CMC wards to the GND map. GNDs are the smallest administrative units in Sri Lanka (Census.lk, 2014) .

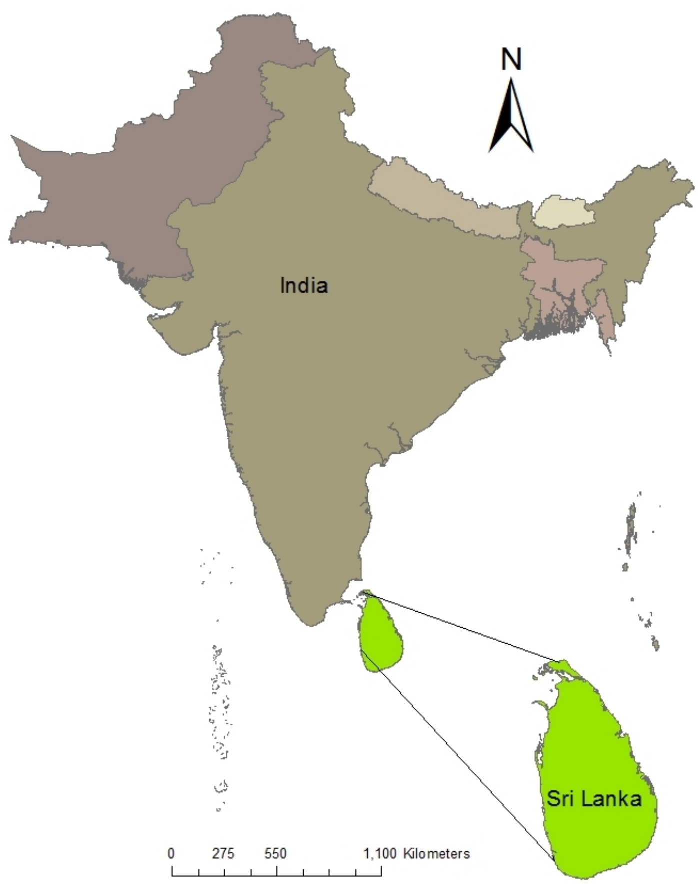


Figure 5: Map of showing South Asia showing Sri Lanka (inset)

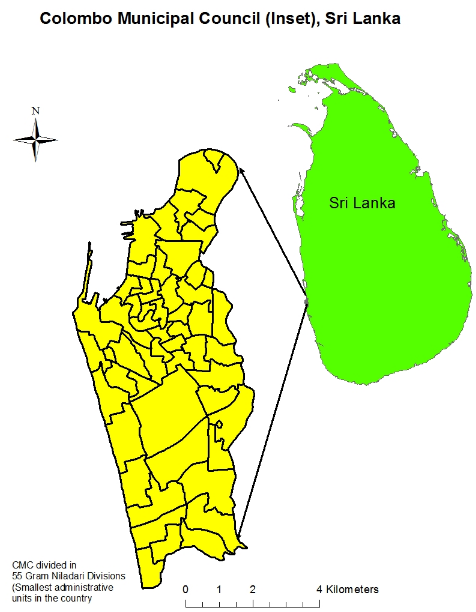


Figure 6: Map of study area

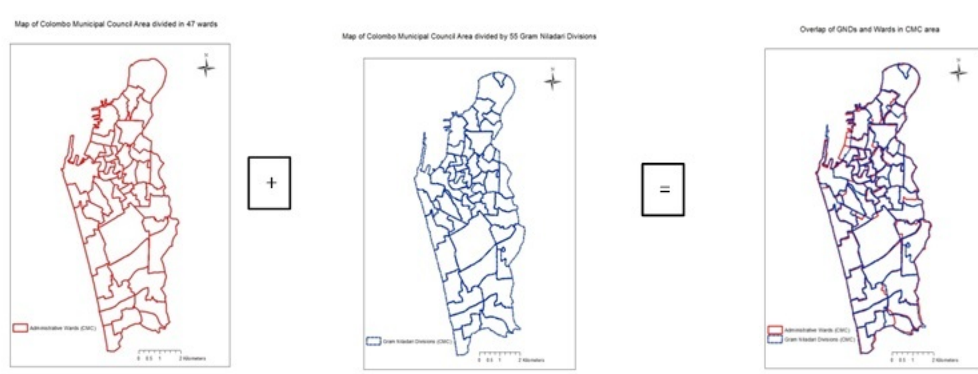


Figure 7: Comparison of GNDs and Wards in CMC area to create one map that linked GND and Wards

3.2 DATA SOURCES

3.2.1 Satellite imagery

The data used in this study is a high spatial resolution QuickBird standard imagery obtained on April 7th, 2007, which includes the study area in entirety. The dataset has 2.4 meter spatial resolution with 4 channels: blue — B1 (0.45–0.52 μm), green — B2 (0.52– 0.60 μm), red — B3 (0.63–0.69 μm), and near infrared — B4 (0.76–0.90 μm). The radiometric resolution of the dataset is 16 bit. The area covered by the satellite image includes urban segments, such as commercial, industrial and residential, and other regions such as, open water, rivers, unmanaged soil, vegetation and trees. This gives a diversity of urban LULC classes. We identified seven LULC classes, including buildings, roads, rivers, trees and shrubs, green open spaces, lakes/ponds and shadows. These particular land-cover classes are important to the ongoing analysis of determining environmental factors associated with risk of dengue incidence in Colombo.

3.2.2 Data management

The methods used in this study include data pre-processing, data processing, image classification (using pixel and object-based feature extraction), and post classification accuracy assessment. Figure 8 provides diagrammatic representation of the methodology implemented in this project.

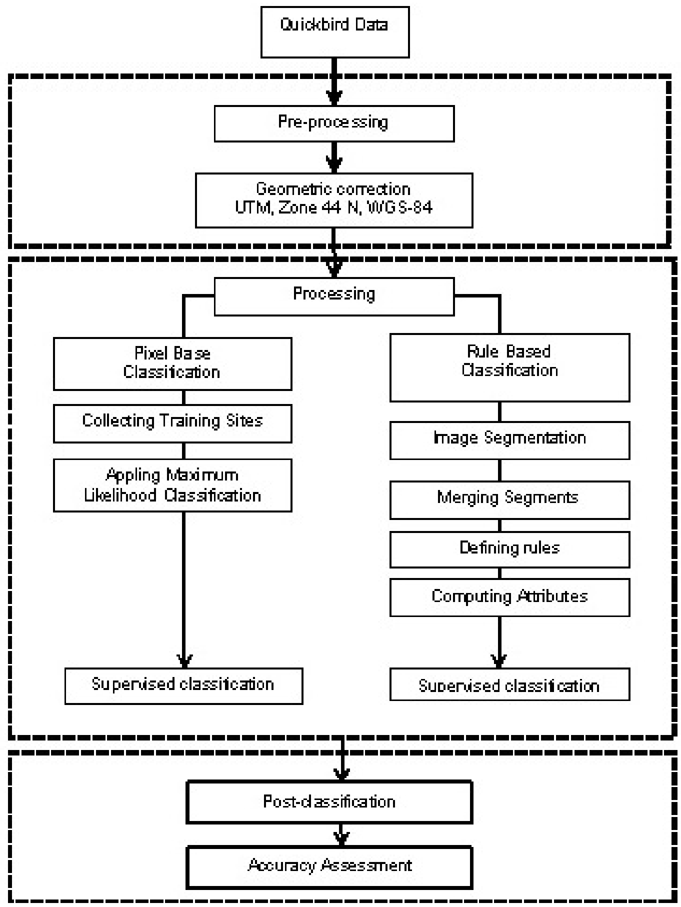
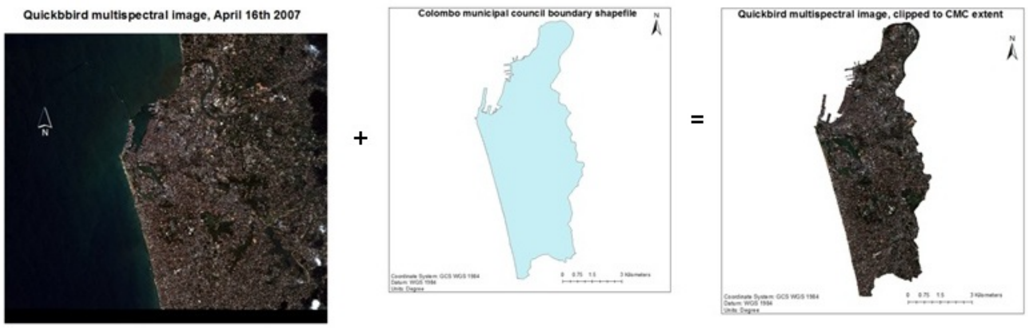


Figure 8: Diagrammatic representation of methods implemented in Aim 2

3.2.2.1 Preprocessing.

1. *Geo-referencing of the images:* The images were imported in ArcGIS. The images had existing coordinate systems of WGS 1984. Adding six control points, the images were aligned with the base imagery from ArcGIS. The georeferenced image was then updated to save the transformation information with the dataset and saved to *.tif* files.
2. *Clipping the raster datasets to the extents of the study area:* After geo-referencing the imagery, the Quickbird multispectral and panchromatic raster images were

clipped, to extract the portion of the raster datasets that included the CMC study area, using the CLIP tool(Figure 9a). The images completely covered the boundaries of the CMC with an approximate area of 37 sq.km. The CMC shapefile used for clipping the raster data had the same geographic coordinate systems as the raster data (WGS 1984) and was projected on the World UTM 1984 44N. It was further aligned with the base imagery in ArcGIS using the transformations to accurately georeference the shapefile.



(a) Clipping of Raster data to specific extents

(b) Pan-sharpening of raster data

Figure 9: Pre-processing of raster data

1. *Pan-sharpening of the multispectral image:* Both the multispectral and panchromatic images were then imported in ENVI 5.1 software. The multispectral image (low spatial-resolution) was pan-sharpened using the panchromatic (high spatial resolution image) using the PC-spectral sharpening technique, see Figure ***9b***. The

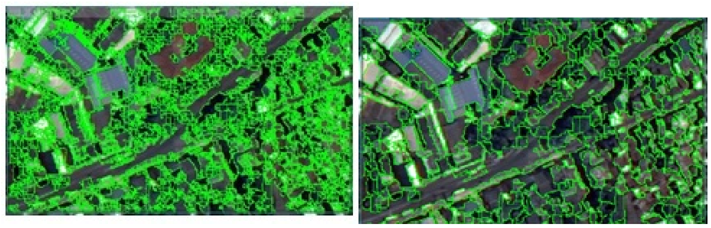
PC-spectral sharpening algorithm assumes that the low spatial resolution spectral bands correspond to the high spatial resolution panchromatic band. If both data sets are geo-referenced, ENVI additionally co-registers them on the fly. ENVI applies PC spectral sharpening by: (1) performing a PC transformation on the multispectral data; (2) replacing PC band 1 with the high spatial resolution band and scaling the high resolution band to match the PC band 1, so no distortion of the spectral information occurs; the PC spectral sharpening method assumes that the first PC band is a good estimate of the panchromatic data. (3) performing an inverse transform; (4) resampling of the multispectral data to the high resolution pixel size using a nearest neighbor, bilinear, or cubic convolution technique. For this study, we used the nearest neighbor technique for re-sampling.

3.2.2.2 Image processing The image was then processed using object-based feature extraction classification and pixel-based maximum-likelihood supervised classification. The classification was conducted in ENVI 5.0 software.

1. *Object-based classification:* Object-based classification extracts information from a high-resolution panchromatic or multispectral imagery, based on spatial, spectral, and texture characteristics of the objects contained in the image. Object-based classification primarily includes two components; first is to 'Find Objects' and second is to 'Extract Features'. The 'Finds Object' task is divided in to four steps: Segmentation, Merge, Refine and Compute Attributes. In this analysis, the pan-sharpened image was loaded in the ENVI FX module along with a shape file of CMC boundary as a mask file.

   - Segmentation: Rule-based classification is started by segmenting the image using an edge-base segmentation algorithm at a scale level of 35. The scale level values range from 0.0 (finest segmentation) to 100.0 (coarsest segmentation, all pixels assigned to one segment). The scale level of 35 was chosen after applying several levels to assess which level segmented the image, such that structures, such as buildings and roads, were delineated. The result of the segmentation is shown in Figure 10a.

- Merging segments: Merging is an optional step used to aggregate small segments within larger, textured areas such as trees, clouds, or fields, where over-segmentation may be a problem. In this analysis, the merge level was chosen at value of 85. The results of merging are shown in Figure 10b .



(a) Image Segmentation (b) Merging of Segments

Figure 10: Image Processing

- Computing attributes: The ENVI feature-extraction module computes spatial, spectral and texture attributes for each segmented object in the image. The image was processed and was readied for feature extraction and classification.
- Classification: Following object identification, rules were defined to extract each class. A total of seven classes were defined including buildings, deep waters/rivers, green spaces, lakes, roads, shadows, and vegetation. Table 2 provides list of rules for all the seven classes. The classification was completed in iterative steps using the rules and the nearest neighbor classifier.

Table 4: Rules for object-based classification

Classes	Rules for object-based classification
River	Band="3" Spectral_Mean between value="12.26362, 150.00000"
Shadows	Band="3" Spectral_Mean between value="50.00000, 170.00000" Band="0" Area" between value="0.75972, 10000.00000" Band="4" Texture_Variance value="0.00020, 0.65943" Band="0" Compactness value="0.10000, 0.28209"
Green Spaces	Band="4" Spectral_Mean > value="0.30000" Band="4" Texture_Variance" < value="0.00200"
Vegetation	Band="4" Spectral_Mean > value="0.30000" Band="4" Texture_Variance between value="0.00200, 0.23751"
Roads	Band="4" Spectral_Mean between value="-0.93563, 0.30000" Band="3" Spectral_Mean between value="170.00000, 1000.00000" Band="6" Major_Length operation" > value="25.00000" Band="0" Compactness < value="0.13000" Band="0" Area between value="350.00000, 50000.00000" Band="0" Rectangular_Fit between value="0.16946, 0.75000"
Building	Band="4" Spectral_Mean between value="-0.93563, 0.30000" Band="0" Area < value="150000.00000" Band="0" Rectangular_Fit > value="0.20000" Band="0" Compactness between value="0.13000, 0.28209" Band="3" Spectral_Mean between value="170.00000, 1000.00000" Band="4" Texture_Variance between value="0.00020, 0.28414"
Lakes	Band="4" Spectral_Mean between value="-0.93563, 0.40000" Band="4" Texture_Variance < value="0.00500" Band="3" Spectral_Mean between value="110.00000, 280.00000" Band="0" Area > value="32000.00000" weight

2. *Pixel based supervised classification:* A pixel-based approach utilizes only the spectral information of the pixels to classify an image in different land cover classes. In this classification, training areas that describe the typical spectral pattern of the land-cover classes are defined. Pixels in the image are compared numerically to the training samples and are labeled to the land cover class that has similar characteristics. The supervised classification was conducted using the

training areas (Table 5) and the maximum likelihood algorithm (MLC). A MLC is a parametric classifier that assumes normal or near normal spectral distribution for each feature of interest. An equal prior probability among the classes is also assumed. This classifier is based on a probability that a pixel belongs to a particular class. It takes the variability of classes into account by using the covariance matrix. Therefore, MLC requires sufficient number of representative training samples for each class to accurately estimate the mean vector and covariance matrix, needed by the classification algorithm. Table 3 lists the total number of training areas selected for each class in the current analysis.

Table 5: Training areas for supervized classification

Class	Number of training areas
Buildings	94
Deep water/rivers	29
Green spaces	21
Lakes	9
Roads	79
Shadows	43
Vegetation	52

3.3 ACCURACY ASSESSMENT

The classified imagery was imported in ArcGIS to conduct accuracy assessment using the confusion error matrix. A confusion error matrix or a classification error matrix compares on a category-by-category basis, the relationship between known reference data (ground truth) and the corresponding results of an automated classification. These error matrices show the contingency of the class to which each pixel truly belongs (columns), on the map unit to which it is allocated, by the selected analysis (rows). From the error matrix, we estimated the overall accuracy, producer's accuracy, user's accuracy, and kappa coefficient, for both supervised and object-based classification methods. We also studied the classification errors of omission (exclusion) and commission (inclusion).

The overall accuracy is computed by dividing the total number of correctly classified pixels (i.e., the sum of elements along the major diagonal in a a error matrix) by the total number of reference pixels. Similarly, the accuracies of the individual categories can be calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in the corresponding row or column. Producer's accuracies refer to the results from dividing the number of correctly classified pixel in each category (on the major diagonal) by the number of reference set pixels used for the category (column total). These numbers indicate how well the reference set pixels of the given land cover class type are classified. The user's accuracies are computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category (row total). This number represents the measured of commission error and indicates the probability that a pixel classified in to a given category actually represent that category on the ground.

The omission error refers to the pixels that should have been classified as a certain class but were omitted that class category. The commission error refers to the pixels that were improperly included in a class category. The Kappa coefficient can be used as a measure of agreement between model predictions and reality (Congalton, 1991). The Kappa coefficient was estimated to assess the percentage to which the percentage correct values of an error matrix are due to true agreement versus chance agreement (Lillesand et al., 2006). Below, we have described the steps taken to complete accuracy assessment.

