An Efficient Approach for Dengue Mitigation
A Computational Framework

Nirosha Sumanasinghe Dinayadura

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS
December 2018

APPROVED:
Armin R. Mikler, Major Professor
Chetan Tiwari, Committee Member
Ranee Bryce, Committee Member
Song Fu, Committee Member
Barrett Bryant, Chair of the Department of Computer Science and Engineering
Yan Huang, Interim Dean of the College of Engineering
Victor Prybutok, Dean of the Toulous Graduate School

COPYRIGHT NOTICE

ACKNOWLEDGEMENTS

My first debt of gratitude must go to my advisor, Dr. Armin R. Mikler. He patiently provided the vision, encouragement and advice necessary for me to move forward through the doctoral program and complete my dissertation. He has been a strong and supportive adviser to me throughout my graduate school career. Completing my PhD degree is the most challenging activity of life. It has been a great opportunity to spend several years in the Department of Computer Science and Engineering, University of North Texas, and I am always glad to have a good relationship with the members of the department.

Special thanks to my committee, Dr. Chetan Tiwari, Dr. Song Fu and Dr. Renee Bryce for their support, guidance and helpful suggestions. Their guidance has served me well and I owe them my heartfelt appreciation. Members of CERL Lab also deserve my sincerest thanks, their friendship and assistance has meant more to me than I could ever express.

TABLE OF CONTENTS

Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Existence of dengue in Sri Lanka links back to the early 1960s and yet dengue has become a major public health issue at present with a high morbidity and mortality. In 2009, dengue infections increased at an alarming rate across Sri Lanka. By the end of the year 2009, 35,095 people were infected, and 346 fatalities reported. The number of infections has never been lower than 28,000 since 2009. In 2013, 47,246 infections were reported. Number of fatalities was 83 in 2013. During the first 10 months of the year 2017, 15,8854 suspected dengue cases have been reported to the Epidemiology Unit of Sri Lanka keeping the mortality rate at an alarming level which was about 300 deaths. There is an urgent need for a comprehensive mitigation plan to manage the impact of the epidemic.

The Presidential Task Force initiated many efforts from Dengue Prevention, to fines for those who neglect breeding grounds, to declaring a national dengue eradication program. Despite these initiatives, the rate of infection is exactly what it was five years ago. Several areas reported a slight reduction in dengue cases after deploying the suggested strategies. The situation is more urgent and alarming in the Western Province, home to over 25 percent of the country’s population of over 20 million people, and to 60 percent of all reported dengue cases since 2009. Western Province (43%) has the highest number of dengue cases reported. The Colombo district is the most affected area.

Trending of dengue cases during recent years indicates that strategies, that have been tried, failed. Experts are now suggesting getting help from the national Meteorological Bureau for the fight against the virus. There is a strong relationship between climate pattern and the spread of dengue disease [1]. The authors further asserted that mosquito breeding grounds increased following heavy rains, pointing out that the two annual peaks in infections were recorded soon after the two annual monsoons. This work also found that warming weather patterns increased the distribution of the dengue-carrying mosquito. Researchers pointed out that detailed weather forecasts could help health authorities to better allocate resources and strategically implement prevention campaigns.

1.1 Global Burden of Dengue

World Health Organization reported that dengue incidence has grown exponentially around the world in recent years. Public awareness on dengue epidemics is low and that results in lower number of cases reported to government offices. It is estimated that 390 million dengue cases reported per year [53]. Another study conducted estimates that 3,900 million people living in 128 countries are at risk [54]. The global risk map of dengue epidemic is shown in Figure 1.1.

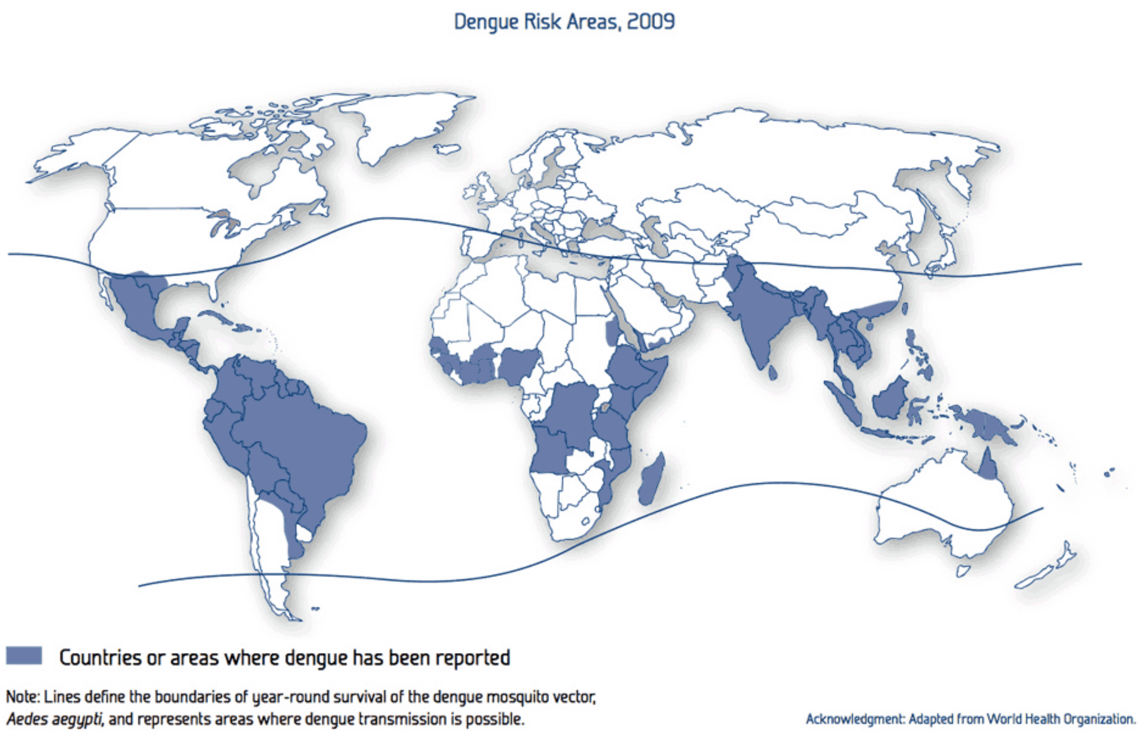


Figure 1.1 Dengue risk map for year 2009

1.2 Status and Trends of Dengue Disease

South Asian and Southeast Asian countries have been fighting dengue for decades. The status and the trends of all these countries are homogeneous. In recent years, Sri Lanka has been hit by several major dengue epidemics. A similar situation can be found in Thailand and it has also been hit by several dengue epidemics in recent years. The most recent outbreak took place in Thailand in 2018. The main goal of the proposed work is to build a general model for modelling of dengue epidemic. Therefore, I conducted statistical analysis of dengue disease in Sri Lanka and apply the findings in Thailand for the comparison.

1.2.1 Dengue Status of Sri Lanka

It is clearly visible in the global risk map of dengue epidemic that Sri Lanka is in the high-risk area. This is not a coincidence and it is clearly reflected in the reports produced by various institutions in Sri Lanka. In the last quarter of the year 2015, 14,776 dengue patients have been reported to the Epidemiology Unit. In the Western Province, that accounted for 47.11% of total reported cases. The recent development of dengue epidemic in Sri Lanka is alarming. During the first 10 months of the year 2017, 158,854 suspected dengue cases have been reported. The mortality rate was at an alarming level which was about 300 deaths. 37,988 dengue cases have been reported January through September in the current year 2018. Distribution of cases by weeks for the year 2017 is given in the Figure 1.2.