1. *Creation of reference shapefile for accuracy assessment.* It has been suggested that a minimum of 50 sample points for each land-use land-cover category in the error matrix be collected for the accuracy assessment of any image classification (Congalton and Green, 1998). A reference database consisting of 50 or more

points for each land cover with a total of 555 test points was prepared using the same image that was classified. Value of each reference point was assigned based on the visual interpretation of the object, aided by Google Earth, and from the knowledge of the area. To be consistent, and for precise comparison purposes, we applied the same sample points generated for the output generated by the object based classifier as the output produced by the pixel-based classification technique (maximum likelihood). A new empty shape file was created in ArcGIS. Two new fields were added to the empty attributable table of the shapefile. The first field created was the reference 'landcover' text field. The second field was a numerical field for designating the land cover class by digits. Using the 'create feature' tools in ArcGIS, 50 or more reference data points were added for each class. Using 'select by attribute' feature and field calculator tool, information for each land-cover class in text and in number were added to the attributable table.

2. *Adding values from the classified images for the reference points.* Using the 'Spatial Analyst Tools/Extraction/Extract Values to Points' tool, the class values for the reference test points were extracted from the classified images, both object-based and supervised. Thus, two datasets were created consisting of reference test points, with ground truth values and their corresponding pixel values from the classified raster, one for the object-based classification, and the other for the supervised classification.

3. *Creation of the confusion or error matrix:* The frequency tool in ArcGIS was used to create a summary table cross-tabulating frequency of the every different value of ground truth point with the frequency of different value for the raster points. A pivot table was then used to create the confusion matrix for each type of the classification.

3.4 RESULTS

The results of the classification using both object-based classification and supervised classification are presented in Table 4. The results show the total area classified under each class in square meters, and the percentage of the total area under classification. The object-based approach classified the data into seven classes with buildings making up the largest land cover class (45.7%) followed by vegetation (25.5%), roads (10.4%), shadows (6.0%), green spaces (5.5%), rivers (2.3%) and lakes (1.6%) (Table 6). About 2.8% of the total area remained unclassified. In supervised classification, the image pixels were classified in seven classes constituting of buildings (31.7%), while the remaining classes constituted of vegetation (23.5%), roads (22.7%), green spaces (9.9%), shadows (9.5%), rivers (1.3%) and lakes (1.3%). Figure11 shows the results of the two classification methods. Results of the accuracy assessment are presented in Tables 5 and 6.

Table 6: Total area and percentage of area covered by each land-cover class per classification method

	Object-based		Supervised		
Land-cover class	Area in m^2	%	Area in m^2	%	P-value
Rivers	941233.2	2.3	521098.4	1.3	0.03
Lakes	642067.05	1.6	537908.7	1.3	<0.001
Green spaces	2216814.4	5.5	4004595.4	9.9	<0.001
Vegetation	10273396	25.5	9460364.4	23.5	0.26
Buildings	18389624	45.7	12759628.6	31.7	<0.001
Roads	4198725.7	10.4	9135826.7	22.7	<0.001
Shadows	2426625.3	6.0	3842642.	9.5	<0.002
Unclassified	1127288.5	2.8	0	0	
Total area	40215774.15	100	40262064.6	100	

3.4.1 Classification accuracies (pixel-based supervised classification)

From Table 7, it can be observed that the supervised classification produced an overall accuracy of 79.3%, and with a kappa coefficient of 0.7399. The lowest producer's accuracy (65.7% was given by the building class indicating that only 66% of the building area was correctly identified, but almost 95% of the areas identified as buildings were actually buildings. It was found that, almost 33% of the buildings were classified as roads. The shadows class produced the second lowest producer's accuracy (74%. One of the highest producer's accuracy was for roads (96, but the user's accuracy for roads was only 45%, suggesting that 96% of the roads were correctly identified, but only 45% of the identified areas were actually roads. Other classes, including rivers (100%), green spaces (86%, , vegetation (88% and lakes (100%, 100%), had high level of accuracy for both producer's and user's respectively.

3.4.2 Classification accuracies (object-based classification)

In contrast to the traditional method, the object-based classification approach generated an overall accuracy of 79.8% and a Kappa coefficient of 0.734. The lowest producer's accuracy was for and vegetation (67.6%), but the user's accuracy for vegetation was almost 92% (**Table 6**). Almost 28% of the area identified as vegetation was classified as green spaces. The roads class had second lowest producer's accuracy (68.7%, as well as the lowest user's accuracy. This accuracy was low because of signature confusion between building and roads.

Comparison of class proportions between individual class types.

We compared between the proportion of each across the two classification approaches. Results of the comparison found that except for vegetation, the proportions of all other six classes were significantly different from each other (Table 6).

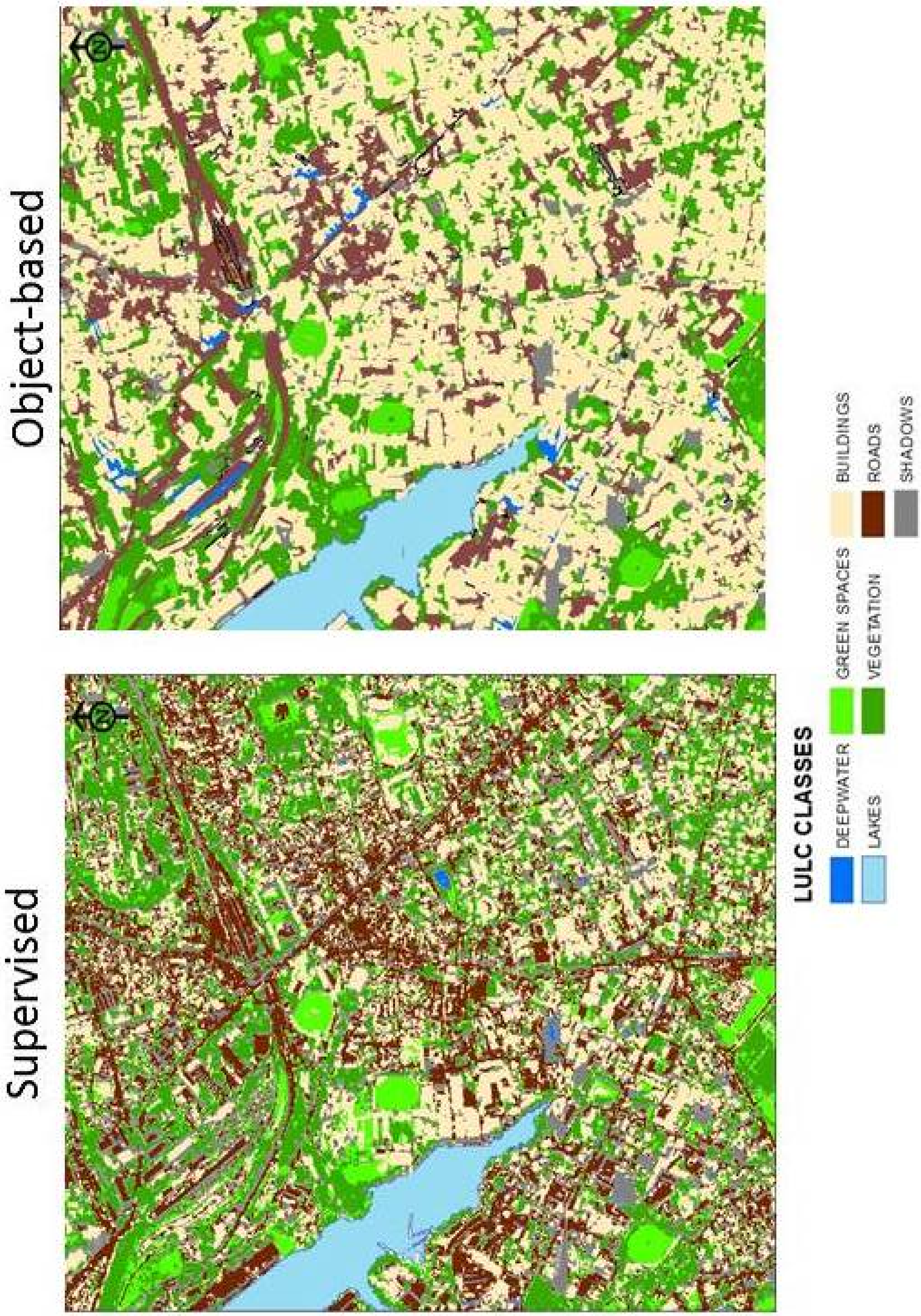


Figure 11: Classified imagery

Table 7: Confusion Error matrix for supervised classification

Classified	Reference								User's	Producer's
	Rivers	Shadows	Green spaces	Vegetation	Roads	Buildings	Lakes	Total	accuracy	Accuracy
Rivers	**31**	7	0	0	0	0	0	38	81.6	100.0
Shadows	11	**50**	0	2	0	0	0	63	79.4	74.0
Green spaces	0	0	**51**	1	0	3	0	55	93.0	86.0
Vegetation	0	0	4	**65**	0	1	0	70	93.0	88.0
Roads	0	0	0	0	**63**	78	0	141	45.0	96.0
Buildings	0	1	3	0	4	**157**	0	165	95.0	66.0
Lakes	0	0	0	0	0	0	**23**	23	100.0	100.0
Total	42	58	58	68	67	239	23	**555**		

Overall Accuracy = 79%; Overall Kappa Statistics = 0.7399

Table 8: Confusion Error matrix for object-based classification

Classified	Reference									User's	Producer's
	Unclassified	Rivers	Shadows	Green spaces	Vegetation	Roads	Buildings	Lakes	Total	accuracy	Accuracy
Unclassified	0	2	0	0	0	3	24	0	29		
Rivers	0	**37**	0	0	0	0	0	0	37	100	88.1
Shadows	0	3	**51**	0	0	0	0	0	54	94.4	87.9
Green spaces	0	0	0	**55**	19	0	2	0	76	72.4	94.8
Vegetation	0	0	1	0	**46**	0	0	3	50	92.0	67.6
Roads	0	0	2	0	0	**46**	25	0	73	63.0	68.7
Buildings	0	0	3	3	3	18	**188**	0	215	87.4	78.7
Lakes	0	0	1	0	0	0	0	**20**	21	95.2	87.0
Total	0	42	58	58	68	67	239	23	**555**		

Overall Accuracy = 79%; Overall Kappa Statistics = 0.7397

3.5 DISCUSSION

In this section, we compared object-oriented with pixel-based classification (supervised) approach using Quickbird image. The result shows that both object-oriented classification and supervised classification produced satisfying results. The overall accuracy of both classifications was approximately 79%. These results are comparable with previous studies in terms of overall accuracy. However, previous studies have found that compared to a pixel, an image objects (segments) can offer important information, that is necessary to interpret an image. Compariosn between the LULC classes from each method of classification found that the proportion of LULC classes differed significantly between both groups except for vegetation. Since previous studies have consistently found that object-based classification is a preferred method for high-resolution imagery such as Quickbird, we decided to use the results of the object-based classification on LULC classes for the next steps of the analysis.

To our knowledge, this is the first study that has used high-resolution imagery to classify land cover in the CMC area. The analysis classified the area in seven distinct LULC classes including buildings, roads, vegetation, open green spaces, rivers, lakes and shadowed areas from buildings and trees. The results of the classification show that CMC area is a densely populated urban area with both residential and commercial areas covering a total area of 37 sq km. One of the limitations of our analysis was the inability to differentiate between commercial and residential properties. Another limitation was the miss-classification between buildings and roads for both supervised and object-based classification. Previous studies have found that extraction of road surfaces in densely populated urban areas is difficult (Repaka et al., 2004). These limitations can be addressed in future with use of next-generation satellite data such as Worldview-I and II which have higher spectral, temporal and spatial resolution. It is alo possible to improve classification results with better geographical knowledge of the study area. Additional data would also allow for assessment of change in LULC over time with satellite data from different time periods within an year and between years.

In sum, the study found comparable results of classification using supervised (pixel) based and object-based. classification. The results of the study show potential to use remotely sensed data and its analysis as a useful tool to obtain information on LULC in places where such data is not readily available. But it also highlights limitations with data analysis with available methods and the need for better quality and frequency of remotely sensed data for densely populated urban areas.

4.0 ENVIRONMENTAL DETERMINANTS OF DENGUE IN CMC AREA

4.1 INTRODUCTION

As discussed in the background, several environmental and social factors including climate, local environmental and, socio-demographic factors have been significantly associated with dengue incidence. In the previous chapter, we discussed the methods used to extract data on the LULC for the CMC area. The information collected from the image classification on LULC served as the data source for estimating some of the local (at the GND level) environmental risk factors of dengue incidence. Information on other risk factors including climate and demographic variables was also obtained. In this chapter, we have described in detail research methods described under Aim 3 including information on all the data sources for independent variables and dependent variables; data management and steps in data analysis.

4.2 STUDY DESIGN

The study is a retrospective panel study using data from 2005 to 2011.

4.3 STUDY POPULATION

The study was conducted in the city of Colombo, the financial capital of Sri Lanka, an island country in the Indian Ocean. The country covers an area of 64,740 square km with the total population of about 21, 481,334. Most people in Sri Lanka are of Sinhalese (74%) ethnicity followed by Sri Lankan Moors (7.2%), Indian Tamil and Sri Lankan Tamil who comprise of 4.6% and 3.9% respectively. The majority of people are Buddhist (70%) followed by Muslims (8%), Hindus (7%) and Christians (6%). The country has a high literacy rate of 91.2%.

4.4 STUDY AREA

Refer to section 3.1.1

4.5 DATA SOURCES

The data used in this study include confirmed dengue cases, high spatial resolution satellite data, daily temperature and rainfall collected from 2005 to 2012, and socio-demographic factors from population census statistics in 2010.

4.5.1 Dengue cases

Dengue is a notifiable disease in Sri Lanka. The surveillance case definition for dengue is Sri Lanka is as follows:

- In Children: An acute febrile illness of 2-7 days duration with 2 or more of the following: headache, retro-orbital pain, myalgia, arthralgia, flushed extremities, tender hepatomegaly, rash, leucopenia and hemorrhagic manifestations.

- In adults: An acute febrile illness of 2-7 days duration with 2 or more of the following; headache, retro-orbital pain, myalgia with one of the following: leucopenia, thrombocytopenia or hemorrhagic manifestations.

All medical practitioners in Sri Lanka who attend to patients with a suspected diagnosis of dengue are expected to report the cases to the proper authorities along information on patient's contact information. On receipt of the notification, a Public Health Inspector visits the home of the patient to collect additional data using a standard surveillance form. These data include detailed information on patient's personal history, results of the clinical and laboratory investigations. In the study area, dengue related preventive activities are controlled by the department of Public Health of the CMC. In addition to the national surveillance form, the CMC Department of Public Health collected additional data on environmental variables. The surveillance form is attached in *appendix 1*. Data on suspected and confirmed dengue cases from January 2005 to June 2012 were obtained from the Department of Public Health for this study.

4.5.2 GIS reference shape files

We obtained road shapefiles of the CMC area from the Department of Public Health. We downloaded an image map of CMC area showing administrative districts (and wards) from the CMC website. We digitized and georeferenced the map using ArcGIS creating a 'ward' shapefile and projected the shapefile to WGS 1984 UTM Zone 44 N. The ward feature was linked to the road shape file in ArcGIS. The road shapefile combined with information on ward, street name, type, zipcodes and GNDs was used to create the address locator file. An address locator is a reference file used in ArcGIS to geocoded addresses.

4.5.3 Temperature and Precipitation

Meteorological data including daily temperature and precipitation for the CMC area were obtained from CMC Area Monitoring stations at Colombo Fort and Meteorological Department (1.5 km apart) for 2005-2011. Only one stream of data was available for the entire area. Since the total area covered under the study is about 37 km^2, and low-lying topography, we expected no variability across the GNDs within CMC area. Dengue surveillance in Sri Lanka reports two dengue peaks in a year, depending on the monsoon seasons. The first peaks occurs June-July following the first inter-monsoon season in March and April. The second peak occurs between October - December following the second inter-monsoon season in October and November. The weekly averages of minimum and maximum temperatures were computed for all years. The climate of Sri Lanka is affected by the topographical features of the country and the Southwest and Northeast monsoons regional scale wind regimes. The climate experienced during 12 months period in Sri Lanka can be characterized in to four climate seasons as follows:

1. First Inter-monsoon Season - March - April
2. Southwest monsoon season - May - September
3. Second Inter-monsoon season - October - November
4. Northeast Monsoon season - December - February

We also obtained data on daily rainfall from the Tropical Rainfall Measuring Mission (TRMM) satellite and daily temperature from Land Data Assimilation System (LDAS), to compare with data from the local monitoring stations.

4.5.4 Socio-demographic data

We obtained data on population characteristics from the Department of Census and Statistics in Sri Lanka. The data on population and housing characteristics following the 2010 census surveys is freely available by GNDS on the Sri Lankan Department for Census and Statistics. The data included information on population demographics (total population, population of males, females, and children less than 15 years, between 15 and 60 years and greater than equal to 16 years). The housing characteristics included information on type of roofs and walls, and the sources of drinking water to the occupied houses in the GNDs. The online data was abstracted in Microsoft excel. It was formatted and checked for any missing data or errors. The data for source for drinking water had nine sources listed. We re-grouped the nine categories to three broad categories included- well water, tap water and other that included supply from the river, tankers, rural project and bottled water. A final dataset for population and housing characteristics was exported in ArcGIS for the 55 GNDs that comprise the study area.

4.6 DATA MANAGEMENT

The software used in this study included ENVI 4.8 and 5.0 version and ENVI EX module for analysis of satellite data, ArcGIS 10.1 and SAS 9.3.