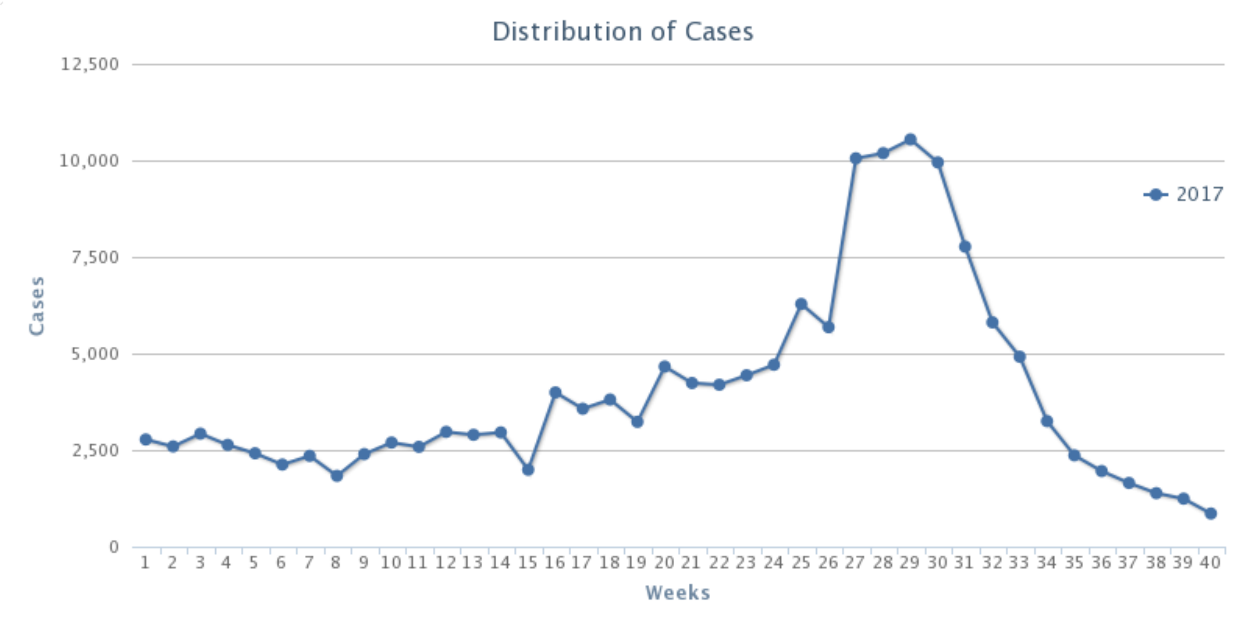


Figure 1.2 Dengue annual case rates reported weekly in year 2017

A committee has been appointed by the government of Sri Lanka to thoroughly study the dengue epidemic and provide recommendations towards the better control of dengue epidemic. The committee comprised of professionals from several fields including medicine, healthcare, environmental and higher research institutions. The committee was first appointed in 2001.The final report was produced in the year 2005 and handed over to the government of Sri Lanka. This report proposed several action plans which included very specific recommendations. Those are listed below,

1. To reduce morbidity and mortality due to DF/DHF.

2. To forecast and prevent dengue epidemics.

3. To strengthen liaison with civil society groups, NGO, media and other relevant stakeholders for social mobilization in dengue control.

4. To identify and mobilize resources to carry out research on dengue.

5. To develop and sustain an effective dengue prevention and control program in Sri Lanka.

Among them, great attention is paid for establishing a forecasting model for the dengue epidemic based on the rainfall, temperature, population density, and other specific factors. Until today, this item in the list of action plan has not been addressed. The effect of the delay of action plan is clearly shown in the dengue trend as shown in Figure 1.3 and Figure 1.4. Figure 1.3 reveals the strong relationship that exists between rainfall and reported dengue cases. The higher peaks occur after every monsoon season.

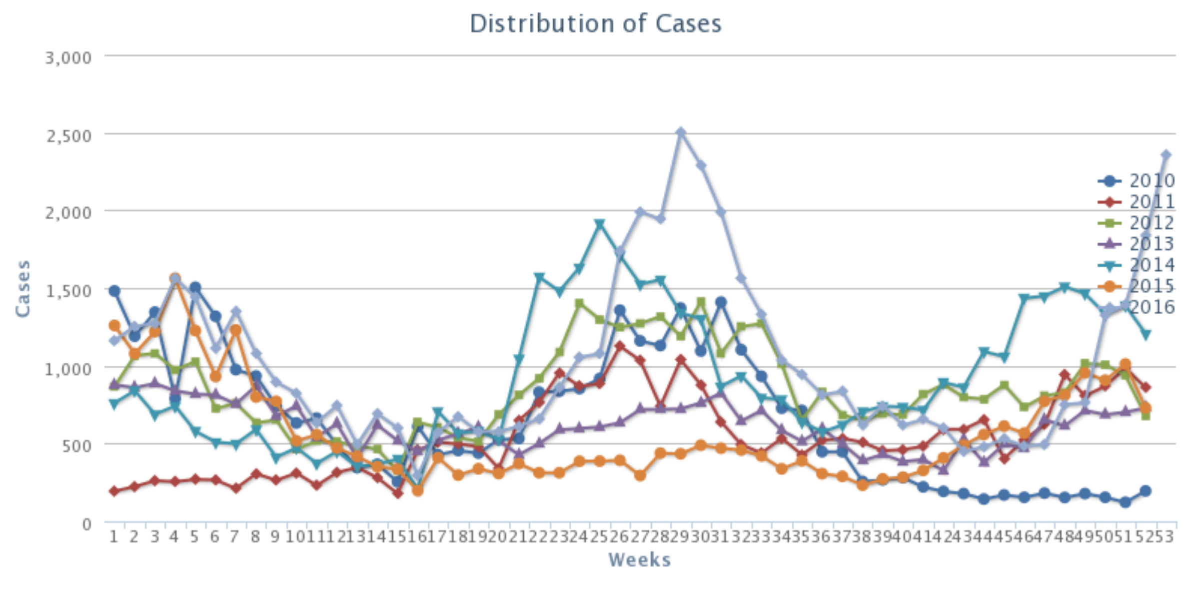


Figure 1.3 Dengue trend for years 2010-2016 in Sri Lanka

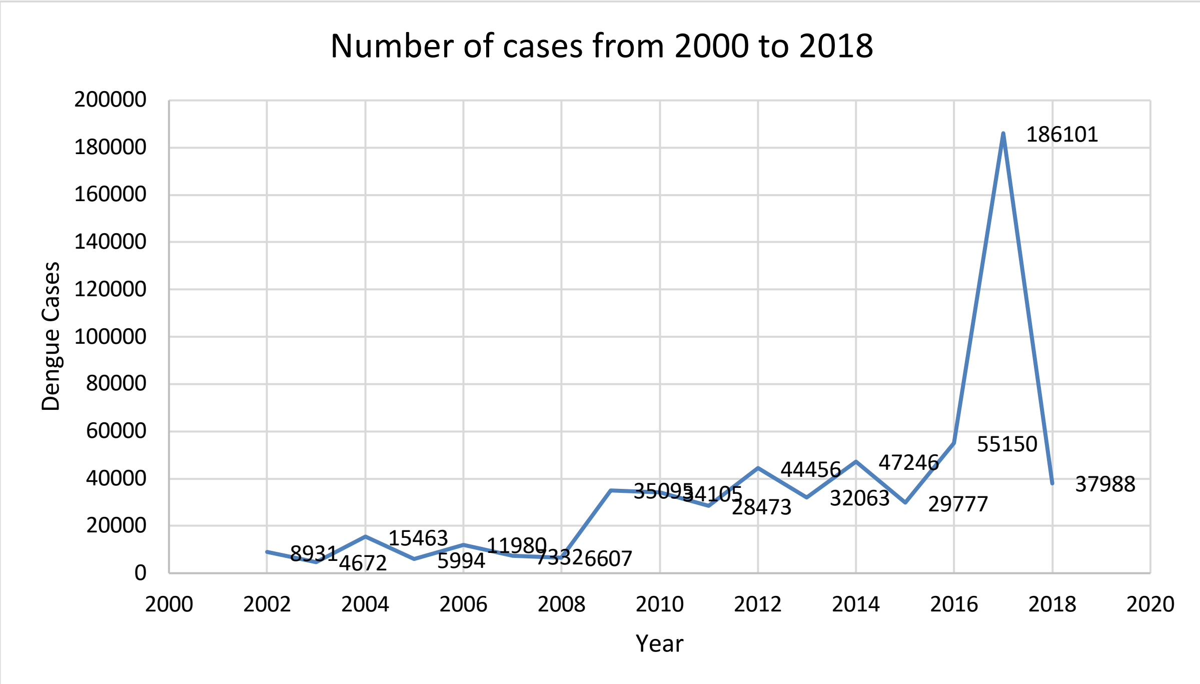


Figure 1.4 Dengue case trends from 2000 to 2018 in Sri Lanka

It is very crucial to design and develop a dengue epidemic forecasting model for dengue epidemic as the rate of increase of reported dengue cases in each year since 2005 is alarming.

1.2.2 Dengue Status of Thailand

Thailand reported 6,565 total dengue cases from 75 of the 76 provinces from January 2018 to April 2018. There were nine dengue fatalities. Out of 6,565 total cases, there were 3,878 dengue fever cases, 2,610 dengue hemorrhagic fever and 2 reported deaths and 77 dengue shock syndrome cases and 7 reported deaths. Figure 1.5 shows the dengue trend in Thailand from year 2000 to 2011. Figure 1.6 shows dengue cases by month from year 2000 to 2012. Data displayed on Figure 1.5 and Figure 1.6 are obtained from the survey study presented in [68].

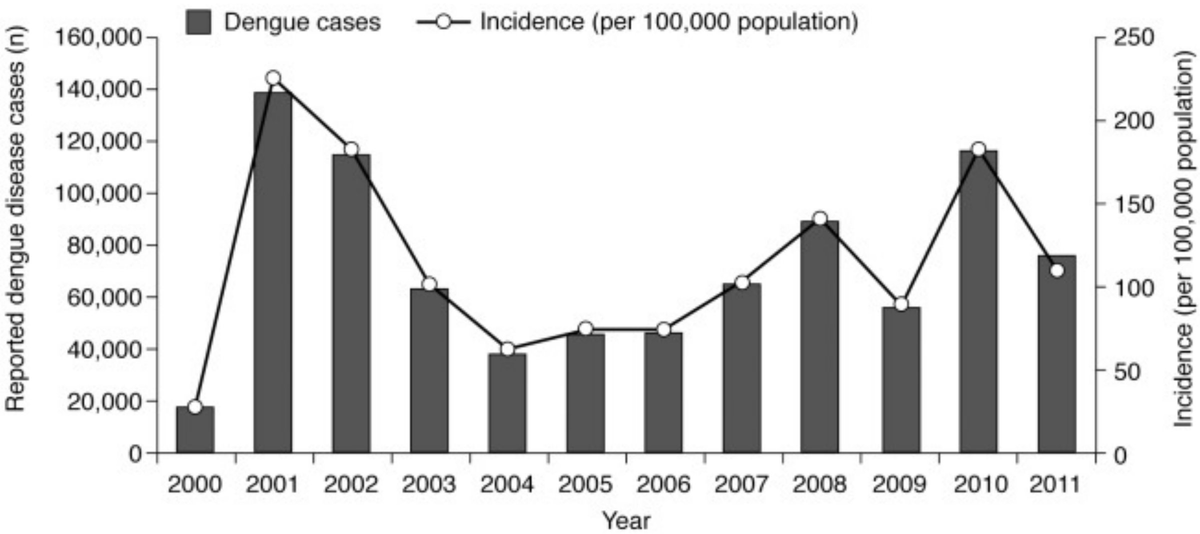


Figure 1.5 Number of reported dengue disease cases and dengue disease incidence, Thailand, 2000–2011

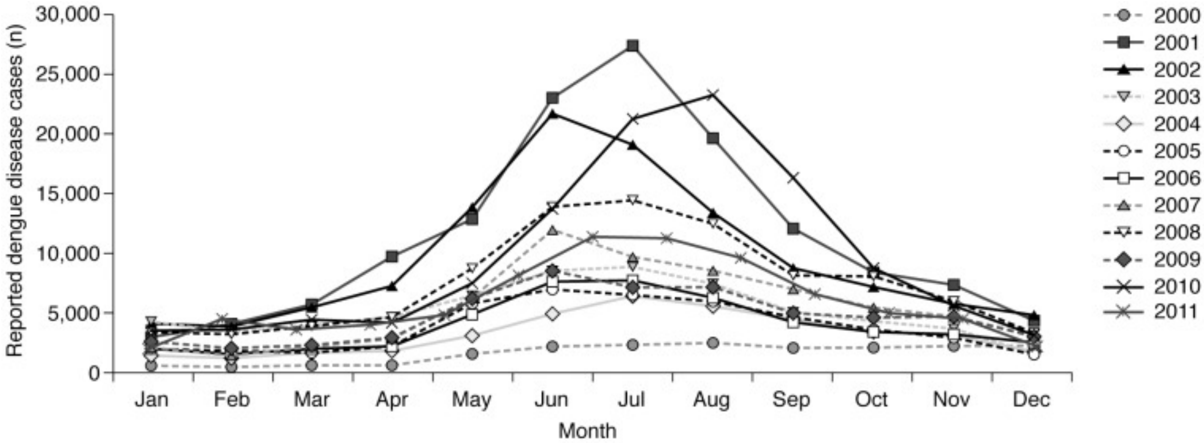


Figure 1.6 Number of reported cases of dengue disease, by month, Thailand, 2000–2012

1.3 Questions to be Addressed

1. What are the factors that affect the spread of dengue epidemic?

2. Are there any local factors that are more important when modeling epidemic and missed when considering only global factors?

3. What is the effectiveness of Support Vector Regression (SVR) and micro ensemble architecture in prediction when considering global and local factors with both vector and human population considered? What are the best parameter settings of the SVR?

4. Can we improve the result by feeding the SVR and ensemble with a combination of geographical, socio economic, and weather data?