4.6.1 Geocoding of dengue cases

Addresses for all dengue cases reported between years 2005 and mid-year of 2012 were reviewed and quality assurance/quality control (QA/QC) was done to check for spelling errors, missing information in Microsoft excel 2010. The final dataset was imported in ArcGIS for further management. Using 'Geocode address' tool in ArcGIS, addresses of dengue cases were geocoded. However, there were difficulties in matching and initial attempts matched 2 to 5 % of the addresses only. Further geocoding was done manually with help of the address locator and Google maps. The geocoding was done at the street level. Since the approximation of the case to the nearest possible location of their homes was critical for this study, the nearest distance from the street was chosen to locate the case to avoid exact location of the case.

4.6.2 Creation of shapefiles and SAS files of LULC classes by GNDs from the classified imagery

ENVI generates the results of the classification in .jpg formats as well as in shapefile format. The results were imported in the ArcGIS. Using the overlay function in ArcGIS, we overlapped the land cover classes with GND shapefile to create a new shapefile that split the land cover classes by each GND. The overlay function splits features in the input layer (Land cover shapefile) where they are overlapped by features in the overlay layer (GND shapefile). The new areas are created where polygons intersect. If the input layer contains polygons, the polygons are split where overlay polygons cross them. The attributes of features in the overlay layer are assigned to the appropriate new features in the output layer, along with the original attributes from the input layer. A final shapefile was created that contained information on GNDs, population data, household characteristics, and land cover classes by GNDs.

4.6.3 Combining population, household characteristics and land cover classes with counts of dengue cases by GND

Using spatial join in ArcGIS, the total number of geocoded dengue cases were linked to the GND that contained the location of these cases. The total counts of dengue cases per GND were also linked to the other environmental variables include population data, household characteristics and land cover classes within the GND that contained the dengue cases. The attribute table of this shapefile containing counts of dengue cases per GND, and other variables were then exported in Dbase files which were then converted to SAS files for further analysis.

4.6.4 Other data

All the other data files included weather data and data on daily dengue cases were checked for errors and consistency and also exported to SAS for further analysis.

4.7 DESCRIPTION OF VARIABLES INCLUDED IN THE ANALYSIS

4.7.1 Dependent Variable

The reported daily dengue fever cases between 2005 and 2011 were geocoded. Depending on the temporal scale of the independent variables, We estimated the dependent variable. We first assessed the total daily, weekly and annual counts of dengue cases per GND. The dependent variable was 'total weekly counts of dengue cases' from 2005 to 2011 for the time series analysis. The dependent variable for assessing the spatio-temporal monthly risk for each GND was monthly incident rate of dengue cases per GND using the formula below:

$$IR_i = \frac{Number\,of\,total\,dengue\,cases\,per\,month}{Total\,population\,of\,GND_i} \times 100,000, \tag{4.1}$$

where, the IR_i is the incidence rate for the i^{th} GND, and where i ranges from 1 to 55.

The dependent variable for assessing overall association between socio- demographic and non-climate local environmental factors was the overall incidence rate of dengue cases per GND estimated using the formula below:

$$IR_i = \frac{Number\,of\,total\,dengue\,cases\,per\,GND}{Total\,population\,of\,GND_i} \times 100,000, \tag{4.2}$$

where, the IR_i is the incidence rate for the i^{th} GND, and where i ranges from 1 to 55.

4.7.2 Independent variables

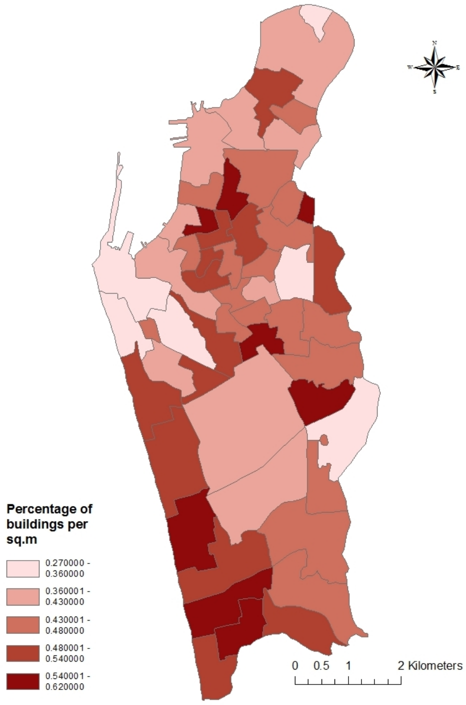
4.7.2.1 Proportion of area covered by buildings in each GND. The data obtained on LULC classess for buildings was further manipulated in ArcGIS 10.1, to generate proportion of built up area in square meters for each GND. Figure 12a provides the distribution of area covered by buildings for each GND. The total proportion ranged 27% to 62%

4.7.2.2 Proportion of area covered by vegetation in each GND.

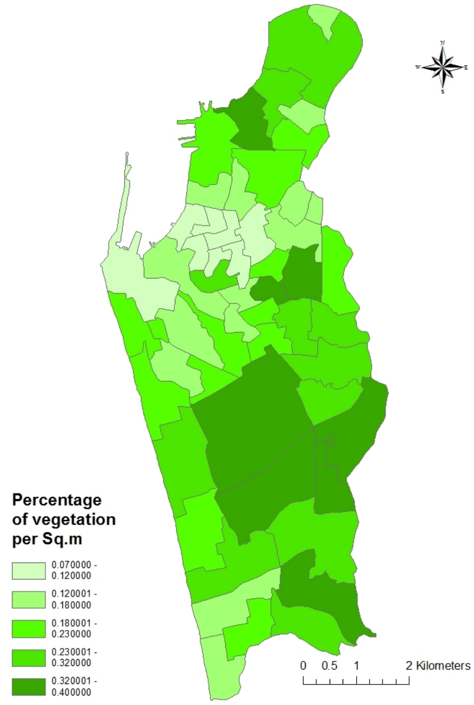
The data obtained on LULC classes for vegetation was further manipulated in ArcGIS 10.1, to generate proportion of area covered by trees and shrubs in square meters in each GND. Figure 12b provides the distribution of area covered by vegetation for each GND. The total proportion ranged from 7% to 40%.

4.7.2.3 Proportion of area covered by roads in each GND.

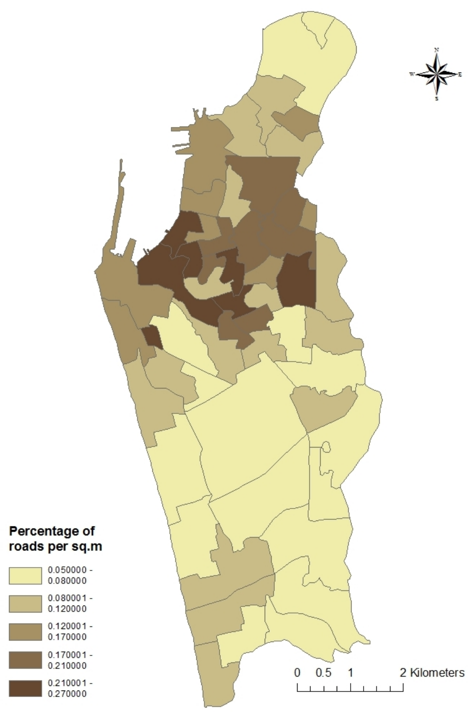
The data obtained on LULC classes for roads was further manipulated in ArcGIS 10.1, to generate proportion of area covered by roads in square meters, in each GND. Figure 12c provides the distribution of area covered by vegetation for each GND. The total proportions ranged from 5% to 27%.



(a) Buildings



(b) Vegetation



(c) Roads

Figure 12: Proportion of land cover classes by GND

4.7.2.4 Population density. The population density of each district was calculated using the total population and the total area for each GND.

4.7.2.5 Housing characteristics. We were able to retrieve data on some of the housing demographic characteristics of the population residing in the CMC area. This data included information on the type of materials used for construction of roof and walls, toilet facility in the homes and types of drinking water supply. This data was retrieved from the Sri Lankan census statistics website (Census.lk, 2014). Description of each of these variables is as follows:

1. Wall materials. The principal materials used in construction of walls in occupied housing units in the CMC district included brick, cement blocks or stone, cabook, pressed soil bricks, kadjan/palymrah, plank or metal sheet and other. The data was available as 'number of houses with walls of particular material for a GND division.' We calculated the proportions of houses for the types of materials. Based on the personal knowledge of the area and the International Wealth Index (IWI), we re-categorized the materials into three categories as brick walls, cement walls and grouped all the other materials in a third group.
2. Roof materials: The principal materials used in the construction of roofs in occupied housing units included tiles, asbestos, concrete, zinc or aluminum sheets, metal sheets, Cadjan or Palymarah or straw and other. We used a similar approach for categorizing roof materials in to three groups. We grouped the houses with having tiled roofs, asbestos roofs and all the other materials as the third group.
3. Toilet facility: The toilet facility for households were divided into seven categories including households having toilets within the unit and exclusively used by the household, household having the toilet within the unit but shared with another household, households having toilet outside the unit but used exclusively by the

household, toilet outside the unit and shared by other households, household having no toilet but sharing the toilets of other household, household using common or public toilets and household not using toilets at all. Households with toilets exclusive for the households and located within the unit were categorized in one group and all other houses with shared toilet facilities were grouped in the second group.

4. Source of drinking water supply: We categorized households into two group: (1) those with piped-water supply within the housing units and, (2) those without piped-water supply directly inside the housing units. The source of drinking water in households without the piped-water included households with tap water outside the housing unit but within the premised of the homes, wells within the household premises or outside, rural water supply projects, tube wells, bowser or river water.

4.7.2.6 Housing density. Data was available for the number of occupied housing units per GND. We estimated the housing density per GND using the following formula:

$$\frac{Number\,of\,occupied\,housing\,units\,per\,GND}{Total\,area\,of\,the\,GND\,in\,m^2} \tag{4.3}$$

4.7.2.7 Weekly average temperature. We computed weekly temperature average in ř$Celcius$ from data on daily average temperature for the years 2005 to 2011 . Weekly averages were computed for both data sources; CMC meteorological department and from LDAS. This data was available at the CMC level.

4.7.2.8 Weekly cumulative precipitation. We computed weekly average of daily cumulative precipitation in millimeters (mm) from data, for the period of 2005 to 2011. This data was available at the entire CMC level.

4.8 DATA ANALYSIS

Analysis began with descriptive statistics using graphs and and summary statistics for all the variables. The continuous variables were summarized using appropriate measures of central tendency (means, medians) and spread (variance, range). The categorical variables were summarized using the proportions. We explored the relationship between all the independent variables and the dependent variable using correlation for continuous variables. We used Pearson's correlation coefficient in case of normally distributed variables and Spearman's correlation coefficient in case of non-normal data. We also assessed for multi-collinearity between the independent variables. If the correlation coefficient $r^2$was 0.75 or more, we then included the correlated variables separately in the model to in the final analysis. We then fitted different models for different set of variables depending on the temporal and spatial scale of the independent and the dependent variables.

4.8.1 Part I: Relationship between climate variables and dengue cases

Weekly counts of dengue cases were computed. Similarly, weekly average temperature and precipitation were computed. Exploratory data analysis for these these variables were performed using box-plots, histograms, and scatter-graphs and autocorrelation graphics. We computed lags from 1 week to up to 25 weeks, for weekly average temperature and weekly average rainfall. We then estimated the correlations between lagged temperature and rainfall, and weekly counts of dengue cases. We analyzed the relationship between weekly temperature and rainfall at various lag periods and dengue incidence independently, using a Poisson multivariate regression model adjusting for seasonality and trend. The four seasons described in section 4.5.3 were included in the model as dummy variables. The year of incident dengue cases was also included in the model as dummy variable. In the final model, both temperature and rainfall were then included in the same model, adjusting for seasonality and trend. The analysis was done using SAS/STAT® software, version 9.3.

4.8.1.1 Sensitivity analysis.

As part of the sensitivity analysis, all the analyses were run using the secondary data sets from LDAS for temperature and TRMM for rainfall. We also subset the data by monsoon seasons to evaluate the relationships between the climate variables and dengue counts for all the datasets.

4.8.2 Part II: Relationship between local environmental variables and incident dengue cases

4.8.2.1 Non-spatial analysis.

We evaluated correlation between all the environmental factors and the incident rates for dengue cases for each GND. Comparisons were made between each subgroup of the risk factors (for eg., roof materials for households included tiles and cement, asbestos,other) and the incident dengue rates. Pearson's correlation coefficient were used for normally distributed data and spearman's correlation coefficient for non-normal data. We further categorized the GNDs into two groups using the median dengue incident rate as the cutpoint. All the variables were compared across the two groups using t-test for continuous and chi-square or Fisher's test for categorical variables.

The variables which were correlated significantly with dengue cases at $p = 0.05$ or had a positive or negative r^2 more than 50% were included in the final regression model. For the final model, we used a Poisson regression and GENMOD procedure. The Akaike Information criterion (AICc) was used to assess the model fit.

4.8.2.2 Spatial analysis

Mapping and modeling the dengue fever risk spatio-temporally by GNDs. Dengue data has both a spatial and a temporal component; i.e., the dengue cases may be clustered in space and in time. We used the space-time hot-spot analysis in ArcGIS 10.1 to assess the spatio-temporal risk of dengue incidence rate within each district. The incidence rate of dengue cases were analyzed spatially and temporally using Getis-Ord GI* statistic to model the monthly risk levels in each district. This approach looks at each feature within the context of neighboring features within a fixed specified distance and fixed specified time interval.

If a feature value is high, and the values for all of its neighboring features are also high, within a specified time window, then the conclusion is that its part of a hotspot. We first used the space-time window option in ArcGIS to conceptualize the spatial relationship of dengue incidence rate in GNDs.

We generated a spacial weights matrix to quantify the spatial and temporal relationships among the GNDs. A threshold distance of 1350 meters and a time interval of one month was specified. Following the conceptualization, we used Euclidean distance to give an output of a z-scores and p-values for each GND in CMC area. GNDs with high z-scores and small p-values indicated spatial clustering of high level of dengue incidence rate hotspots (a high temporal risk in a given period). GNDs with low-scores and small p-values indicated a spatial clustering of low-level of dengue incidence rate hotspots i.e., cold spots for dengue incidence rate (low temporal risk in a given period). The space-time modeling generated one hotspot model for each month of the 7-year period to identify the areas of very low, low, medium, and high incidence probability in recorded dengue cases. The risk areas were classified based on the z scores values: z scores ≥ 3 indicating high risk areas; z-score $2-3$ indicating medium risk, z-scores $1-2$ indicating low risk and z scores ≤ 1 indicating very low risk areas.

Spatial relationship between local environmental factors and socioeconomic factors. The data for the local environmental factors including proportion area covered by buildings, roads and vegetation, green spaces, neighborhood quality, types of roof and wall materials, toilet facility and piped-water supply were available at the GND level for one period in time. Thus, the dependent variable in this analysis was the incidence rate per GND across all seven years. The incidence rate was estimated as described in equation 4.2. We began the analysis by correlating all the independent and dependent variables. The correlation statistics provided information on the type of relationship between the predictor variables and the outcome variable, whether it is positive or negative. Variables that were significantly correlated with the outcome and those that have been found in the literature as potential risk factors of dengue incidence rate were then included in the next step of the spatial analysis.

Spatial data, such as in the current study do not fit traditional, non-spatial regression requirement because (1) they are spatially autocorrelated (i.e., features near each other are more similar than those further away), and (2) the data is non-stationery (ie., feature behave differently based on their location variation). The spatial relationship between the local environmental and socioeconomic factors, and the incident dengue cases was then evaluated in ArcGIS. There are a number of methods that can be used to determine spatial relationships. Two of the common spatial methods are Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR). OLS is a global regression method. GWR is a local, spatial, regression method that allows the relationships you are modeling to vary across the study area. This means that the relationship between the predictor variables and the outcome variables may vary across different GNDs and thus, it allows us to identify these relationships specific to the GND.

- Ordinary Least Squares Regression.

OLS provides a global model of the variable or process you are trying to understand or predict (incident dengue cases); it creates a single regression equation to represent that process. A OLS regression can be denoted as follows:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...\beta_n X_n + \epsilon \tag{4.4}$$

- Geographically Weighted Regression (GWR).

GWR is a local spatial statistical technique used to analyze spatial non-stationarity, defined as when the measurement of relationships among variables differs from location to location (Fotheringham et al., 2002). Unlike conventional regression, which produces a single regression equation to summarize global relationships among the explanatory and dependent variables, GWR generates spatial data that express the spatial variation in the relationships among variables. The conventional regression equation can be expressed as:

$$\hat{y}_i = \beta_0 + \sum_k \beta_k x_i + \varepsilon_i \tag{4.5}$$

where $\hat{y}_i$ is the estimated value of the dependent variable for observation i, β_0 is the intercept, β_k is the parameter estimated for variable k, x_{ik} is the value of the k^{th} variable for i,and $\varepsilon_i$is the error term. Instead of calibrating a single regression equation, GWR generates a separate regression equation for each observation. Each equation is calibrated using a different weighting of the observations contained in the data set. Each GWR equation may be expressed as

$$\hat{y}_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{4.6}$$

where (u_i, v_i) captures the coordinate location of i (Fotheringham et al., 2002). The assumption is that observations nearby one another have a greater influence on one another's parameter estimates than observations farther apart. The weight assigned to each observation is based on a distance decay function centered on observation i.In the case of areal data, the distance between the observation is calculated as the distance between polygon centroid. The distance decay function, which may take a variety of forms, is modified by a bandwidth setting at which distance the weight rapidly approaches zero. The bandwidth may be chosen by minimizing the AICc score, give as

$$AIC_c = 2nlog + e(\hat{\sigma}) + nlog_e(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} \tag{4.7}$$

where $tr(S)$is the trace of the hat matrix.

Because the GWR regression equation is calibrated independently for each observation (each GND in the current scenario), a separate parameter estimate, t-value, and goodness-of-fit is calculated for each observation. These values can thus be mapped, allowing the analyst to visually interpret the spatial distribution of the nature and strength of the relationships among explanatory (local environmental and socioeconomic variables for each GND) and the dependent variables (incidence rate of dengue cases per GND).

We examined both the OLS and the GWR to determine which method would provide a better fit to the observations. This was determined through the results of the AICc. The AIC resulting from GWR was compared to the AIC resulting from OLS. Since the AIC of GWR (836.9) was lower than the AIC of OLS (848.4), GWR provided a better fit to the observed data.

Before applying the GWR, the explanatory or independent variables were identified using the explanatory regression in ArcGIS. Exploratory regression tool tries every combination of possible variables that explain the variability in the dependent variable. It identifies the model that satisfies all the threshold criteria for minimum acceptable adjusted R squared (R^2), Maximum Coefficient p-value Cutoff, Maximum VIF value cutoff and Minimum Acceptable Jarque-Bera p-value. It also runs the Spatial Autocorrelation (Global Moran's I) tool on the model residuals to see if the under/over predictions are not clustered. The explanatory regression process identified two models with three variables each that best predicted the dengue incidence rates. These variables included proportion of piped-water-supply, proportion of households with brick walls, housing density and vegetation cover.

Thus, for GWR, we included incident dengue cases per GND as the dependent variable and the proportion of piped-water supply, proportion of households with brick walls, housing density and vegetation cover as the explanatory variables. The kernel was specified as a fixed distance to solve each regression analysis. The bandwidth was specified as AIC (ESRI, 2010) to determine the extent of the kernel. This was the bandwidth or the number of neighbors used for each local estimation, and was the most important parameter for the GWR as it controlled the degree of smoothing in the model. The significance between the predictor variables and the dengue IR was determined by t-values. The t-values were estimated by dividing the coefficient of parameter estimate by the coefficient of standard error. Values above 1.96 and below -1.96 were considered as statistically significant.

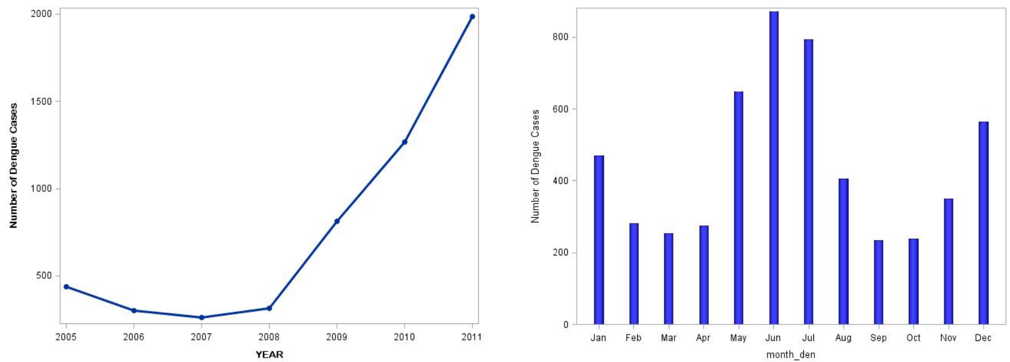
5.0 RESULTS

In this chapter, we present the results of all the analyses discussed in Chapter 4. The results begin with descriptive statistics, followed by bivariate analysis and multivariable regression models, characterizing the environmental factors and dengue cases. We present the results in two parts: Part I will characterize the relationship between climate variables and dengue incidence; Part II will present the results of the non-spatial and spatial analysis evaluating the relationship between local environmental factors and dengue incidence rate.

5.1 RELATIONSHIP BETWEEN CLIMATE VARIABLES AND DENGUE INCIDENCE (PART I)

5.1.1 Descriptive statistics

Over the seven-year period of the study from 2005 to 2011, a total of 5,379 cases of dengue were reported. The number of reported cases of dengue varied by year. The highest number of dengue cases was reported in 2011 and the lowest in 2007. There was a gradual increase in the overall number of cases from 2008 onwards (Figure 13a). In each year, the dengue cases showed a similar pattern of occurrence. The total number of cases peaked during the south-west monsoon season which lasts from May to September (Figure 13b). There was another peak in dengue cases in December and January which coincides with the North-West monsoon season and precede the inter-monsoon season (Figure 13b).



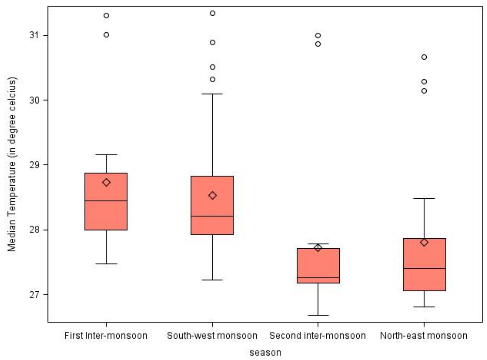
(a) Total number of reported dengue cases, by year (2005-2011), CMC area, Sri Lanka (n=5379)

(b) Total number of cases reported, by month (January to December), CMC Area, Sri Lanka (n = 5379)

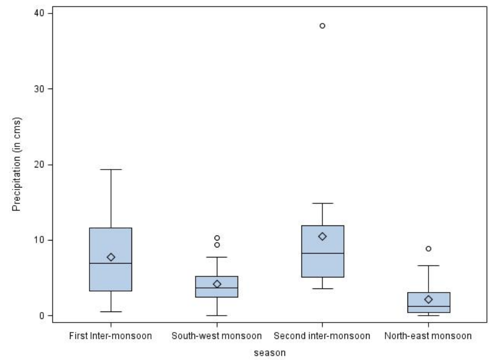
Figure 13: Distribution of dengue cases

Figure 14 displays the variation in weekly average of temperature, cumulative rainfall and total number of dengue cases across the four seasons. The mean temperatures are higher in first inter-monsoon and south-west monsoon seasons as compared to second inter-monsoon and north-east monsoon seasons. The rainfall show cyclical pattern with high rainfall during the first and second inter-monsoon seasons and low rainfall during the south-west and north-west monsoon seasons. There is no clear relationship observed between the temperature and total dengue counts. However, there is distinct pattern with rainfall; the dengue incidence peaks during south-west and north-east monsoon follow the high rainfall periods in the preceding seasons.

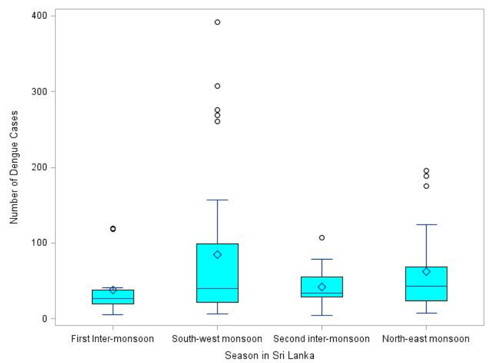
The average weekly counts of dengue cases were reported as 14.7 cases ($SD =$ 18.3; $median = 8.0$; $range = 124$) (Table 9). The average weekly mean temperature was reported as $28.2°C$ ($SD = 1.3$; $median = 27.9°C$; $range : 7.4°C$). The highest average weekly mean temperature was reported in year 2011 ($30.7°C$). The average weekly cumulative precipitation was about $6.8cms$ ($SD = 9.7$; $median$=4; $range =$ 69).



(a) Temperature



(b) Rainfall



(c) Total dengue counts per week

Figure 14: Box plots for temperature, rainfall and total dengue counts across four seasons

Table 9: Distribution of temperature, rainfall and dengue cases

Variable	Mean (SD)	Median	Range
Weekly average of mean temperature (°C)	28.2 (1.3)	27.9	7.4
Weekly average of cumulative rainfall (in mm)	6.8 (9.7)	4	69
Weekly counts of DF cases	14.7 (18.3)	8	124

DF= dengue fever; SD = standard deviation

5.1.2 Correlations

The results of the correlations between temperature, precipitation and dengue cases at lags 0 to 25 weeks are presented in Table 8. We found that the total weekly counts of dengue cases were moderately correlated with temperature from lag 0 to lag 18 weeks; and weakly correlated with precipitation between lag 5 and 11 weeks. The temperature and precipitation have weak negative correlation with each other for the same week but it approaches to zero with weeks further along. Table 11 display the correlations between rainfall and temperature at various lag periods.

Table 8: Correlation between weekly mean temperature (°C), average cumulative rainfall (in mm) and DF cases

Lag periods	Temperature	Rainfall
(in weeks)	Correlation Coefficient (R^2)	Correlation Coefficient (R^2)
0	0.45*	-0.07
1	0.44*	-0.04
2	0.46*	-0.01
3	0.47*	0.04
4	0.47*	0.08
5	0.48*	0.13*
6	0.49*	0.17*
7	0.49*	0.19*
8	0.53*	0.15*
9	0.51*	0.15*
10	0.51*	0.15*
11	0.52*	0.14*
12	0.52*	0.08
13	0.50*	0.06
14	0.43*	-
15	0.51*	-
16	0.49*	-
17	0.45*	-
18	0.40*	-
19	0.36*	-
20	0.32*	-
21	0.21*	-
22	0.21*	-
23	0.16*	-
24	0.16*	-
25	0.10	-

*Significant $p < 0.05$

Table 11: Correlation between rainfall and temperature at various lag periods

Lag period (in weeks)	Temperature$^{\Upsilon}$	Lag period (in weeks)	Rainfall$^{\Psi}$
Rainfall at lag 0	-0.26*	Temperature at lag 0	-0.26
" at lag 1	-0.15*	" at lag 1	-0.06
" at lag 2	-0.09	" at lag 2	0.00
" at lag 3	-0.08	" at lag 3	-0.01
" at lag 4	-0.09	" at lag 4	0.02
" at lag 5	-0.08	" at lag 5	0.00
" at lag 6	-0.06	" at lag 6	-0.03
" at lag 7	-0.13*	" at lag 7	0.01
" at lag 8	-0.07	" at lag 8	0.02
" at lag 9	-0.10	" at lag 9	0.00
" at lag 10	-0.09	" at lag 10	0.00
" at lag 11	-0.04	" at lag 11	
" at lag 12	-0.08	" at lag 12	
" at lag 13	-0.04	" at lag 13	
" at lag 14	0.00	" at lag 14	
" at lag 15	0.03	" at lag 15	
" at lag 16	0.03	" at lag 16	
" at lag 17	0.07	" at lag 17	
" at lag 18	0.11	" at lag 18	
" at lag 19	0.06	" at lag 19	
" at lag 20	0.06	" at lag 20	
" at lag 21	0.04	" at lag 21	
" at lag 22	0.12*	" at lag 22	
" at lag 23	0.10	" at lag 23	
" at lag 24	0.06	" at lag 24	
" at lag 25	0.06	" at lag 25	

$^{\Upsilon}$Temperature at lag 0 is correlated with rainfall at various lag periods.$^{\Psi}$Rainfall at lag 0 is correlated with temperature at various lag periods.

5.1.3 Multivariable regression model

The results of the regression analysis suggested a weak association between weekly mean temperature, precipitation and dengue incidence after adjusting for seasons and year of dengue incidence. In the initial analysis, only one climate variable was included in the model adjusting for season and year of dengue incidence. The relative risk of dengue fever for temperature ranged from 1.11 to 1.21 for lag periods 0 to 16 weeks, and became non-significant for lagged 17th week (Table 9). The relative risk of dengue incidence for weekly average rainfall significant from lag 3 to lag 12 and ranged between 1.02 to 1.08 (Table 10).

In the final model, we evaluated the effect of temperature at various lag periods and rainfall at lag 8th week, adjusting for seasons and year of dengue incidence. The results found no significant changes in the relative risk of dengue fever for all other lag period except for lag 8 (Table 11). Similarly, we evaluated the effect rainfall at various lag periods and temperature at lag 7. We found no significant changes in relative risk between one climate variable model and two climate variables model (Table 12).

5.1.4 Results of sensitivity analysis

Results of the analyses using the LDAS data for temperature and TRMM data for rainfall found similar results as the original analysis. Results of the sensitivity analyses are included in appendix 2.

Table 12: Relationship between weekly average temperature and total weekly dengue counts (CMC data)

Variable	Crude RR^{ϕ}	95% CI		Pr > ChiS
Temperature (°C) Lag 0^{δ}	1.02	0.98	1.05	0.39
Lag 1	1.03	0.99	1.07	0.10
Lag 2	1.11	1.07	1.14	<.0001
Lag 3	1.11	1.07	1.14	<.0001
Lag 4	1.12	1.07	1.14	<.0001
Lag 5	1.15	1.09	1.16	<.0001
Lag 6	1.15	1.11	1.18	<.0001
Lag 7	1.18	1.12	1.18	<.0001
Lag 8	1.21	1.15	1.22	<.0001
Lag 9	1.15	1.17	1.24	<.0001
Lag 10	1.17	1.12	1.19	<.0001
Lag 11	1.15	1.13	1.20	<.0001
Lag 12	1.14	1.12	1.19	<.0001
Lag 13	1.15	1.10	1.17	<.0001
Lag 14	1.11	1.11	1.18	<.0001
Lag 15	1.06	1.08	1.15	<.0001
Lag 16	1.03	1.03	1.09	<.0001
Lag 17	1.02	1.00	1.05	0.06
Lag 18	1.02	0.99	1.04	0.18

§ Weekly average temperature with lag period in weeks. GENMOD, DIST=POISSON, LINK=LOG model. Model includes total dengue count (dependent variable), temperature (independent), dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent). ¥Risk ratios and 95% confidence intervals for an increase of approximately 1°C.

Table 13: Relationship between weekly average rainfall and total weekly dengue counts (CMC data)

Variable	Crude RR^{ϕ}	95% CI		Pr > ChiS
Rainfall (mm)	1.00	0.98	1.01	0.39
Lag 0^{δ}				
Lag 1	1.01	1.00	1.03	0.10
Lag 2	1.02	1.00	1.03	<.0001
Lag 3	1.04	1.03	1.05	<.0001
Lag 4	1.04	1.03	1.06	<.0001
Lag 5	1.06	1.09	1.08	<.0001
Lag 6	1.08	1.06	1.09	<.0001
Lag 7	1.08	1.07	1.09	<.0001
Lag 8	1.06	1.05	1.08	<.0001
Lag 9	1.05	1.03	1.06	<.0001
Lag 10	1.05	1.03	1.06	<.0001
Lag 11	1.04	1.03	1.05	<.0001
Lag 12	1.02	1.00	1.03	0.01
Lag 13	1.01	1.00	1.02	0.12

$^{\delta}$Weekly average temperature with lag period in weeks.
GENMOD, DIST=POISSON, LINK=LOG model.
Model includes total dengue count (dependent variable), temperature (independent), dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent).
$^{\phi}$Risk ratios and 95% confidence intervals for an increase of approximately 1°C.

Table 14: Relationship between weekly average temperature and weekly total dengue fever counts when adjusted for rainfall

Variable	Adjusted RR^{ϕ}	95% CI		Pr > ChiS
Temperature (°C)$^{\delta}$				
Lag 4	1.16	1.13	1.20	<.0001
Lag 5	1.21	1.17	1.25	<.0001
Lag 8	1.33	1.29	1.38	<.0001
Lag 10	1.16	1.13	1.19	<.0001
Lag 15	1.13	1.10	1.16	<.0001

$^{\delta}$Weekly average rainfall with lag period in weeks.
GENMOD, DIST=POISSON, LINK=LOG model. Model includes total dengue count (dependent variable), temperature (independent), weekly average rainfall at lagged at week 8, dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent).
$^{\phi}$Risk ratios and 95% confidence intervals for an increase of approximately 1°C.

Table 15: Relationship between weekly average rainfall and weekly total dengue fever counts when adjusted for temperature

Variable	Adjusted RR^{ϕ}	95% CI		Pr > ChiS
Temperature (°C)$^{\delta}$				
Lag 0	1.00	0.98	1.01	<.0001
Lag 5	1.06	1.05	1.07	<.0001
Lag 8	1.07	1.06	1.08	<.0001
Lag 10	1.06	1.05	1.07	<.0001
Lag 15	0.97	0.95	0.98	<.0001

$^{\delta}$Weekly average rainfall with lag period in weeks.
GENMOD, DIST=POISSON, LINK=LOG model. Model includes total dengue count (dependent variable), temperature (independent), weekly average temperature at lagged at week 9, dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent).
$^{\phi}$Risk ratios and 95% confidence intervals for an increase of approximately 1°C.

5.2 RELATIONSHIP BETWEEN LOCAL ENVIRONMENTAL VARIABLES, AND DENGUE INCIDENCE RATE (PART II)

5.2.1 Non-spatial analysis

5.2.1.1 Descriptive statistics. A total of 5,555 confirmed cases of dengue fever were reported between the years 2005 and 2011, from the 55 GNDs in CMC area, Sri Lanka. Of these, 5379 cases were successfully geocoded and included in the analysis. For descriptive analysis we presented results by comparing characteristics of dengue cases across two groups 'above median' and 'below median' incidence rate per GND (n= 937 per 100,000 population) over the seven year period. Table 16 provides the descriptive statistics for dengue cases. The average age among all dengue cases ranged from 0.1 to 89 years. The mean age was 13.7 years (SD=13.7). Almost 75% of the total cases were children. Distribution of cases between the above median and below median groups found that the average age of cases in the above median group $(14.9; \pm 14.3\ )$ was older than the average age of cases in the below median group (12.1 ± 12.8), $(p = 0.03)$. Among all the cases, about 54 % cases were male and 46 % were females. There were no significant gender differences among the cases in both above and below median groups.