Based on the information presented above, I propose a well formulated computational approach to predict and provide guidance in mitigating dengue epidemic in Sri Lanka and use the same modeling technique to address the dengue burden in Thailand. In the proposed model, I consider all the environmental factors as well as geographical factors that may affect dengue vector population and the reported dengue cases. Contributions of my research proposal are as given below.

1.4 Contribution

The contribution of the proposed work is three-fold and span into three major research areas. Those are epidemic analysis, epidemic prediction, and resource allocation. Contribution to each area is explained in detailed below.

1. To provide a comprehensive insight into dengue epidemic and spread of the vector population based on stimulating factors such as rainfall, temperature, and population density.

   The foremost task to be completed before building a framework to identify and mitigate dengue epidemic is to thoroughly analyze the epidemic. There may be various factors stimulating the spread of the dengue epidemic. These factors may contain directly related factors and hidden factors that are playing a major role in spread of the epidemic. Several major contributors of dengue epidemic have been identified by several research works conducted [1]. Among them, temperature, rainfall, and land use at the top of the list. All the research work conducted so far considered only the global factors and treat every part of the country homogenously. In reality, there is a great variation in climate, land elevation, population, temperature, and rainfall parameters among different regions in two countries. In this study, we consider global parameters and treat each region with different strategy to profoundly represent the state of the particular region throughout the country.

2. To predict the upcoming dengue epidemic and its severity using a support vector regression and micro ensemble architecture.

There are a handful of research projects conducted to find the major stimulating factor of the dengue epidemic. Despite the work that has been done, there is still a lack of dengue mitigation strategy implemented based on the research findings available in Sri Lanka and Thailand. The government of Sri Lanka is deploying numerous projects to mitigate the dengue epidemic. The lack of a method to measure the severity of the epidemic and the failure to identify the epicenter of the epidemic led to failure in the dengue mitigation effort. It is questionable to use global parameters to predict the epidemic acknowledging that each region is different from every other region in terms of population density, rainfall, temperature variation, land use, etc. For the identification of upcoming dengue epidemic, we propose a micro ensemble architecture in which each district is modeled with a small-scale ensemble. The result is obtained by applying a combining strategy on all the results obtained from each output of the ensemble.

3. To allocate limited resources efficiently to effectively mitigate the dengue epidemic

Resources are sparse especially in countries like Sri Lanka. It is important to utilize the available resources effectively. This poses a challenge of how to allocate limited resources among facilities which demand for resources. In major outbreaks such as the one happened in year 2017 where all the hospitals over filled with patients, posed a significant threat to patients due to lack of resources. All the resources were lacking including, a number of hospital beds during major outbreaks. Resource allocation plan is a major component in response planning. Hence, I proposed a resource allocation strategy based on a modified genetic algorithm (GA). The proposed resource allocation scheme can generate a resource allocation plan in less time compared to the standard GA. The proposed GA is producing a result closer to the optimum.

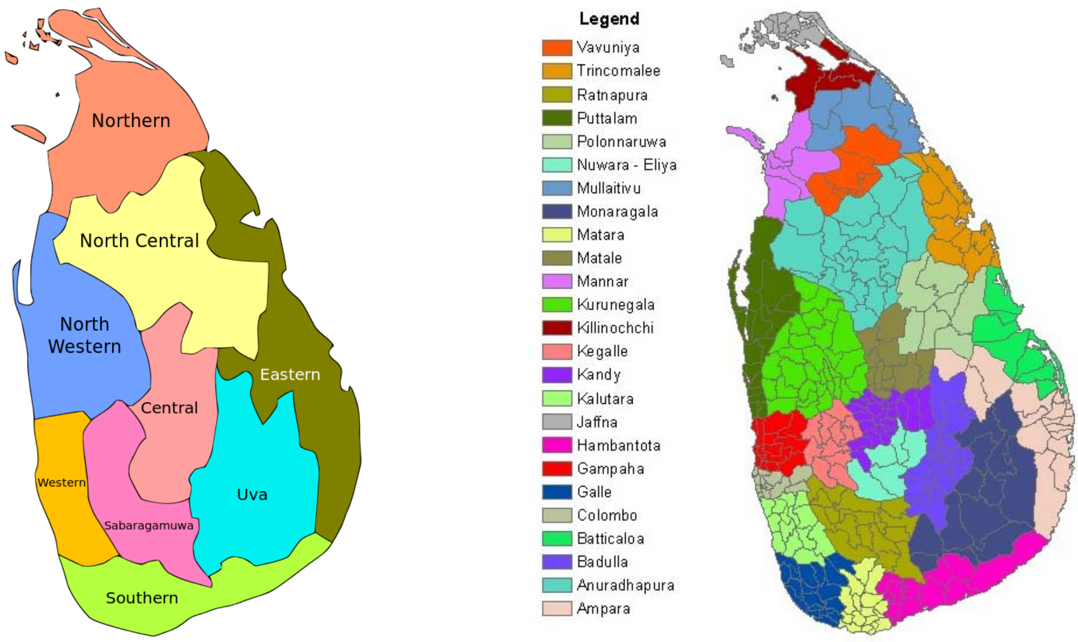
CHAPTER 2

BACKGROUND

The proposed study was evaluated on data sets obtained from Sri Lanka and Thailand. The geographical structure of both countries plays an important role in the variation of climate patterns. Climate pattern is directly related to the spread of the dengue virus. The geography of Sri Lanka and Thailand is described in the following sections.

2.1. The Geography of Sri Lanka

Sri Lanka is an island in the Indian Ocean and located in Southern Asia. It has 64,740 km^2 of land and 870 km^2 of water. Sri Lanka's climate is tropical. There are two main rainfall seasons which are the northeast monsoon (from December to March), and the southwest monsoon (from June to October). Majority of Sri Lanka’s land is flat and at sea level. The highest point is Pidurutalagala which is 2,524.13 m high. Sri Lanka is divided into 9 administrative regions (provinces). Each province is divided into several sub regions (districts) resulting in 25 districts. Administrative regions are shown in Figure 2.1.



(a) (b)

Figure 2.1(a) Provinces (b) Districts of Sri Lanka

2.2 Rainfall of Sri Lanka

The main sources of rainfall in Sri Lanka are Monsoonal, Convectional and expressional rain. The mean annual rainfall varies between 900mm to 5,000mm (Figure 2.2).

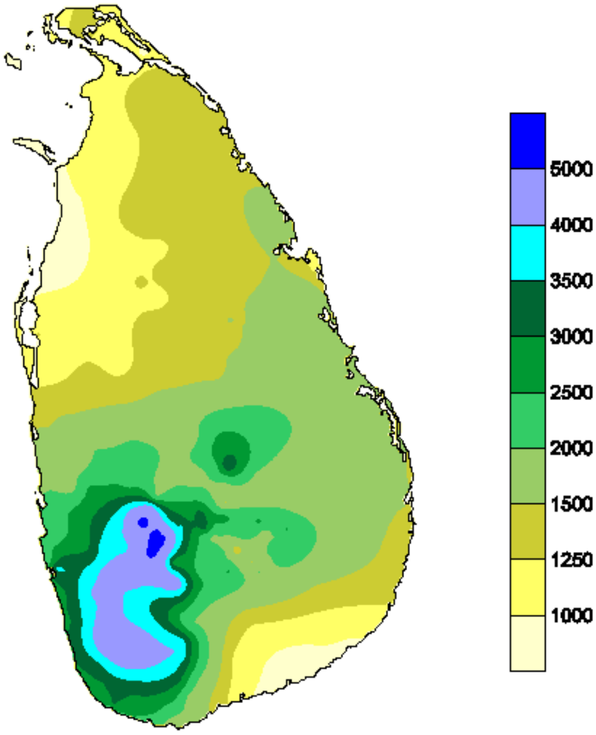


Figure 2.2 Annual rainfalls in Sri Lanka (Courtesy: Department of Meteorology Sri Lanka)

2.3 Temperature of Sri Lanka

Altitude is the main cause of regional differences observed in air temperature over Sri Lanka. The mean monthly temperatures slightly differ time to time based on the seasonal changes due to movement of the sun. The mean annual temperature in Sri Lanka is rapidly decreasing when moving towards highlands from low lands. At the altitude of 100 m to 150 m, the mean annual temperature is between 26.5 ^{0}C to 28.5 ^{0}C. The temperature falls rapidly as the altitude increases. The town Nuwara Eliya is at 1,800 m from sea level and its mean annual temperature is 15.9 0C. The coldest month is January, and April and August are the warmest.