Table 16: Characteristics of dengue cases across 'above median' and 'below median' dengue incidence rate per GNDs

Variables	Overall	> Median$^{\Phi}$	$\leq$Median$^{\Phi}$	P-value
Number of Cases (n)	5379			
Number of GND	55			
Age (mean, SD)	13.7	14.9 (14.3)	12.1 (12.8)	<0.0001
Age range (in years)	0.1-89	0.1 -89	0.1 -81	
Age Categories (n, %)				
0-5	1688 (31.4)	884 (28.6)	804 (35.2)	
5.1 - 9	1222 (22.7)	687 (22.2)	535 (23.4)	
9.1 to 19	1168 (21.7)	643 (20.8)	525 (23.0)	
>19	1302 (24.2)	882 (28.5)	419 (18.4)	<0.0001
Sex				
Males	2897 (53.9)	1658 (46.5)	1239 (45.7)	
Females	2482 (46.1)	1438 (53.6)	1044 (54.3)	0.6017

$^{\Phi}$Median incidence rate: 937 per 100,000 population; SD = standard deviation; GND= Gram niladari divisions

5.2.1.2 Bivariate analysis. The comparison for local environmental characteristics between 'above median' and 'below median' groups found that proportion of households with brick walls (p = 0.0001) and the proportion of households with cement walls (p = 0.0169) were significantly different between the two groups (Table 14). 'Above median' incidence rate GNDs' had higher proportion of households with brick walls as compared to below median group; while the reverse was true for households with cement walls. piped-water supply, population density and housing density were significantly different between the two groups at $p < 0.05$.

Table 17: Distribution of local environmental factors across GNDs above and below dengue median incidence rate

Variables	Overall	$>$ Median$^{\theta}$	$\leq$ Median$^{\theta}$	P-value
Environmental Characteristics (mean, SD)				
Buildings	0.5 (0.1)	0.5(0.1)	0.5(0.1)	0.15
Vegetation	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.71
Roads	0.1 (0.04)	0.1 (0.1)	0.1 (0.1)	0.73
Shadow	0.1 (0.1)	0.1 (0.04)	0.1 (0.04)	0.46
Green Space	0.04 (0.1)	0.04 (0.1)	0.04 (0.03)	0.95
Household Characteristics				
Brick Walls	0.4 (0.1)	0.6 (0.1)	0.5 (0.1)	0.001*
Cement Walls	0.6 (0.1)	0.3 (0.1)	0.4 (0.1)	0.02*
Other wall materials	0.1 (0.1)	0.1(0.1)	0.1 (0.7)	0.52
Tile Roofs	0.1 (0.1)	0.2 (0.1)	0.1 (0.1)	0.06
Concrete roofs	0.3 (0.1)	0.3 (0.2)	0.3 (0.1)	0.37
Asbestos Roof	0.5 (0.1)	0.5 (0.1)	0.6 (0.1)	0.39
Other wall materials	0.1 (0.1)	0.04 (0.1)	0.1 (0.1)	0.03*
Toilets exclusively for household	0.7 (0.1)	0.8 (0.2)	0.7 (0.1)	0.54
Toilets shared	0.3 (0.1)	0.3 (0.1)	0.3 (0.1)	0.54
Piped-water supply	0.8 (0.1)	0.8 (0.1)	0.9 (0.1)	0.03*
Population Characteristics				
Population density (per 1000.sqm)	20 (12.7)	16.8 (11.8)	23.8(19.2)	0.04*
Housing density	4.2 (2.5)	4.9 (2.2)	3.5 (2.6)	0.03*

$^{\theta}$Median incidence rate of DF: 937 per 100,000 population;

5.2.1.3 Regression model. All the risk factors significant in the bivariate analysis were included in the multivariable regression analysis (Table 18). Results of the adjusted model found that piped-water was negatively associated with dengue incidence rate ($RR = 0.90$; 95% $CI = 0.87 - 0.92$); higher proportion of brick walls was associated with increased risk ($RR = 1.04$; 95%$CI = 1.00 - 1.07$). Since housing density and population density were highly correlated (r^2=0.95; $p = 0.001$),we included both the variables one at a time in the model. Higher housing density was negatively associated with increased dengue incidence rate ($RR = 0.997$; 95%$CI = 0.96 - 1.03$).

Table 18: Results of multivariable regression model

Variables*	Risk Ratio	95% CI		P-value
Piped-water supply	0.90	0.87	0.92	<.0001
Brick wall	1.04	1.00	1.07	0.03
Housing density	0.997	0.995	0.998	0.0011
Cement wall	0.99	0.96	1.03	0.76

*Variables included as continuous in the model; Model: Poisson regression with repeated statement QIC = 5124.9

5.2.2 Spatial analysis

5.2.2.1 Monthly spatio-temporal models for 2005-2011. The results are presented on a monthly basis for 84 months (January 2005 to December 2011) with Figures 15, 16 and 17, indicating the location of the hotspots. The spatio-temporal dengue risk was categorized into four risk levels. Risk levels 1 denoted 'very low risk' for dengue incidence rate; 2 denoted 'low risk', 3 denoted 'medium risk' and 4 denoted 'very high risk'. The spatio-temporal monthly hotspots were distributed in GNDs located in the center of the CMC area for most months. The maps also demonstrated that in years 2005, 2007, 2008 and 2009, the patterns for the risk levels of dengue incidence rate were almost similar. Risk patterns for years 2010 and 2011 were similar, but different than previous years. For example, for Kurunduwatta GND, located in the center of the CMC, the risk pattern in the month of January was low-risk for years 2005, 2007 and 2010. However, it was very-high risk in years 2006, 2008 and 2011. The GND had a medium risk level in 2009. Looking at the pattern changes over the span of the year for Kurunduwatta, for year 2011, a risk level 4 was observed in the months of January, March, April, June, and July. Risk level 3 was found in month of August. Low, and very-low risks levels were found in months of May, and September, October, November and December respectively.

In years 2005,2010 and 2011, the hotspot patterns and high risk areas displayed a shift to new GNDs, not identified in previous years, particularly in the months of June and July.

5.2.2.2 Overall spatio-temporal model based on monthly risk models over the 7-year study period. The average monthly risk over the 7-year period showed most of the hotspots for dengue incidence rates in the central and western region of the CMC area. Risk levels ranged from very low, low to medium. None of the GNDs had a overall average risk above 3. In particular, the risks levels were highest consistently in Kurunduwatta and the Fort, Pettah and Galle Face GNDs (Figure 18).

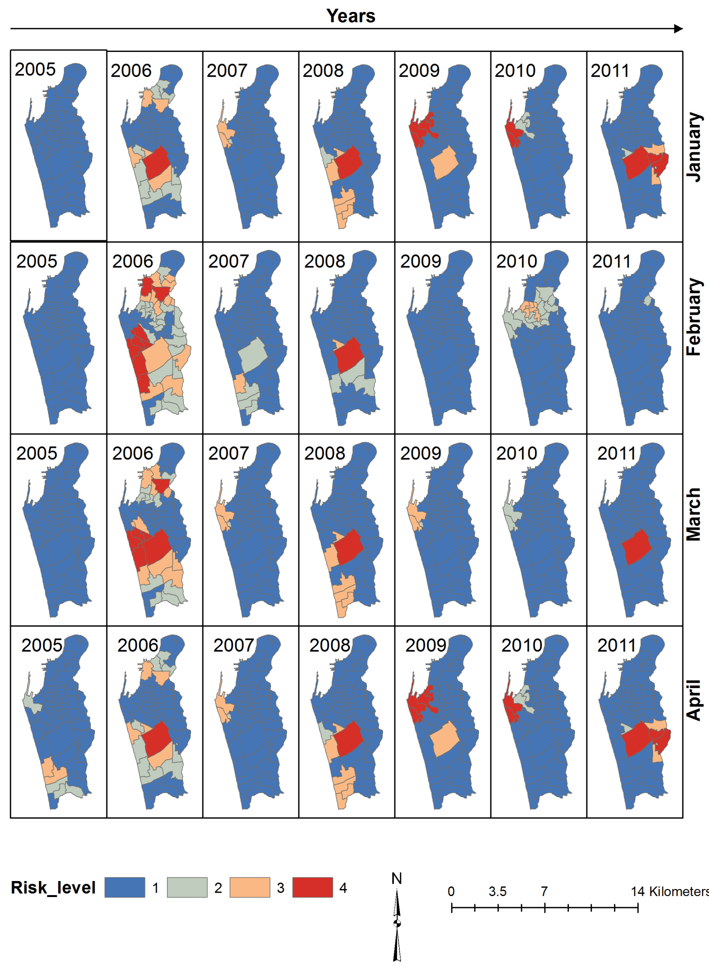


Figure 15: Spatio-temporal monthly risk models of the GNDs, January to April, 2005 to 2011

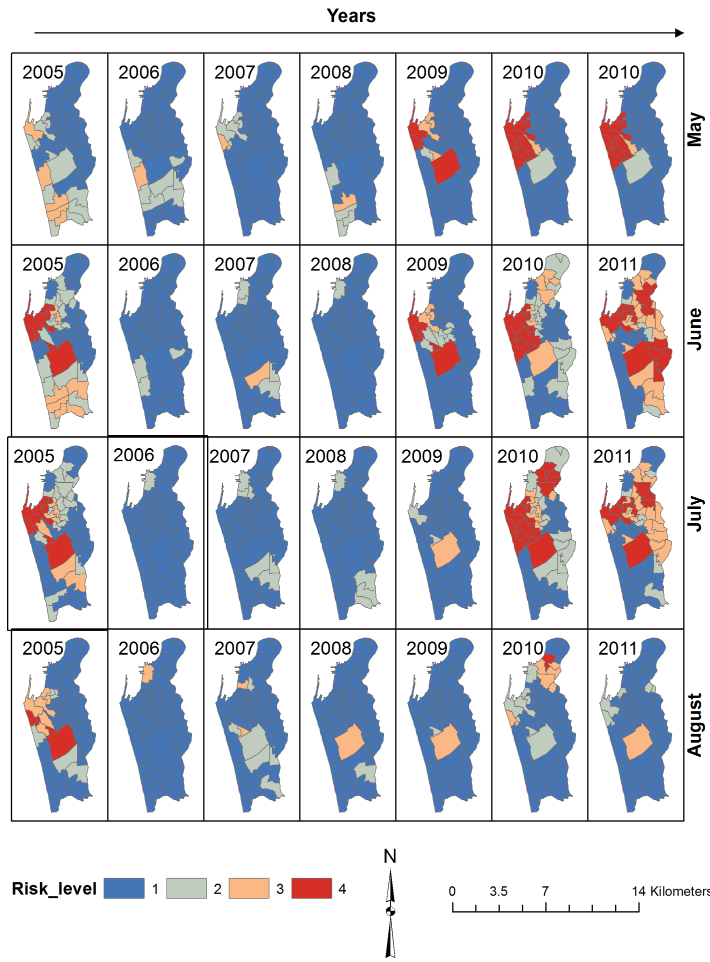


Figure 16: Spatio-temporal monthly risk models of the GNDs, May - August, 2005 to 2011

Figure 17: Spatio-temporal monthly risk models of the GNDs, September - December, 2005 to 2011

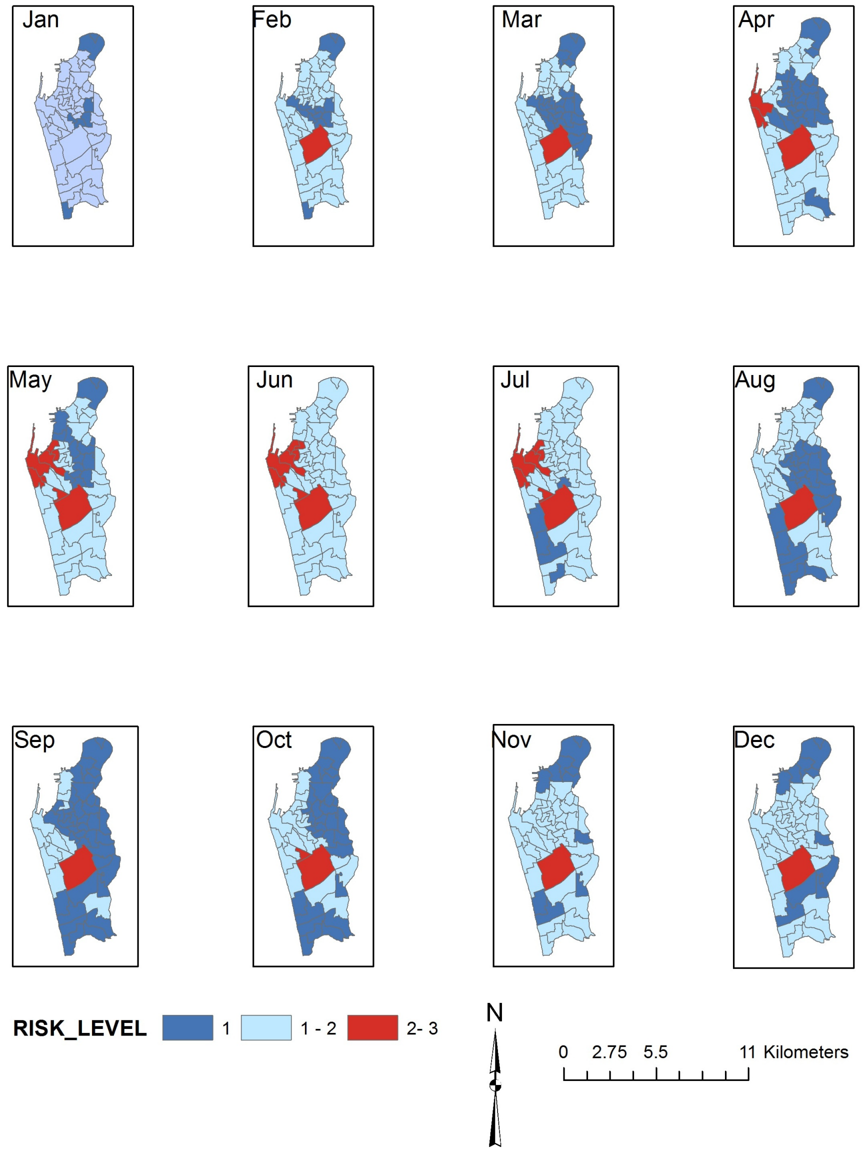


Figure 18: Overall spatio-temporal model of monthly temporal risk over the 7-year study period.Risk level <1 =Low; 1-2 = Medium and 2-3= High

5.2.2.3 Spatial regression analysis. The summary of the independent and dependent variables used in the OLS and GWR are described in Table 19.

Table 19: Variables included in the spatial regression analysis

Dependent variable	Independent variables	
Source: DPH	Source: LULC classes (%)	Source: Census (%)
Incidence rate (IR) of DF per 100,000 population per GND	Buildings	Brick walls
	Vegetation	Cement walls
	Green spaces	Other walls
		Tiled roofs
		Asbestos roofs
		Other roof materials
		Piped-water supply
		Population density
		Housing density

DPH = CMC, Department of Public Health

The results of OLS regression found that proportion of tiled or concrete roofed households, proportion of household with piped-water supply and neighborhood quality ratio were significantly associated with incident dengue cases. The Variance Inflation Factor (VIF) values (lower than 7.5) indicated that the OLS estimations were not biased by multi-collinearity. We further examined the residuals of the OLS model, and found that the residuals had no spatial autocorrelation (Moran's I = 0.06;$p = 0.214$) . The OLS model explained about than the 76% of the total variance of the IR with an AICc of 842.8. The Jarque-Bera Statistic is a test for whether the model predictions are biased (i.e., the residuals are not normally distributed). This test was statistically non-significant suggesting that the residuals of the model were normally distributed. However, the Koenker (BP) statistic was statistically significant ($p = 0.003$),suggesting that the relationships modeled are not consistent (possibility due to non-stationarity or heteroskedasticity). We thus used the GWR to model the relationship between the independent variables and the incident dengue cases locally at each GND level. Two models were selected for GWR with highest adjusted $R^2 = 0.61$and second highest $R^2 = 0.59$.

$$\gamma_{dengue\, incidence\, rate} = \beta_{piped\, water\, supply} + \beta_{Brick\, wall} + \beta_{Housing\, density} \ldots Model\, I$$

$$\gamma_{dengue\, incidence\, rate} = \beta_{piped\, water\, supply} + \beta_{Brick\, wall} + \beta_{Vegetation} \ldots Model\, II$$

Figure 19 provides the spatial description of all the independent and dependent variables included in the final GWR model.

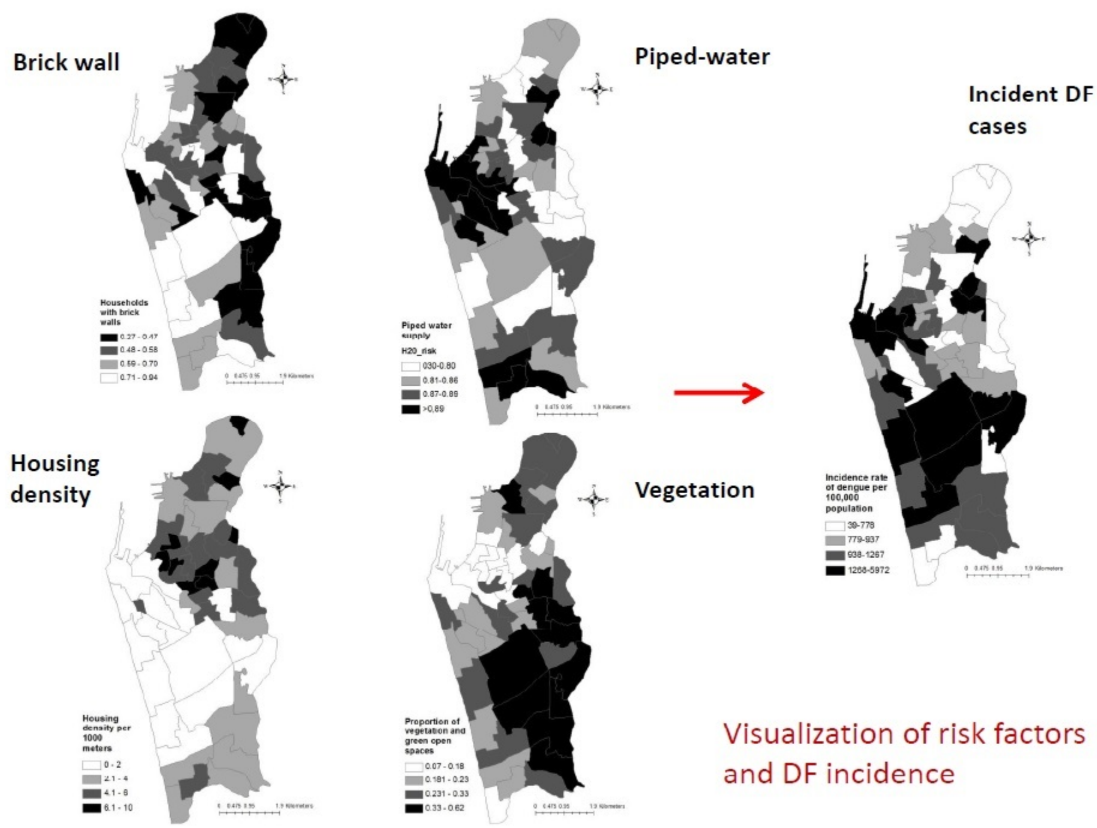


Figure 19: Visualization of environmental risk factors and dengue fever incidence

Results of model I show both the magnitude and direction of association of the parameter estimates. The grey areas indicate that the parameter estimates are not significant for risk factor in that particular area (Figure 20). The results found that proportion of piped-water supply and housing density were negatively associated with dengue incidence rate. This association was significant for all GNDs. The increased proportion of households with brick walls were significantly associated with higher dengue incidence rate but only in one GND. The locally weighted R^2between the observed and the fitted values for each GND in figure 20 indicated how well the GWR model fitted for each GND. GNDs in north-western part of the CMC area showed higher R^2. These areas overlapped the areas with higher magnitude of parameter estimates for piped-water supply.

Results of model II are presented in figure 21. In this model, we included proportion of piped-water supply in household by GNDs, proportion of brick walled households and proportion of area covered by vegetation. As in model I, piped-water supply and brick walled household were negatively and positively associated with increased dengue risk, respectively. Increasing vegetation cover was positively associated with increased dengue incidence rate. All the three risk factors were significant in most parts of the CMC area. The locally weighted R^2 was highest in the north-western parts of the CMC areas as in model I. These areas overlapped with areas with higher magnitudes of parameter estimates for piped-water supply and brick walls.

Overall, results from both model I and II found strong association between lower proportion of piped-water supply and higher proportion of brick walled households and dengue incidence rate. Both housing density and vegetation were weakly associated with dengue incidence rate.

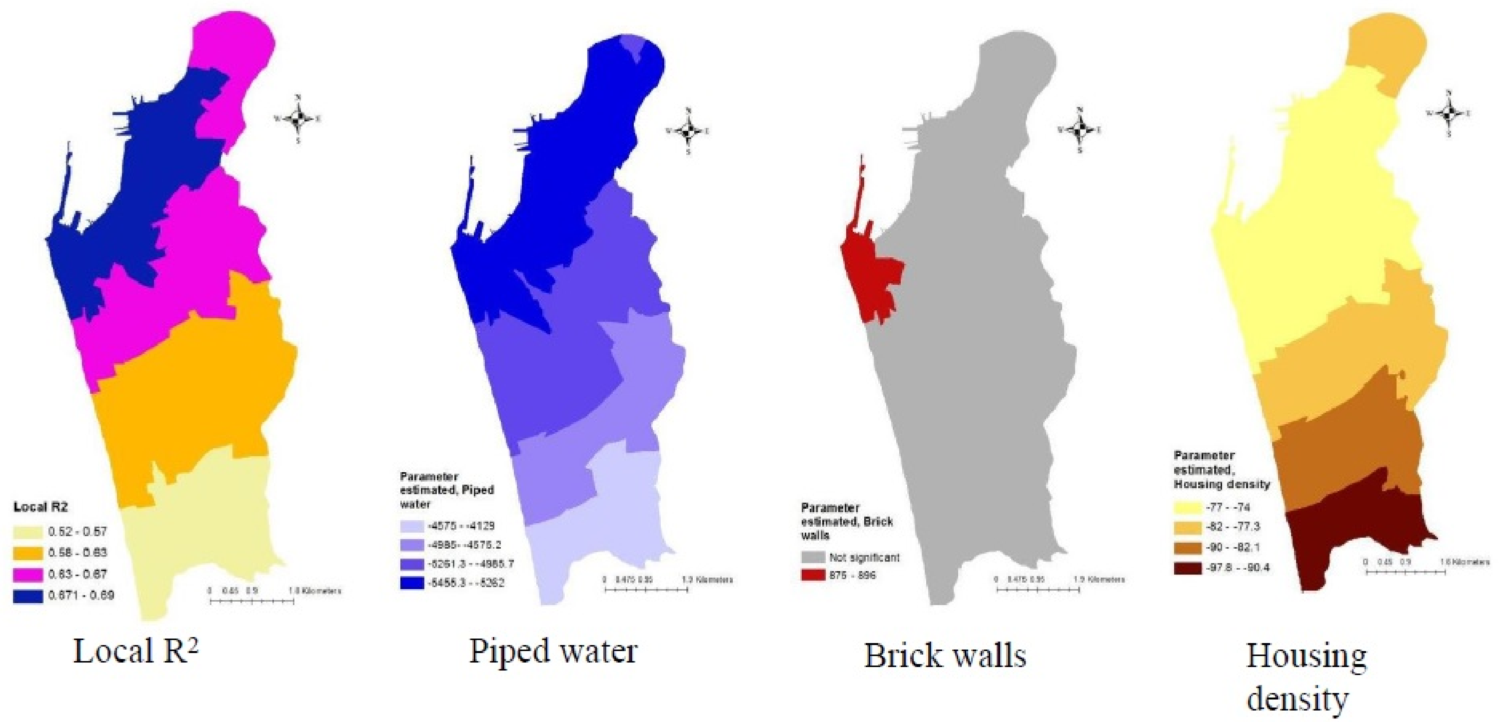


Figure 20: Model I, Choropleth maps displaying both the magnitude and parameter estimate (PE) by GNDs; a grey mask applied to those GNDs where PE that are not significant at 95% CI; GWR (adjusted $r^2 = 0.64$; AICc =839.8)

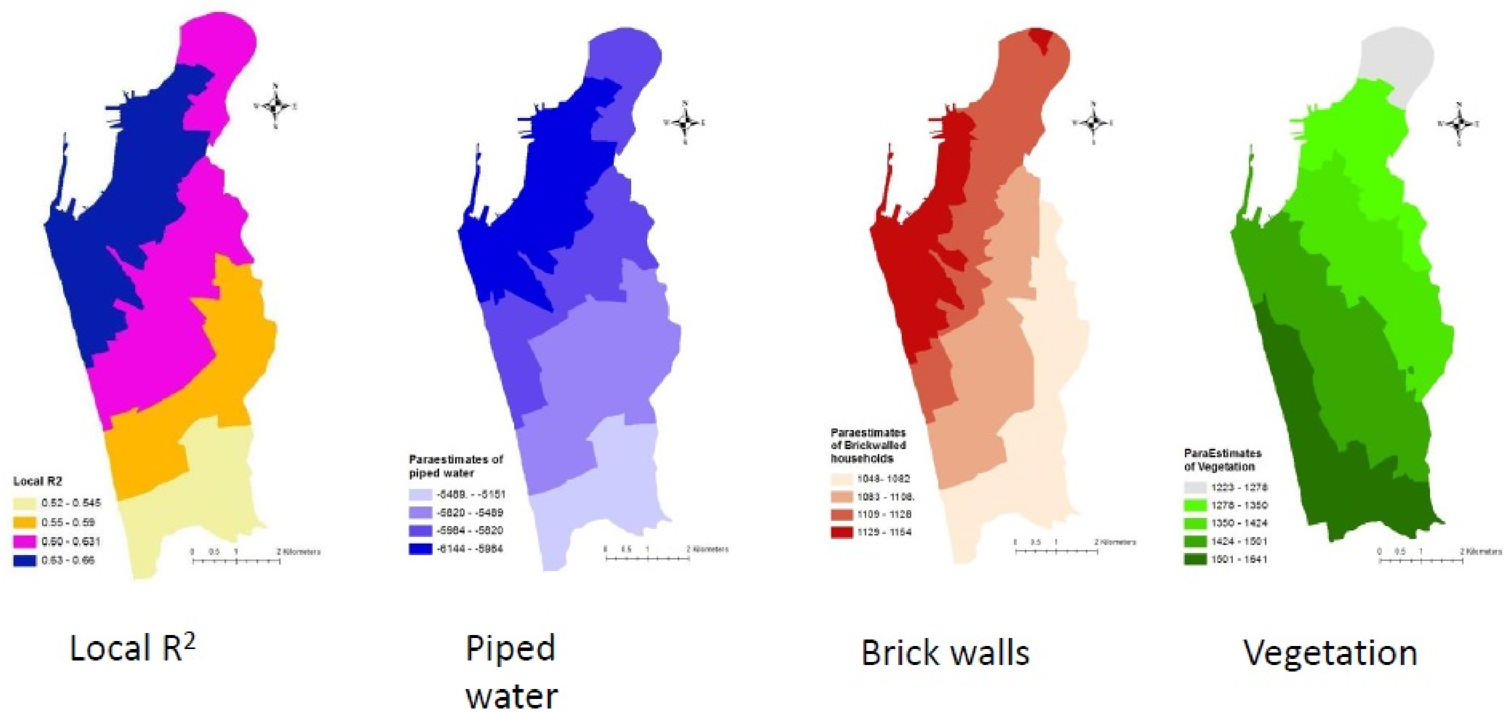


Figure 21: Choropleth maps displaying both the magnitude and parameter estimate (PE) by GNDs; a grey mask applied to those GNDs where PE that are not significant at 95% CI; GWR (adjusted r2 = 0.59; AICc =836.9)

5.2.3 Comparison of results between spatial and non-spatial analyses

The comparison between non-spatial (Poisson) regression and spatial (GWR) analysis found that results of GWR were mostly consistent with Poisson regression (Table 20). In both analyses, piped-water supply and brick walled households were significantly associated with increased dengue incidence rate. Housing density was negatively associated with dengue incidence rate in Poisson regression and in GWR model I. However, the association was weak in both cases. Increased vegetation in an area was significantly associated in GWR model II but not in the non-spatial analysis.

Table 20: Comparison between non-spatial (Poisson) and spatial analysis (GWR)

Type of analysis	Piped-water supply	Brick wall	Housing density	Vegetation
Non- spatial	Negative[£]	Positive[£]	Negative	No association
Spatial (model I)	Negative[£]	Positive[#]	Negative	-
Spatial (model II)	Negative[£]	Positive[£]	-	Positive[¥]

GWR = Geographic weighted regression; * = Not included in the model [#] Significant for 7% GNDs at p = 0.05; [£]Significant for 100% GNDs; [¥] Significant for 93% GND

6.0 DISCUSSION

The main purpose of this research was to investigate the association between climate, local environmental and socio-demographic risk factors, and dengue incidence rate in CMC, Sri Lanka. The study was done with the expectation that the results will delineate the areas of high risk for dengue incidence within the CMC area, using data that is currently available to the public health authorities. The study utilized several secondary data sources that were publicly available to extract information on risk factors. It also demonstrated the application of innovative tools such as remote sensing, GIS and spatial statistics for modeling, and visualization of relationships between the risk factors and dengue incidence rates.

The study found that: 1) there is a weak association between weekly temperature and rainfall and higher risk of dengue. The effect of climate variables is significant at various lag periods but does not vary much in magnitude; 2) it identified monthly hotspots of dengue risks based on the z-scores resulting from Getis-Ord Gi* for each year in the study period as well as the estimated the overall spatio-temporal risk; and 3) it identified local environmental factors such as decreased piped-water supply, increased brick-walled housing, decreased housing density and increased vegetation were significantly associated with high incidence of dengue fever cases.

6.1 EFFECTS OF CLIMATE ON DENGUE

This study found that increase in temperature preceded increase in dengue cases from week 0 to week 18, but the effect remained constant. Increase in rainfall preceded the increase in dengue cases by 5 to 11 weeks but the risk was found to be constant across the period. Other studies have shown positive association between temperature and rainfall and increase in dengue cases (Hii et al., 2009, 2012; Karim et al., 2012; Johansson et al., 2009). These studies also found a lag effect of temperature and rainfall on dengue incidence ranging between 8 to 15 weeks. Studies in Sri Lanka reported positive association between climate variables and dengue vectors and dengue incidence, but the results were non-significant. Surendran et al. (2007a) found a positive association between monsoon rains and increase in *Ae. Aegypti* and *Ae. Albopictus* populations in northern Sri Lanka. Similarly, Pathirana et al. (2009) found a positive association with increased rainfall and increased vector population; but the association became non-significant when other factors such as age and SES were included in the model.

Dengue infections are climate sensitive and so it is important to better understand how the changing climate factors affect the potential for dengue epidemics. Studies have found that temperatures between 26řC and 36řC were highly conducive for vector development (Morin et al., 2013). Others have reported that increasing temperature decreases the time required for the dengue virus to become transmissible to another host after initial mosquito infection (Rohani et al., 2009; Chang et al., 1997; Lambrechts et al., 2011). Lambrechts et. al., (2011) suggested that large changes in the daily temperature decrease the probability of vector infection. The average daily temperature in Sri Lanka ranged from 23.6 to 36.10 (mean = 28.2; SD = 1.45), for years 2005 to 2011. The daily temperature fluctuation between minimum and maximum ranged between 0 to 14 with a mean of 5°C. This suggest a favorable environment for vector infection for most part of the year.

The climate in Sri Lanka is dominated by two monsoon seasons within a year and two pre-monsoon season. The abundant rainfall in Sri Lanka sustains the mosquito population creating ample breeding sites. During the first inter-monsoon season, (March and April), there are warm conditions with thunderstorm-type rains. Over most parts of the island, the amount of rainfall various between 10 cm and 250 cm . In our study, we found that the median rainfall during the first inter-monsoon season was highest as compared to other seasons between 2005 to 2011. This period preceded the season when the dengue incidence was highest across all the other seasons (see Figure 14). During the Southwest monsoon season, varies from about 10 cm to over 30 cm over the entire country. The area where the study was conducted, experience rains between 10 cm to 16 cm.

In our study, we found that the precipitation during the south-west monsoon season ranged between 0 to 10 cm. This season coincided with peaks in the incidence of dengue cases across all years. Studies have shown that very high rainfall potentially washes away the breeding areas for dengue. Even though, the relationship between the rainfall and dengue incidence is weak in our study, the lag of 5-11 weeks suggest that amount of rainfall in the first inter-monsoon season and the reduction of rainfall in the south-west monsoon season influences the incidence of dengue. Similarly, increased rainfall during the second-inter monsoon season, followed by drier period during the Northeast monsoon season results in another peak in the dengue incidence.

The effects of the mean temperature was found to be uniform on the dengue incidence across the year. Increase in temperature 8-9 weeks prior showed slight increase in risk of dengue incidence. This corresponds to the two dengue peaks observed in the monsoon seasons each year, following the higher temperatures during the inter-monsoon season.

Results of the sensitivity analysis using alternate data sources, were similar to the original results. The analysis allowed comparison between data from single station (i.e., from CMC area) with reanalyzed data (LDAS and TRMM data) that take in to consideration large scale conditions that may influence weather in a given area.

The frequency and span of weeks with high temperature and moderate rainfall play an important role in sustainable dengue transmission. Therefore, monitoring of these weather conditions can provide early warning for dengue outbreaks. The time lags occurring between the exposure to temperature and rainfall and the occurrence of increasing dengue case or outbreaks offers a window for dengue forecast. However, further investigations with additional data both in terms of time and space are needed to develop a strong forecast model.

6.2 SPATIO-TEMPORAL RISK MAPPING

The spatio-temporal mapping of dengue risk was estimated by using the hotspot analysis, for each month, for all years in the study (2005 to 2011). The hotspot analysis modeled the monthly incidence rates of each GND. The study identified extreme index of hotspots across the 55 GNDs in the CMC area, thus improving the visualization of the spread of disease over time and space. These results provide useful information about the seasonal trends in the incidence of dengue fever as well as the overall temporal trend during the study period. The monthly hotspot patterns of dengue incidence were similar for most months but changes did occur. In particular, the monthly risk patterns changed in years 2010 and 2011 reflecting the increase in the overall dengue incidence rates during those two years as compared to the previous years.

The results showed that these methods and tools can be beneficial for public health officers to visualize and understand the distribution and trends of diffusion patterns of diseases and to prepare warnings and awareness to the community (Khormi et al., 2011). To our knowledge, this is the first study in Sri Lanka which has modeled monthly spatio-temporal risks of dengue fever. Dengue spatio-temporal diffusion patterns and hotspot detection may provide useful information to support public health officers to control and predict dengue spread over critical hotspot areas only rather than the entire CMC area.