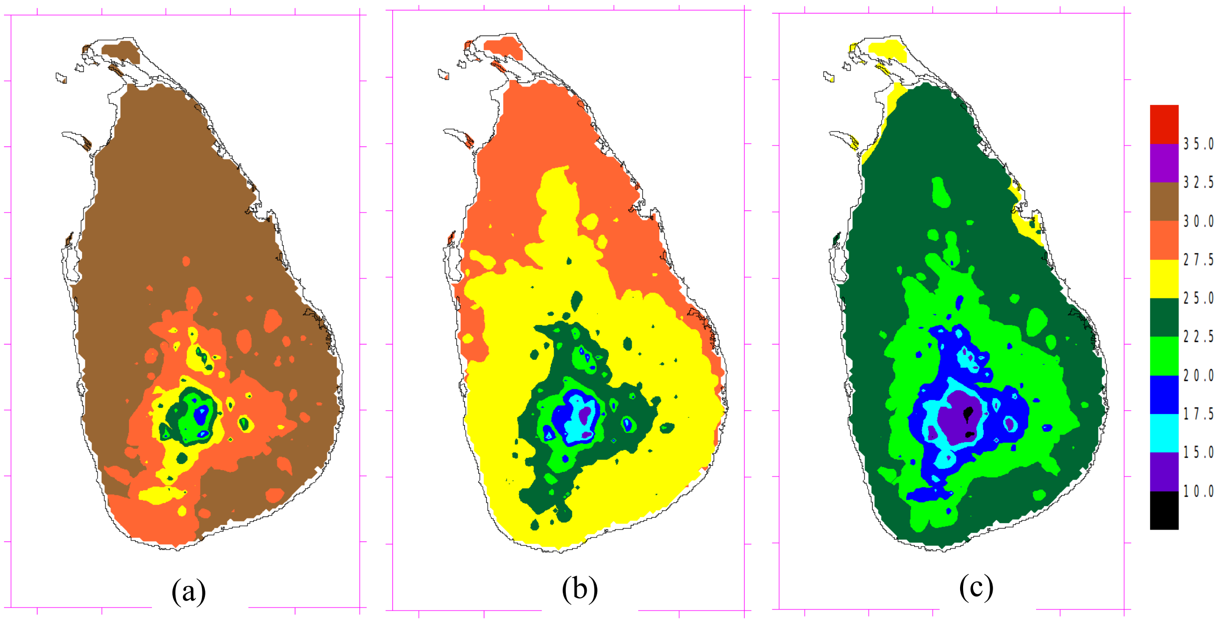


Figure 2.3 Average temperatures from 1961 to 2015 for (a) April (b) August and (c) January (Courtesy: Department of Meteorology Sri Lanka)

2.4 Climate Seasons of Sri Lanka

The Climate experienced for 12 months period in Sri Lanka can be characterized in to 4 climate seasons as follows.

First Inter-monsoon Season (March - April)

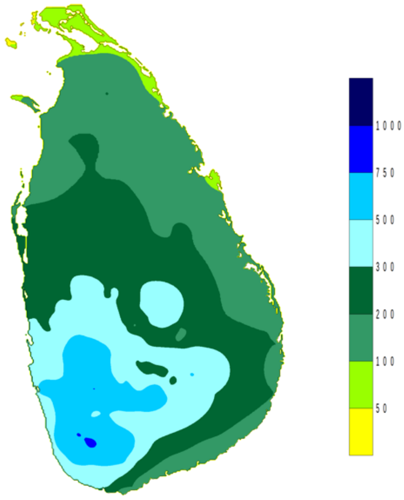


Figure 2.4 Distribution of rainfall in First Inter-monsoon Season (Courtesy: Department of Meteorology Sri Lanka)

Southwest -monsoon Season (May - September)

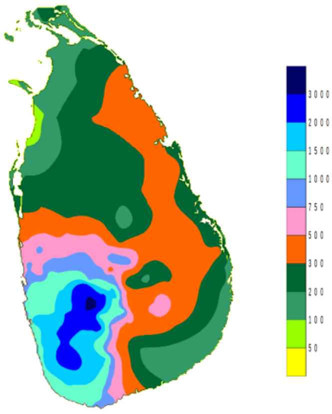


Figure 2.5 rainfall distributions for Southwest -monsoon Season (Courtesy: Department of Meteorology Sri Lanka)

Second Inter-monsoon Season (October-November)

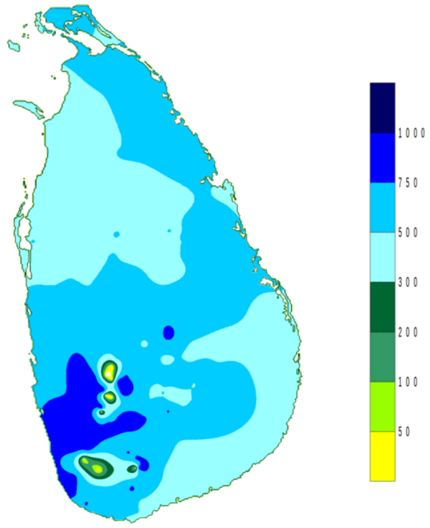


Figure 2.6 rainfall distributions for Second Inter-monsoon Season (Courtesy: Department of Meteorology Sri Lanka)

Northeast -monsoon Season (December - February)

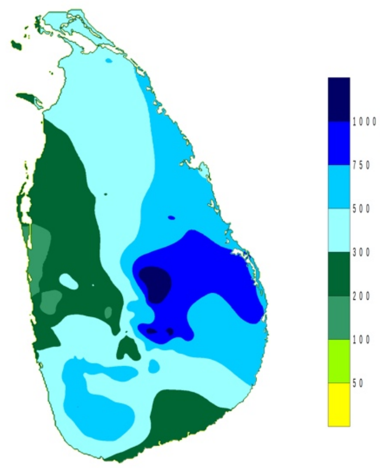


Figure 2.7 rainfall distributions for Northeast -monsoon Season (Courtesy: Department of Meteorology Sri Lanka)

2.5 The Geography of Thailand

Thailand is located in Southeast Asia. With a total area of 513,000 km2 (198,000 sq mi), Thailand is the world's 50th-largest country. Its population ranked the 20^{th} in the world, with 69 million individuals. Thailand is divided into 76 provinces as given in the Figure 2.8. The figure also shows the dengue incidence for the month May of year 2014, which are gathered into five groups of provinces by location. Bangkok (Krung Thep Maha Nakhon) and Pattaya are considered as two special districts. Bangkok is considered both a district and a province. The country of Thailand is given in the Figure 2.8.

Figure 2.8 Provinces of Thailand (Courtesy: Wikipedia)

2.6 The Climate of Thailand

There are three seasons of climate in Thailand. Those are Southwest monsoon season or rainy season (May - October), Northeast monsoon season or Winter season (October - February) and Pre-monsoon season or Summer (February – May).

The five divisions of Thailand are Northern, Northeastern, Central, Eastern and Southern Parts. The rainfall for each part is given in Table 2.1.

Table 2.1 Seasonal Rainfall (mm) in Thailand

Region	Winter	Summer	Rainy
North	100.4	187.3	943.2
Northeast	76.3	224.4	1,103.80
Central	127.3	205.4	942.5
East	178.4	277.3	1,433.20
South			
- East Coast	827.9	229	680
- West Coast	464.6	411.3	1,841.30

The temperature of Thailand for each part is given in Table 2.2.

Table 2.2 Temperature (Celsius Degree - °C) in Thailand

Temperature	Region	Winter	Summer	Rainy
Mean	North	23.4	28.1	27.3
	Northeast	24.2	28.6	27.6
	Central	26.2	29.7	28.2
	East	26.7	29.1	28.3
	South			
	- East Coast	26.3	28.2	27.8
	- West Coast	27	28.4	27.5

3.3 The Dengue Epidemic of Sri Lanka

Dengue virus is a mosquito-borne flavivirus. Dengue existed and has overwhelmed human population for a long period of time. Dengue transmission is supported by the urbanization. Human population growth is another reason for the quick transmission of dengue virus between continents. Specially in the tropical regions of the world, these conditions generated a favorable environment for successful Dengue transmission. *Aedes* *aegypti* and *Aedes albopictus* are the vectors that transmit DENV among humans. Currently, 48 Aedes species in 11 subgenera have been identified in Sri Lanka. Sri Lanka has been suffered by DF/ DHF epidemics for more than two decades. DF was officially confirmed in Sri Lanka in 1962. In 1966, presence of DF in all major cities in Sri Lanka was confirmed.

Sri Lankan Government targets its control efforts on the disease and vector control, social mobilization, clinical management of DF/DHF patients, and public awareness using media. A president task force on DF/DHF has been established to moderate the DF/DHF control activities. Training professionals to effectively address concerns and intern bringing the DF/DHF mortality to zero, or to a minimum level has also been taken place.

3.4 The Virus and the Vector

DENV is a flavivirus which is transmitted by *Aedes aegypti* mosquitoes. There are four distinct DENV serotypes, DENV 1–4 [70]. Infection with a single DENV serotype leads to long-term protective immunity against that serotype. The immunity obtained from one serotype will not protect from other serotypes. The geographical distribution of DENV in the world is an indication of the spread of dengue transmitting mosquitos that causes frequent outbreaks [71]. Mosquitoes (female) lay their eggs in water containers such as tires, cans, and in any object that collects water. Rising in the number of dengue cases after rainy seasons directly link to the water requirement of breading of dengue mosquito. The *Ae. aegypti* mosquito is adapted to urban environments. The *Ae. aegypti* is abundant in close proximity to humans and causing to have multiple host contacts within short period of time. The female mosquito (female mosquito bites humans and male mosquito does not depend on human blood) bites multiple hosts in order to complete a single meal. The

Aedes mosquitoes are active during the day and protective clothing is recommended where DENV is prominent. DENV infection is replicated in the mosquito midgut and disseminates and replicates in other tissues. Once the DENV infects the salivary glands of the mosquito, it transmits to the host in the next meal of the mosquito [72].

All four serotypes of DENV have been co-existed in Sri Lanka for more than three decades. Despite its long-existence, their distribution has not changed in the last 30 years. Studies found the existence of two or more DENV serotype in different parts of the country. There was an epidemic of DF associated with DENV serotypes 1 and 2 from 1965 to 1968. This island-wide epidemic caused 51 DHF cases and 15 deaths.

3.5 REPLAN Framework

RE-PLAN is a computational framework developed to create, analyze and optimize emergency response plans for public health emergencies. Specially, RE-PLAN facilitates the placement of PODs across the region of interest and establishes the geographic region that is being served by each POD. POD locations are selected to minimize the distance that the public has to travel to receive emergency services. Population distribution and geospatial data of the region are utilized for the purpose of response plan creation. Data pertaining to the infrastructure of the region, such as the road network, are utilized in analyzing the effectiveness of the resulting response plan. Specific methods have been developed as part of the RE-PLAN framework to enable creation, analysis and optimization of response plans for different scenarios [2].

A response plan developed in RE-PLAN consists of a set of PODs and their respective service areas. Each POD is a location in the region of interest defined by its geographic coordinates and attributes such as the number of service booths that the facility may accommodate. A service area of a POD is a portion of the region of interest that is serviced by the POD. Service areas consist of groups of contiguous population blocks, which are geographic entities such as represented by polygons with associated population counts. Population blocks, for instance, can be geographic entities such as census blocks or block groups used by the United States Census Bureau to represent populations. RE-PLAN facilitates response plan creation by either establishing the service areas for a set or subset of user-supplied PODs or by recommending a partitioning of the region into service areas and selecting available POD locations for each of the service areas. Methods which determine the service areas for a given set of POD locations are referred to as constrained methods. Unconstrained methods partition the region into service areas and map suitable PODs to these service areas.

The POD placement and determination of catchment areas of RE-PLAN can be used in the mitigation of dengue epidemics. The proposed study is capable of predicting the high-risk areas of the upcoming epidemic. The RE-PLAN framework will be used to place POD facilities in the most needed areas and determine the catchment area of the POD facility. The information generated, such as the population that needs to be served, the total area of the catchment area of the POD facility, from the POD placement can be used in the proposed resource allocation.

CHAPTER 3

RELATED WORK

3.1 Dengue Epidemic

There are multiple serotypes involved in increased disease transmission in Asia. The existence of multiple serotypes is also responsible for more frequent outbreaks [8]. There are 2.5 billion people around the world living in dengue-endemic countries with a risk of getting contact with DF/DHF and half of them live in 10 countries of the Southeast Asian region. The Southeast Asian region and the Western Pacific region account for 75% of the global dengue burden. In 2002, DF/DHF was ranked as the third most common noticeable disease in Sri Lanka (first and second were malaria and tuberculosis) [9, 10]. In recent years, deaths due to DF/DHF have been greater than those due to malaria, and DF/DHF is becoming the number one killer mosquito-borne infection in Sri Lanka [9]. At present, DF and DHF are prevalent in many urban and semi-urban areas of Sri Lanka with seasonal and periodic epidemics occurring regularly in the island [5]. In recent decades, a higher incidence of DF/DHF has been reported in the districts of Colombo, Gampaha, Kalutara, Kurunegala, Kegalle, Ratnapura, and Kandy [11]. The reported number of suspected and serologically positive DF/DHF cases, from the epidemic, occurred in 2004, varied from 4,749 to 15,643, involving 25–88 deaths [6]. Jaffna and Batticaloa from northern and eastern provinces respectively reported that DF/ DHF became endemic in those cities with a high dengue incidence [12].