6.3 LOCAL ENVIRONMENTAL AND SOCIO-DEMOGRAPHIC VARIABLES

The results of non-spatial and spatial analysis found decreased piped-water supply, increased brick-walled housing, decreased housing density and increased vegetation were significantly associated with high incidence of dengue fever cases. Previous studies have investigated one or more of these factors in relation with dengue incidence and found similar results in most cases.

The number of households with piped-water supply was negatively associated with dengue incidence. These results are consistent with studies in Vietnam (Schmidt et al., 2011), India (Fulmali et al., 2008), and other countries (Bowman et al., 2014). In Sri Lanka, one study investigated the relationship between decreased piped-water supply and *Ae. Aegypti* population and found positive but non-significant results (Arunachalam et al., 2010). These studies have found that areas with poor tap water supply have higher incidence rate of dengue fever. People without tap-water tend to store water for drinking and cleaning purposes. These storage containers possibly become breeding sites for *Ae.Aegypti*.

In our study, the association between piped-water supply and dengue incidence was significant across all GNDs, suggesting that it was a strong predictor of dengue fever in the CMC area. This information can help the public health authorities to conduct dengue surveillance targeted in areas with limited or no piped-water supply. Piped-water supply is also an indicator of socio-economic status (SES) in an urban population. In a metropolitan city like Colombo, almost 70% of the total population has piped-water supply in their own units. Houses in slums and lower SES neighborhoods are more likely to have tap-water supply outside the units or outside the premises of their housing. It is possible that houses with access to well water supply belong to higher SES. We evaluated the relationship between well-water supply and dengue incidence in a separate analysis but did not find any association.

Our study found that GNDs with higher proportion houses with brick walls had higher incidence rate of dengue fever. Brick-walled housing has not been associated with dengue incidence previously. However, we included it in our study as potential surrogate for SES. In Sri Lanka, concrete plastered brick walls may indicate high SES, while unplastered could indicate lower SES. SES may also depend of the quality of bricks used. For example, hand-made bricks are used by families with lower incomes as compared to commercially available bricks purchased by higher income families. However, no data were available to verify this further.

Housing density was inversely associated with dengue incidence. This result is inconsistent with other studies that have found that closely spaced housing units such as apartment complexes and walled to walled housing are likely to have higher incidence of dengue (Schmidt et al., 2011; Vanwambeke et al., 2007; Khormi and Kumar, 2012). Dengue vectors have a short flight distance (ranging from 50 meters to 100 meters) in densely populated urban areas. Thus, closely packed housing or high population density provides dengue mosqitoes higher number of hosts to feed on, resulting in increased number of dengue infections. Previous studies in Sri Lanka found that increasing population density was associated with increased dengue incidence (Kannathasan et al., 2013). In our study, we found that housing density was highly correlated with population density, and both had similar association on dengue incidence. One of reason we were observing this effect was possibly due to higher number of cases reported from GNDs with the location of National Hospital of Sri Lanka. Several cases in the study provided addresses that were located within the hospital campus. Due to lack of access to individual data other than what was provided by the department of public health, we were unable to assess whether the reported cases were residents within hospital quarters or patients admitted to the hospital. However, this could not be verified further at this time.

Vegetation was found significantly associated with increased dengue incidence in the spatial analysis. Previous studies in Sri Lanka have found that discarded water containers with vegetation cover provide favorable breeding habitat for dengue vectors (Ramasamy et al., 2011). Similarly, studies from Costa Rica and Brazil have found similar results (Barrera et al., 2006; Bisset Lazcano et al., 2006). Localities with less built area and more tree cover were found to have higher virus transmission (Troyo et al., 2009). In Sri Lanka, most affluent houses have backyards with trees and shrubs. It is possible that shade provided by trees in larger backyards and open areas such as parks may protect mosquito habitats from heating and direct sunlight. This can result in higher vector densities in localities with more tree cover(Barrera et al., 2006; Bisset Lazcano et al., 2006)).

6.4 STRENGTHS

The study included comprehensive data on dengue cases, over a period of seven years. Data were collected using standardized forms thus, reducing interviewer bias as well as misclassification of dengue cases. For predictor variables, all data were obtained from secondary sources publicly available and mostly free. This reduced the overall cost of the study.

Results of the climate-dengue analysis were further validated by comparing them with results of alternate data sources for temperature and rainfall. Climate data at the local level was compared with the reanalyzed data, indicating the potential to utilize other data sources such as LDAS and TRMM for future climate studies. We conducted both geospatial and non-spatial analysis to evaluate the relationship between environmental risk factors and dengue fever. In most studies using geospatial tools, results are reported only for the spatial analysis. Consistency of results using both approaches supported internal validity of the study.

6.5 LIMITATIONS

There are certain limitations of this study, which needs consideration while interpreting the results. The study is a ecologic in design. Thus, no temporal or causal relationship can be determined between the risk factors and dengue incidence.

The study was conducted in a small geographical area covering 37 sq. km. This offered limited spatial variability in terms of LULC classes, to discern relationship between various surface characteristics and dengue incidence. Quickbird data, obtained from DigitalGlobe, Inc, used for LULC classification was a high resolution imagery but was limited in spectral resolution to only four bands between visible and near-infra red spectrum. This may lead to misclassification between some of the impervious surfaces such as built-up area and roads. It is possible that with the upcoming newer satellite imagery such as Worldview I and II, we can address some of these limitations. The Worldview satellites have higher spectral, temporal and spatial resolution than the Quickbird data, and are available to the public.

Another limitation in the study was the lack of field validation of LULC classification and geocoding dengue cases. Field validation was not possible as result of limited funding for the study. This is a common limitation of several remote sensing studies, which then rely on Google Earth and pre-exisiting maps to validate their findings.

The data on local environmental factors including LULC classes and population and household characteristics were available only at one point in time. It is possible that changes had occurred in these factors over the 7-year study period and were not taken in to account in our analysis leading to possible misclassification of data for risk factors.

7.0 CONCLUSION

Dengue fever is a significant public health concern in Sri Lanka. The current study is one of the few studies in Sri Lanka that evaluate the relationship between several environmental risk factors and dengue incidence. Only three studies have been conducted in Sri Lanka previously that have used GIS and remote sensing in context of dengue research. The use of geospatial tools for data management and analysis in the current study thus, contributes towards building the evidence-base for dengue research using these tool in the country and in South Asia region.

Overall, the results of our analyses, both spatial and non-spatial are mostly comparable with previous studies. The study explored several risk factors including climate, environmental and socio-demographics to model risk of dengue. Future studies could use similar methods applied in this study to guide rapid population surveys in high risk areas. These survey can help in determining appropriate interventions including behavioral, educational and vector management, to targeted areas where the risk of dengue was found to be high. The study also paves way for population-based epidemiologic studies to validate the findings of this study. It shows the ability to develop vulnerability and risk maps using existing data resources; and the potential to incorporate geospatial tools for routine dengue surveillance in the country.

As part of our agreement with the CMC public health department, we will report the results of the study to the department. The overall approach of this research can be used by the decision makers in CMC department of public health for planning and prioritizing future work in prevention and control of dengue. It could also inform capacity building in use of geospatial tools at the local public health department. With the help of local public health authorities we could build the foundation for developing dengue risk prediction model which in turn will can inform early warning system for dengue.

8.0 BIBLIOGRAPHY

Ali, M., Wagatsuma, Y., Emch, M., and Breiman, R. F. (2003). Use of a geographic information system for defining spatial risk for dengue transmission in Bangladesh: role for Aedes albopictus in an urban outbreak. *The American journal of tropical medicine and hygiene*, 69(6):634–640.

Ali, S., Corner, R. J., and Hashizume, M. (2014). Spatiotemporal Analysis of Dengue Infection Between 2005 and 2010. In Dewan, A. and Corner, R., editors, *Dhaka Megacity, Geospatial Perspectives on Urbanisation, Environment and Health*, pages 367–384. Springer Netherlands, Dordrecht.

Alto, B. W. and Bettinardi, D. (2013). Temperature and dengue virus infection in mosquitoes: independent effects on the immature and adult stages. *The American journal of tropical medicine and hygiene*, 88(3):497–505.

Arunachalam, N., Tana, S., Espino, F., Kittayapong, P., Abeyewickreme, W., Wai, K. T., Tyagi, B. K., Kroeger, A., Sommerfeld, J., and Petzold, M. (2010). Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban Asia. *Bulletin of the World Health Organization*, 88(3):173–184.

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. *The American journal of tropical medicine and hygiene*, 69(2):135–140.

Banu, S., Hu, W., Hurst, C., Guo, Y., Islam, M. Z., and Tong, S. (2012). Space-time clusters of dengue fever in Bangladesh. *Tropical medicine & international health : TM & IH*, 17(9):1086–1091.

Banu, S., Hu, W., Hurst, C., and Tong, S. (2011). Dengue transmission in the Asia-Pacific region: impact of climate change and socio-environmental factors. *Tropical medicine & international health : TM & IH*, 16(5):598–607.

Barrera, R., Amador, M., and Clark, G. G. (2006). Ecological factors influencing Aedes aegypti (Diptera: Culicidae) productivity in artificial containers in Salinas, Puerto Rico. *Journal of medical entomology*, 43(3):484–92.

Bergquist, R. (2011). New tools for epidemiology: a space odyssey. *Memórias do Instituto Oswaldo Cruz*, 106(7):892–900.

Bhandari, K. P., Raju, P. L. N., and Sokhi, B. (2008). Application of GIS modeling for dengue fever prone area based on socio-cultural and environ- mental factors- a case study of Delhi city zone. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 165–170.

Bisset Lazcano, J. A., Marquetti, M. d. C., Portillo, R., Rodríguez, M. M., Suárez, S., and Leyva, M. (2006). [Ecological factors linked to the presence of Aedes aegypti larvae in highly infested areas of Playa, a municipality belonging to Ciudad de La Habana, Cuba]. *Revista panamericana de salud pública = Pan American journal of public health*, 19(6):379–84.

Bohra, A. and Andrianasolo, H. (2001). Application of GIS in Modeling Dengue Risk Based on Sociocultural Data: Case of Jalore, Rajasthan, India. *Dengue Bulletin*, 25:92–102.

Bowman, L. R., Runge-Ranzinger, S., and McCall, P. J. (2014). Assessing the relationship between vector indices and dengue transmission: a systematic review of the evidence. *PLoS neglected tropical diseases*, 8(5):e2848.

Braga, C., Luna, C. F., Martelli, C. M., de Souza, W. V., Cordeiro, M. T., Alexander, N., de Albuquerque, M. d. F. P. M. a., Júnior, J. C. S., and Marques, E. T. (2010). Seroprevalence and risk factors for dengue infection in socio-economically distinct areas of Recife, Brazil. *Acta tropica*, 113(3):234–240.

Census.lk (2014). Department of Census and Statistics-Sri Lanka.

Chang, M. S., Hii, J., Buttner, P., and Mansoor, F. (1997). Changes in abundance and behaviour of vector mosquitoes induced by land use during the development of an oil palm plantation in Sarawak. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 91(4):382–6.

Chen, Y., Wen, D., Jing, L., and Shi, P. (2007). Shadow information recovery in urban areas from very high resolution satellite imagery. *International Journal of Remote Sensing*, 28(15):3249–3254.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46.

Congalton, R. G. . and Green, K. (1998). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition (Mapping Science): Russell G. Congalton, Kass Green: 9781420055122: Amazon.com: Books.

Dare, P. M. (2005). Shadow analysis in high-resolution satellite imagery of urban areas.

Eisen, L. and Eisen, R. J. (2011). Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. *Annual review of entomology*, 56:41–61.

Epidemiology Unit, M. o. H. (2014). Distribution of Notification(H399) Dengue Cases by Month.

Focks, D. A., Daniels, E., Haile, D. G., and Keesling, J. E. (1995). A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *The American journal of tropical medicine and hygiene*, 53(5):489–506.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Fulmali, P. V., Walimbe, A., and Mahadev, P. V. M. (2008). Spread, establishment & prevalence of dengue vector Aedes aegypti (L.) in Konkan region, Maharashtra, India. *The Indian journal of medical research*, 127(6):589–601.

Gubler, D. J. (1998). Dengue and Dengue Hemorrhagic Fever. *Clin. Microbiol. Rev.*, 11(3):480–496.

Guzman, M. G., Halstead, S. B., Artsob, H., Buchy, P., Farrar, J., Gubler, D. J., Hunsperger, E., Kroeger, A., Margolis, H. S., Martínez, E., Nathan, M. B., Pelegrino, J. L., Simmons, C., Yoksan, S., and Peeling, R. W. (2010). Dengue: a continuing global threat. *Nature reviews. Microbiology*, 8(12 Suppl):S7—-16.

Heukelbach, J., de Oliveira, F. A., Kerr-Pontes, L. R., and Feldmeier, H. (2001). Risk factors associated with an outbreak of dengue fever in a favela in Fortaleza, northeast Brazil. *Tropical medicine & international health : TM & IH*, 6(8):635–642.

Hii, Y. L., Rocklöv, J., Ng, N., Tang, C. S., Pang, F. Y., and Sauerborn, R. (2009). Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Global health action*, 2.

Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., and Rocklöv, J. (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS neglected tropical diseases*, 6(11):e1908.

Jayasooriyaa, G., Senaratnea, S. M. L., Wijesingheb, W., Kusumawathiea, P. H. D., and Gunatilakec, J. (2009). Use of geographical information system (GIS) and global positioning system (GPS) for dengue and dengue haemorrhagic fever control in Sri Lanka. *Dengue Bulletin*, 33:11–20.

Johansson, M. A., Dominici, F., and Glass, G. E. (2009). Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS neglected tropical diseases*, 3(2):e382.

Kanakaratne, N., Messer, W. B., Tissera, H. A., Shahani, A., Abeysinghe, N., de Silva, A. M., and Gunasekera, M. (2009). Severe Dengue Epidemics in Sri Lanka, 2003 - 2006. *Emerging Infectious Diseases*, 15(2):192–199.

Kannathasan, S., Suthakar, K., Jude, P. J., and Surendran, S. N. (2013). Potential risk areas for dengue in the Jaffna Municipal Area in Northern Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*, 41(2):165–168.

Karim, M. N., Munshi, S. U., Anwar, N., and Alam, M. S. (2012). Climatic factors influencing dengue cases in Dhaka city: a model for dengue prediction. *The Indian journal of medical research*, 136(1):32–9.

Kholedi, A. A. N., Balubaid, O., Milaat, W., Kabbash, I. A., and Ibrahim, A. (2012). Factors associated with the spread of dengue fever in Jeddah Governorate, Saudi Arabia. *Eastern Mediterranean health journal*, 18(1):15–23.

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

Khormi, H. M. and Kumar, L. (2012). Assessing the risk for dengue fever based on socioeconomic and environmental variables in a geographical information system environment. *Geospatial health*, 6(2):171–176.

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.

Kyle, J. L. and Harris, E. (2008). Global spread and persistence of dengue. *Annual review of microbiology*, 62:71–92.

Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., and Scott, T. W. (2011). Impact of daily temperature fluctuations on dengue virus transmission by Aedes aegypti. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18):7460–5.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2006). *Remote Sensing and Image Interpretation*. John Wiley & Sons, 6th editio edition.

Lu, D. and Weng, Q. (2009). Extraction of urban impervious surfaces from an IKONOS image. *International Journal of Remote Sensing*, 30(5):1297–1311.

Messer, W. B., Vitarana, U. T., Sivananthan, K., Elvtigala, J., Preethimala, L. D., Ramesh, R., Withana, N., Gubler, D. J., and De Silva, A. M. (2002). Epidemiology of dengue in Sri Lanka before and after the emergence of epidemic dengue hemorrhagic fever. *The American journal of tropical medicine and hygiene*, 66(6):765–773.

Moran, E. F. (2010). Land Cover Classification in a Complex Urban-Rural Landscape with Quickbird Imagery. *Photogrammetric engineering and remote sensing*, 76(10):1159–1168.

Morin, C. W., Comrie, A. C., and Ernst, K. (2013). Climate and dengue transmission: evidence and implications. *Environmental health perspectives*, 121(11-12):1264–72.

Murray, N. E. A., Quam, M. B., and Wilder-Smith, A. (2013). Epidemiology of dengue: past, present and future prospects. *Clinical epidemiology*, 5:299–309.

Naish, S., Dale, P., Mackenzie, J. S., McBride, J., Mengersen, K., and Tong, S. (2014). Spatial and temporal patterns of locally-acquired dengue transmission in northern Queensland, Australia, 1993-2012. *PLoS one*, 9(4):e92524.

Nisha, V., Gad, S. S., Selvapandian, D., Suganya, V., Rajagopal, V., Suganti, P., Balraj, V., and Devasundaram, J. (2005). Geographical information system (GIS) in investigation of an outbreak. *The Journal of communicable diseases*, 37(1):39–43.

Ooi, E.-E., Goh, K.-T., and Gubler, D. J. (2006). Dengue prevention and 35 years of vector control in Singapore. *Emerging infectious diseases*, 12(6):887–893.

Pathirana, S., Kawabata, M., and Goonatilake, R. (2009). Study of potential risk of dengue disease outbreak in Sri Lanka using GIS and statistical modelling. *Journal of Rural and Tropical Public Health*, 8:8–17.

Pham, H. V., Doan, H. T. M., Phan, T. T. T., and Minh, N. N. T. (2011). Ecological factors associated with dengue fever in a Central Highlands province, Vietnam. *BMC infectious diseases*, 11(1):172.