There is a significant relationship between the dengue disease and the age group of the population. In many age groups, males have been affected. According to a regional study done by the World Health Organization (WHO) in Sri Lanka. The study was carried out based on reported cases from 1996 to 2005. The proportion of infections of DF/DHF among young male adult aged 15 years was significant. This male dominance was reported in every province of the country. Among those aged 1–4 and 5–14 years, there were significantly fewer male cases than expected, although there was some annual variation [13]. The highest incidence occurred in the 5–9 years age-group [3]. Children population was the main target of DF/DHF before 2000 and young adult was added to the risk list after the year 2000. An increase in mean age affected by DF/DHF was reported. The mean age was increased from 15 to 25 in 1996 to 2006 respectively [4].

Climate change such as temperature, rainfall, and humidity can expand the geographical range of vector mosquitoes. And also, it can extend the length of the disease transmission season. Ideal climate conditions will also reduce the time taken for the virus to get matured and develop into infective stages in mosquitoes. This might increase the propagation rates of diseases transmitted by *A. aegypti and A. albopictus*. In addition, these works clearly pointed out that there is a strong positive correlation between rainfall and the dengue cases reported [7, 14, 15, and 16]. Two DF/DHF peaks occur annually in Sri Lanka. These peaks are in association with the monsoon rains. During peak times, the densities of two mosquito vectors (*A. aegypti and A. albopictus*) are high. The first peak appears in June/July, along with the south-western monsoon that commences in late April. The second peak usually occurs in December and it is associated with the north-eastern monsoon rains (October to December) [3].

Temperature affects DF/DHF outbreaks in many different ways. *A. aegypti* has been shown to transmit DENV when the temperature is above 20 ^{0}C. The vector is inactive in temperatures below 16 ^{0}C. There is a positive correlation between temperature and the vector growth, especially female vector. And also, the feeding frequency is increased due to the fast reduction of food reserves in mosquitoes in high temperatures [17]. It is predicted that countries that have a mild temperature will have a rapid distribution of DENV due to global warming [18]. This brings serious concerns to temperate countries due to the distribution of *A. albopictus* [19]. Altitude is also a major factor that limits distribution of *A. aegypti*. Lower elevation (less than 500 m) is a favorable location for mosquitoes and hence moderate to heavy mosquito populations are expected. The higher elevation has a low population of mosquitos [20].

There is a couple of researches conducted to study the control measures of the dengue epidemic in Sri Lanka. These studies have revealed couple of strategies to be used in controlling the epidemic. A study conducted in the Kandy District of Sri Lanka showed that the mechanical and biological measures alone are not sufficient to prevent *Aedes* breeding. The prevention of *A. aegypti* and *A. albopictus* breeding in water storage containers would help to control DF/DHF. Therefore, DF/DHF control programs should pay more attention to the control of *Aedes* breeding in domestic water storage containers [5]. More importantly, public education on preventing dengue epidemic will reduce the mosquito breeding sites and hence will be very effective in the dengue mitigation [21, 22]. This will call for a system to integrate all these findings and develop a methodology as proposed in this study to effectively deal with the dengue epidemic in Sri Lanka. We propose to study and use rainfall, temperature, and population density in forecasting/predicting system to clearly identify the high-risk areas. And also, the proposed system is capable of estimating the casualties of the upcoming dengue epidemic and the government officials will be able to prepare for the epidemic. This information will help to educate the general public and to put controlling measures into action.

3.2 Forecasting /Prediction

In this section, several state-of-the-art prediction tools are categorized and presented.

3.2.1 GIS and Statistical Models

The authors of [23] and [24] studied the prediction of dengue outbreak in Sarawak and Johor respectively by using statistical models. The work presented in [23] analyzes the interaction between environmental, entomological, socio-demographic factors. GIS technology was used to generate geographic and environmental data on *Aedes albopictus* and dengue transmission. A total of 32,838 *Aedes albopictus* eggs were collected from trapping that spans 56 days. Cluster sampling was conducted to determine whether any of the risk factors (entomological or geographical) were influenced by geographical location. SPSS 10.1 was used on the data collected to perform the analysis. Descriptive analysis tools such as frequency, mean, and median were used. Two-sample t-test, and Pearson's Chi-Square were used to determine the association between independent variables and dengue cases reported. Use of differential Global Positioning System in mapping sites of 1m accuracy is also highlighted. Analysis of the data revealed there are major differences in clusters of villages. These differences include the number of *A. albopictus* eggs from ovitraps set indoor, outdoor and in dumping sites, container density, house density, and distance of the house from the main road. T-test conducted showed that the house density, container density, indoor mosquitoes egg count, outdoor mosquitoes egg count, and dumping sites mosquitoes egg count were higher at the roadside villages compared to border villages.

The work presented in [24] links mosquito survival and reproduction with various environmental factors such as rainfall, temperature, living conditions, demography structure domestic waste management and population distribution. A geostatistical method has been used in this study to analyze the correlation between dengue fever, population distribution and climate factors. Authors showed that the spatial variation of dengue incidence can be mapped by combining GIS with geostatistical analysis and space-time permutation scan statistic tools. They support their claim with the fact that Geographically weighted regression (GWR) analysis produced a strong (R2= 0.87) positive spatial correlation between dengue fever and population distribution. Vaidya A. at el. [61] introduces a mathematical, compartmental model to forecast the population dynamics of a mosquito and its life cycle in relation to seasonal variations of temperature and rainfall. Populations within the compartments were expressed in the form of a system of coupled differential equations, which describe changes in the mosquito population through processes of maturation and mortality. By using regression tools, maturation and mortality rates at various temperatures were estimated.

3.2.2 Neural Network

A group of researchers had predicted the dengue confirmed cases by using the neural network [32]. The average temperature, average humidity, total rainfall and the number of confirmed dengue cases were used in model training. There are 14,209 dengue cases were used in the training of the model. Authors reported the results are encouraging and the proposed prediction model can be used worldwide. The model was kept time agnostic by eliminating time factors in the model training.

An automatic prediction system of Dengue Hemorrhagic Fever outbreak by using entropy and ANN is proposed in the research study presented in [25]. In this study, authors mandate the information preprocessing prior feeding into ANN. This step will eliminate redundant data and noises. Temperature, relative humidity, and rainfall were considered in the information extraction phase. Then, a supervised neural network was used to predict the possible risk of Dengue Hemorrhagic Fever outbreak. The performance of the proposed system was evaluated based on the experiments conducted with weather data and Dengue Hemorrhagic Fever cases from January 1999 until December 2007. Authors claim 85.92% accuracy.

3.2.3 Cellular Automata

Cellular automata models began from the concept of John von Neumann to make the machine copies itself. Cellular means "consist of cells". Cellular automata can be multidimensional. If there are two dimensions, it resembles a checkerboard. Each cell has some adjacent cells and called "Neighborhood". Changing the status of a cell in a one-time step depends on local rules. The local rules may be the probability [35]. This research work uses Moore neighborhood with radius=1 and uses the probability in changing status. The cellular automaton model is used with SIR and SEIR infection propagation models. For SIR model, each cell has only one status in one-time step such as 'S' represents susceptible, 'I' represents infected and is able to transmit the disease to the others, 'R' represents recovered. Some diseases have a latent period, a status for this period is 'E' and is called SEIR model [34]. An outbreak of dengue fever is characterized by a SEIR model. Some people are not sick when exposed to the dengue virus. The patient will have an incubation period of about a week after exposure to the virus and before symptoms to appear. Another study using cellular automata created a model of Hepatitis B Virus (HBV) [31]. The cellular automata (CA) lattice size was 300x5000. The status of a cell in lattice might be "susceptible”, "infected”, "core" or "immune". The local rules were the probability.