Quintero, J., Brochero, H., Manrique-Saide, P., Barrera-Pérez, M., Basso, C., Romero, S., Caprara, A., De Lima Cunha, J. C., Beltrán-Ayala, E., Mitchell-Foster, K., Kroeger, A., Sommerfeld, J., and Petzold, M. (2014). Ecological, biological and social dimensions of dengue vector breeding in five urban settings of Latin America: a multi-country study. *BMC infectious diseases*, 14:38.

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):e1648.

Raheel, U., Faheem, M., Riaz, M. N., Kanwal, N., Javed, F., Zaidi, N., and Qadri, I. (2011). Dengue fever in the Indian Subcontinent: an overview. *Journal of infection in developing countries*, 5(4):239–247.

Ramasamy, R., Surendran, S. N., Jude, P. J., Dharshini, S., and Vinobaba, M. (2011). Larval development of Aedes aegypti and Aedes albopictus in peri-urban brackish water and its implications for transmission of arboviral diseases. *PLoS neglected tropical diseases*, 5(11):e1369.

Repaka, S. R., Truax, D. D., State, M., Kolstad, E., and Manager, C. (2004). COMPARING SPECTRAL AND OBJECT BASED APPROACHES FOR CLASSIFICATION AND TRANSPORTATION FEATURE EXTRACTION FROM. (May).

Rodenhuis-Zybert, I. A., van der Schaar, H. M., da Silva Voorham, J. M., van der Ende-Metselaar, H., Lei, H.-Y., Wilschut, J., and Smit, J. M. (2010). Immature dengue virus: a veiled pathogen? *PLoS pathogens*, 6(1):e1000718.

Rohani, A., Wong, Y. C., Zamre, I., Lee, H. L., and Zurainee, M. N. (2009). The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in Aedes aegypti (L.). *The Southeast Asian journal of tropical medicine and public health*, 40(5):942–50.

Sarfraz, M. S., Tripathi, N. K., Tipdecho, T., Thongbu, T., Kerdthong, P., and Souris, M. (2012). Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. *BMC public health*, 12(1):853.

Schmidt, W.-P., Suzuki, M., Thiem, V. D., White, R. G., Tsuzuki, A., Yoshida, L.-M., Yanai, H., Haque, U., Tho, L. H., Anh, D. D., and Ariyoshi, K. (2011). Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. *PLoS medicine*, 8(8):e1001082.

Sirisena, P. and Noordeen, F. (2014). Evolution of dengue in Sri Lanka-changes in the virus, vector, and climate. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 19:6–12.

Surendran, S. N., Kajatheepan, A., Sanjeefkumar, K. F. A., and Jude, P. J. (2007a). Seasonality and insecticide susceptibility of dengue vectors: an ovitrap based survey in a residential area of northern Sri Lanka. *The Southeast Asian journal of tropical medicine and public health*, 38(2):276–82.

Surendran, S. N., Kannathasan, S., Kajatheepan, A., and Justin Jude, P. (2007b). Chikungunya-type fever outbreak: some aspects related to this new epidemic in Jaffna district, northern Sri Lanka. *Tropical Medicine and Health*, 35(3):249–252.

Troyo, A., Fuller, D. O., Calderón-Arguedas, O., Solano, M. E., and Beier, J. C.

(2009). Urban structure and dengue fever in Puntarenas, Costa Rica. *Singapore journal of tropical geography*, 30(2):265–282.

Van Benthem, B. H. B., Vanwambeke, S. O., Khantikul, N., Burghoorn-Maas, C., Panart, K., Oskam, L., Lambin, E. F., and Somboon, P. (2005). Spatial patterns of and risk factors for seropositivity for dengue infection. *The American journal of tropical medicine and hygiene*, 72(2):201–208.

Vanwambeke, S. O., Lambin, E. F., Eichhorn, M. P., Flasse, S. P., Harbach, R. E., Oskam, L., Somboon, P., van Beers, S., van Benthem, B. H. B., Walton, C., and Butlin, R. K. (2007). Impact of Land-use Change on Dengue and Malaria in Northern Thailand. *EcoHealth*, 4(1):37–51.

Vanwambeke, S. O., van Benthem, B. H. B., Khantikul, N., Burghoorn-Maas, C., Panart, K., Oskam, L., Lambin, E. F., and Somboon, P. (2006). Multi-level analyses of spatial and temporal determinants for dengue infection. *International journal of health geographics*, 5:5.

Walker, K. R., Joy, T. K., Ellers-Kirk, C., and Ramberg, F. B. (2011). Human and environmental factors affecting Aedes aegypti distribution in an arid urban environment. *Journal of the American Mosquito Control Association*, 27(2):135–141.

WHO (2009). Dengue: Guidelines for Diagnosis,Treatment, Prevention and Control. Technical report.

WHO (2012). WHO | Dengue vaccine research.

WHO (2014a). WHO | Dengue and severe dengue.

WHO (2014b). WHO | Impact of Dengue.

WHO-SEARO and WHO-WPRO (2008). The Dengue Strategic Paln for the Asia Pacific Region 2008-2015. Technical report.

Wilder-Smith, A., Renhorn, K.-E., Tissera, H., Abu Bakar, S., Alphey, L., Kittayapong, P., Lindsay, S., Logan, J., Hatz, C., Reiter, P., Rocklöv, J., Byass, P., Louis, V. R., Tozan, Y., Massad, E., Tenorio, A., Lagneau, C., L'Ambert, G., Brooks, D., Wegerdt, J., and Gubler, D. (2012). DengueTools: innovative tools and strategies for the surveillance and control of dengue. *Global health action*, 5.

Zhou, W., Huang, G., Troy, A., and Cadenasso, M. L. (2009). Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. *Remote Sensing of Environment*, 113(8):1769–1777.

APPENDIX A

IRB APPROVAL FOR RESEARCH PROJECT

The ethical approval was obtained from both University of Alabama at Birmingham and University of Kelaniya, Sri Lanka

OMB No. 0990-0263
Approved for use through 3/31/2015

Protection of Human Subjects
Assurance Identification/IRB Certification/Declaration of Exemption
(Common Rule)

Policy: Research activities involving human subjects may not be conducted or supported by the Departments and Agencies adopting the Common Rule (56FR28003, June 18, 1991) unless the activities are exempt from or approved in accordance with the Common Rule. See section 101(b) of the Common Rule for exemptions. Institutions submitting applications or proposals for support must submit certification of appropriate Institutional Review Board (IRB) review and approval to the Department or Agency in accordance with the Common Rule.

Institutions must have an assurance of compliance that applies to the research to be conducted and should submit certification of IRB review and approval with each application or proposal unless otherwise advised by the Department or Agency.

1. Request Type	2. Type of Mechanism	3. Name of Federal Department or Agency and, if known, Application or Proposal Identification No.
[ ] ORIGINAL [X] CONTINUATION [ ] EXEMPTION	[ ] GRANT [X] CONTRACT [ ] FELLOWSHIP [ ] COOPERATIVE AGREEMENT [ ] OTHER:____________________	

4. Title of Application or Activity	5. Name of Principal Investigator, Program Director, Fellow, or Other
Integration of Remote Sensing and GIS Technologies for Dengue Surveillance in Sri Lanka [UAB-International Training and Research in Environmental and Occupational Health (ITREOH)]	TIPRE, MEGHAN

6. Assurance Status of this Project *(Respond to one of the following)*

[X] This Assurance, on file with Department of Health and Human Services, covers this activity:
Assurance Identification No. FWA00005960 , the expiration date 01/24/2017 IRB Registration No. IRB00000726

[ ] This Assurance, on file with *(agency/dept)*____________________, covers this activity.
Assurance No.__________, the expiration date__________ IRB Registration/Identification No.__________*(if applicable)*

[ ] No assurance has been filed for this institution. This institution declares that it will provide an Assurance and Certification of IRB review and approval upon request.

[ ] Exemption Status: Human subjects are involved, but this activity qualifies for exemption under Section 101(b), paragraph__________.

7. Certification of IRB Review (Respond to one of the following IF you have an Assurance on file)

[X] This activity has been reviewed and approved by the IRB in accordance with the Common Rule and any other governing regulations.
by: [ ] Full IRB Review on (date of IRB meeting) __________ or [X] Expedited Review on (date) 3-28-14
[ ] If less than one year approval, provide expiration date __________

[ ] This activity contains multiple projects, some of which have not been reviewed. The IRB has granted approval on condition that all projects covered by the Common Rule will be reviewed and approved before they are initiated and that appropriate further certification will be submitted.

8. Comments
Protocol subject to Annual continuing review.
HIPAA Waiver Approved?: Yes

Title X120326012
Integration of Remote Sensing and GIS Technologies for Dengue Surveillance in Sri Lanka [UAB-International Training and Research in Environmental and Occupational Health (ITREOH)]

IRB Approval Issued: 3-28-14

IRB Approval No Longer Valid On: 3-28-15

9. The official signing below certifies that the information provided above is correct and that, as required, future reviews will be performed until study closure and certification will be provided.

10. Name and Address of Institution
University of Alabama at Birmingham
701 20th Street South
Birmingham, AL 35294

11. Phone No. *(with area code)* (205) 934-3789

12. Fax No. *(with area code)* (205) 934-1301

13. Email: irb@uab.edu

14. Name of Official
Marilyn Doss, M.A.

15. Title
Vice Chair, IRB

16. Signature Marilyn Doss

17. Date 3-28-14

Authorized for local Reproduction

Sponsored by HHS

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 0990-0263. The time required to complete this information collection is estimated to average 30 minutes per response. If you have comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Health & Human Services, OS/OCIO/PRA, 200 Independence Ave., S.W., Suite 336-E, Washington D.C. 20201, Attention: PRA Reports Clearance Officer.

APPENDIX B

SENSITIVITY ANALYSIS

Results of the analyses using data from CMC meteorological departments and alternate data sources for temperature and rainfall, were compared. There were no significant difference between the results. Table 21-26 display these results.

Table 21

Comparison between data from CMC meteorological department and Land Data Assimilation System (LDAS) for temperature, and Tropical Rainfall Measuring Mission (TRMM) data for rainfall for years 2005-2011

Variable	Mean	Median	Min	Max	Range
Weekly Mean Temperature (CMC)	28.2	27.9	25.7	33.2	7.5
Weekly Mean Temperature (LDAS)	25.8	25.6	23.9	28.2	4.3
Rainfall (mm, CMC)	0.8	4.0	0.0	69.0	69.0
Rainfall (mm, LDAS)	7.3	5.7	0.0	41.8	41.8
Rainfall (mm, TRMM)	6.98	4.11	0.0	57.21	57.21

Table 22

Comparison between data from CMC meteorological department and Land Data Assimilation System (LDAS) for temperature, and Tropical Rainfall Measuring Mission (TRMM) data for rainfall for years 2005-2010

Variable	Mean	Median	Min	Max	Range
Weekly Mean Temperature (CMC)	27.8	27.8	25.7	29.9	4.2
Weekly Mean Temperature (LDAS)	25.7	25.5	24.2	27.9	3.7
Rainfall (mm, CMC)	7.2	4	0	69	69
Rainfall (mm, LDAS)	7.7	6.2	0	41.8	41.8
Rainfall (mm, TRMM)	0.2	4.3	0.0	57.2	57.2

Table 23

Comparison between data from CMC meteorological department and Land Data Assimilation System (LDAS) for temperature, and Tropical Rainfall Measuring Mission (TRMM) data for rainfall for year 2011

Variable	Mean	Median	Min	Max	Range
Weekly Mean Temperature (CMC)	30.7	30.7	27.0	33.2	6.2
Weekly Mean Temperature (LDAS)	26.7	26.9	23.9	28.2	4.3
Rainfall (mm, CMC)	4.9	2.5	0.0	22.6	22.6
Rainfall (mm, LDAS)	5.0	4.2	0.3	19.2	18.9
Rainfall (mm, TRMM)	5.9	3.6	0.0	29.4	29.4

Table 23

Correlation between weekly counts of dengue case and weekly average temperature (CMC), and weekly average temperature (LDAS) at lag period 0-25

	Temperature (CMC)		Temperature LDAS	
Lag period	Correlation coefficient	P-value	Correlation coefficient	P-value
Lag 0	0.45	<.0001	0.16	0.0029
Lag 1	0.45	<.0001	0.19	0.0003
Lag 2	0.47	<.0001	0.20	0.0002
Lag 3	0.47	<.0001	0.22	<.0001
Lag 4	0.47	<.0001	0.24	<.0001
Lag 5	0.48	<.0001	0.26	<.0001
Lag 6	0.49	<.0001	0.28	<.0001
Lag 7	0.49	<.0001	0.31	<.0001
Lag 8	0.51	<.0001	0.33	<.0001
Lag 9	0.53	<.0001	0.35	<.0001
Lag 10	0.51	<.0001	0.33	<.0001
Lag 11	0.52	<.0001	0.33	<.0001
Lag 12	0.52	<.0001	0.31	<.0001
Lag 13	0.50	<.0001	0.29	<.0001
Lag 14	0.51	<.0001	0.27	<.0001
Lag 15	0.49	<.0001	0.23	<.0001
Lag 16	0.45	<.0001	0.19	0.0005
Lag 17	0.43	<.0001	0.14	0.0114
Lag 18	0.41	<.0001	0.09	0.097
Lag 19	0.36	<.0001	0.05	0.3762
Lag 20	0.32	<.0001	-0.01	0.8429
Lag 21	0.21	<.0001	-0.04	0.4952
Lag 22	0.21	<.0001	-0.08	0.1556
Lag 23	0.16	0.0033	-0.09	0.0944
Lag 24	0.16	0.0033	-0.11	0.0475
Lag 25	0.10	0.0635	-0.14	0.011

Table 24

Correlation between weekly counts of dengue case and weekly average rainfall (CMC), and weekly average rainfall (TRMM) at lag period 0-25

	Rainfall (CMC)		Rainfall (TRMM)	
Lag period	Correlation coefficient	P-value	Correlation coefficient	P-value
Lag 0	-0.08	0.12	-0.06	0.28
Lag 1	-0.05	0.38	-0.03	0.61
Lag 2	-0.02	0.78	0.01	0.85
Lag 3	0.04	0.49	0.07	0.19
Lag 4	0.07	0.20	0.10	0.05
Lag 5	0.12	0.02	0.13	0.01
Lag 6	0.16	0.00	0.18	0.00
Lag 7	0.19	0.00	0.20	0.00
Lag 8	0.18	0.00	0.18	0.00
Lag 9	0.14	0.01	0.15	0.00
Lag 10	0.14	0.01	0.13	0.02
Lag 11	0.13	0.01	0.11	0.03
Lag 12	0.08	0.15	0.06	0.26
Lag 13	0.06	0.25	0.05	0.32
Lag 14	0.01	0.80	0.02	0.72
Lag 15	-0.03	0.57	-0.04	0.43
Lag 16	-0.06	0.30	-0.06	0.26
Lag 17	-0.09	0.10	-0.09	0.10
Lag 18	-0.14	0.01	-0.13	0.02
Lag 19	-0.14	0.01	-0.15	0.01
Lag 20	-0.14	0.01	-0.13	0.01
Lag 21	-0.15	0.01	-0.11	0.04
Lag 22	-0.13	0.01	-0.12	0.02
Lag 23	-0.11	0.04	-0.10	0.07
Lag 24	-0.14	0.01	-0.12	0.03
Lag 25	-0.07	0.22	-0.07	0.18

Table 25

Comparison between adjusted Relative risk (RR) of dengue incidence using data from CMC meteorological department and LDAS, for weekly average temperature (°C) at various lag periods, when rainfall is included in the model

Lag period	Adjusted RR¥	95% CI		Adjusted RR¥	95% CI	
Lag 1§	1.09	1.05	1.13	1.05	1.01	1.10
Lag 2	1.14	1.10	1.18	1.05	1.01	1.09
Lag 3	1.14	1.11	1.18	1.06	1.02	1.11
Lag 5	1.16	1.13	1.20	1.12	1.08	1.17
Lag 7	1.21	1.17	1.25	1.20	1.15	1.25
Lag 8	1.33	1.29	1.38	1.21	1.16	1.25
Lag 9	1.23	1.19	1.26	1.18	1.13	1.23
Lag 10	1.16	1.13	1.19	1.12	1.08	1.17
Lag 15	1.13	1.10	1.16	0.95	0.91	0.98

§ Weekly average temperature with lag period in weeks. GENMOD, DIST=POISSON, LINK=LOG model. Model includes total dengue count (dependent variable), temperature (independent), dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent) and rainfall at lag period 8th week.
¥Risk ratios and 95% confidence intervals for an increase of approximately 1°C.

Table 26

Comparison between adjusted relative risk (RR) of dengue incidence using data from CMC meteorological department and TRMM, for weekly average rainfall (mm) at various lag periods, when temperature is included in the model

Lag period	Adjusted RR[¥]	95% CI		Adjusted RR[¥]	95% CI	
Lag 1	1.01	0.99	1.02	1.01	0.99	1.03
Lag 2	1.01	0.99	1.02	1.02	1.01	1.04
Lag 3	1.03	1.02	1.04	1.05	1.03	1.06
Lag 5	1.06	1.05	1.07	1.07	1.05	1.08
Lag 7	1.07	1.06	1.09	1.10	1.08	1.11
Lag 10	1.06	1.05	1.07	1.05	1.03	1.06
Lag 12	1.04	1.03	1.06	1.02	1.01	1.04

[§] Weekly average rainfall with lag period in weeks. GENMOD, DIST=POISSON, LINK=LOG model. Model includes total dengue count (dependent variable), rainfall (independent), dummy variable for seasons (season 3 is referent) and year of dengue case occurrence (year 2005 is the referent) and temperature at lag period 11^{th} week.
[¥]Risk ratios and 95% confidence intervals for an increase of approximately 1°C.