A time series model to predict the number of patients with Chickenpox by using Probabilistic Cellular Automata was proposed by a group of researchers [33]. A chromosome of the genetic algorithm consists of the state changing probability. Experimental results showed that the bigger number of cells in the lattice is better than fewer numbers of cells. A different approach is taken in the research work presented in [30]. The proposed model considers the number of people in each status of the epidemic model SIER. In this respect, CA take a Genetic Algorithm (GA) to generate the factor weight chromosomes and ANN to determine the probability of state transition 'S' to 'E' at time step t (Pt (s, e)). In addition, other related probabilities are obtained by expert knowledge; P (e, i) = 0.15 and P (i, s) = 0.001. P (r, s) is determined by GA. These probabilities were used to calculate the cell number of each state at the next time step of GA. GA compute the fitness for one-time step and repeat every time step finally to compute RMSE. For performance evaluation, 32 factors of dengue causes are used in the model. The dataset collected from 2005 to 2011 consisting of 359 weeks in which 287 and 72 are used to train and test the model, respectively. Authors claim, with the results obtained that their method, outperforms the artificial neural network approaches.

3.2.4 Support Vector Machine

Support vector machine is used in various fields to perform pattern recognition successfully. These areas include face detection/recognition, object detection, image retrieval, information retrieval, speech recognition, and prediction/forecasting. SVM is also used as a regression model in which a value for the dependent parameter is given instead of the class of the parameter. The aim of many nonlinear forecasting methods [26, 27, 28, and 29] is to predict next points of time series. Tay and Cao [29] proposed C-ascending SVMs by increasing the value of C, the relative importance of the empirical risk with respect to the growth of regularization term. This research assumed, assigning more weights on recent data than distant data results in better performance. C-ascending SVMs produce better results than standard SVM. Fan et al. [28] adopted the SVM approach to the problem of predicting corporate distress from financial statements. In this problem domain, the performance is affected by the choice of input variables. Authors also claimed that selecting suitable input variables has a positive impact on the performance. Input variable must be selected in a way that maximizes the distance of vectors between different classes and minimizes the distance within the same class. Euclidean distance-based input selection generated a better performance.

3.3 Response Planning

Dispensing treatments to the general public during an emergency is an important task. There are numerous works have been done on various aspects of response planning. Every plan must adopt strategies to distribute supplies and dispense medication to the affected population within a specified time frame [36]. There are other important factors such as vulnerabilities of the population must be considered when developing a response plan. Each of these factors is considered separately in different works. Routing and scheduling the distribution of supplies have been addressed in different ways, and management of treatment facilities has been studied. Different strategies to distribute medication among the facilities have been introduced in the research work [37]. Distributing medications and treatment supplies to each local agency is a challenge. The concept of Point of Dispense (PODs) was introduced by the Center for Disease Control (CDC). PODs strategy is been recognized by the authorities as an effective method of planning an emergency. The CDC maintains a warehouse of treatment supplies and medications and delivers them in accordance with the demand to the local authorities. It is the responsibility of each local authority to develop their own response plan adheres to the guideline setup by the CDC [36]. The PODs concept is well utilized in the RE-PLAN framework, developed at the Center for Computational Epidemiology and Response Analysis (CeCERA) [38]. The framework is capable of producing an effective response plan based on PODs placement where necessary. There are couple of different PODs placement methods introduced in the framework. Each of which is suitable for different scenarios. The RE-PLAN is a better candidate for the proposed work, to be integrated with, based on the available features of the RE-PLAN. An additional strain is imposed when dealing with mass events involving mass number of people [39]. Public health and policy studies stressed the mandating the managing of limited resources during an emergency [40]. Ethic that involve in allocating resources during a mass casualty event is presented in [36] [41].

The spatial data of the region is a critical component in response planning. This will allow localization of data and provide a visual feedback to the plan designer. The spatial data may include population distribution and road infrastructure, and census data with arbitrary census blocks. Geographical Information System (GIS) is needed to set up an effective management and manipulation of spatial data. Integration of GIS data and usage of spatial tools in response planning are widely studied and recommended by various studies [42]. RealOpt [43] proposed a simulation and decision support system created to support planning, designing and placing large-scale emergency dispensing clinics for emergency responses. The Centers for Disease Control and Prevention created BioSense, a surveillance system [44], is targeting at early detection of biological emergency events. Coombes [45] and Schneider et al [46] proposed several methods of defining boundaries for the response planning area. The authors proposed algorithms to solve problems associated with discrete and continuous PODs allocation [47, 48, 49, 50, 51 and 52].

3.4 Resource Allocation

Resource allocation is mostly a subpart of response planning for a natural or manmade disaster. There are several research efforts that focused on the optimum allocation of available resources among different entities during disasters. They address different disaster types such as wildfires [64], earthquake [65], and public health emergencies [66].

The research study presented in [64] introduced a minimal stochastic process to model wildfire progression that captures the statistical distribution of fire sizes. The authors then coupled the model to a series of response models that are targeting the measurement of performance of timing and suitability of response plans and also dealing with the distribution of limited resources among multiple entities. The authors proposed to use the framework to compute the optimal strategies for decision-making scenarios. The proposed framework is a set of guidelines and does not involve a method of computation for resource allocation.

A dynamic optimization model can be used in resource allocation [65]. The model was fed with available resources and a detailed description of the concerning areas to calculate the resource performance. In [66], a solution to overwhelming demand for healthcare resource during a large-scale public health emergency was proposed. The authors discussed a resource allocation approach for optimizing regional aid during public health emergencies. They presented a relationship between the optimal response and delaying the distribution of resources from the central stockpile. And also presented that policy level decisions that alter the objectives of pandemic relief efforts can significantly impact the allocations to affected regions. This study is not presenting a computational framework that can be used to allocate resources among facilities.

The agent-based simulations have been using in resource allocations [63]. The authors proposed an agent-based simulation for allocation of resources for a response plan that involved two main facilities. In this study, they try to minimize the hospital arrival times for critically injured casualties. Further, they investigate how the optimal resource allocation depends on the distribution of casualties across the two sites. The author also claimed that this study is tested only on two sites and hence further improvements are needed to apply for a larger number of sites.

I propose a solution to analyze and predict the dengue epidemic with a trained model which considers socio-economical and geographical factors. The main simulating factor for the proposed work is the differences in the administrative regions. Each province has its own geographical and socio-economical characteristics and can be different from each other. Assuming all the districts are homogeneous may result in a biased conclusion. We need to overcome this shortcoming with the introduction of local characteristics in the model that are specific for each administrative unit. And also, a mechanism, in which the differences in geographical regions is considered, must be integrated into the solution. There is no direct way of feeding geographical differences into the SVR model. I proposed to use micro-ensemble for each district to handle the differences effectively and getting the differences into the model indirectly.

CHAPTER 4

METHODOLOGY

The methodology of obtaining an efficient prediction system involves several steps. It goes through a distinct set of thorough processing from data processing to resource allocation. System flow diagram for the proposed study is shown in the Figure 4.1.

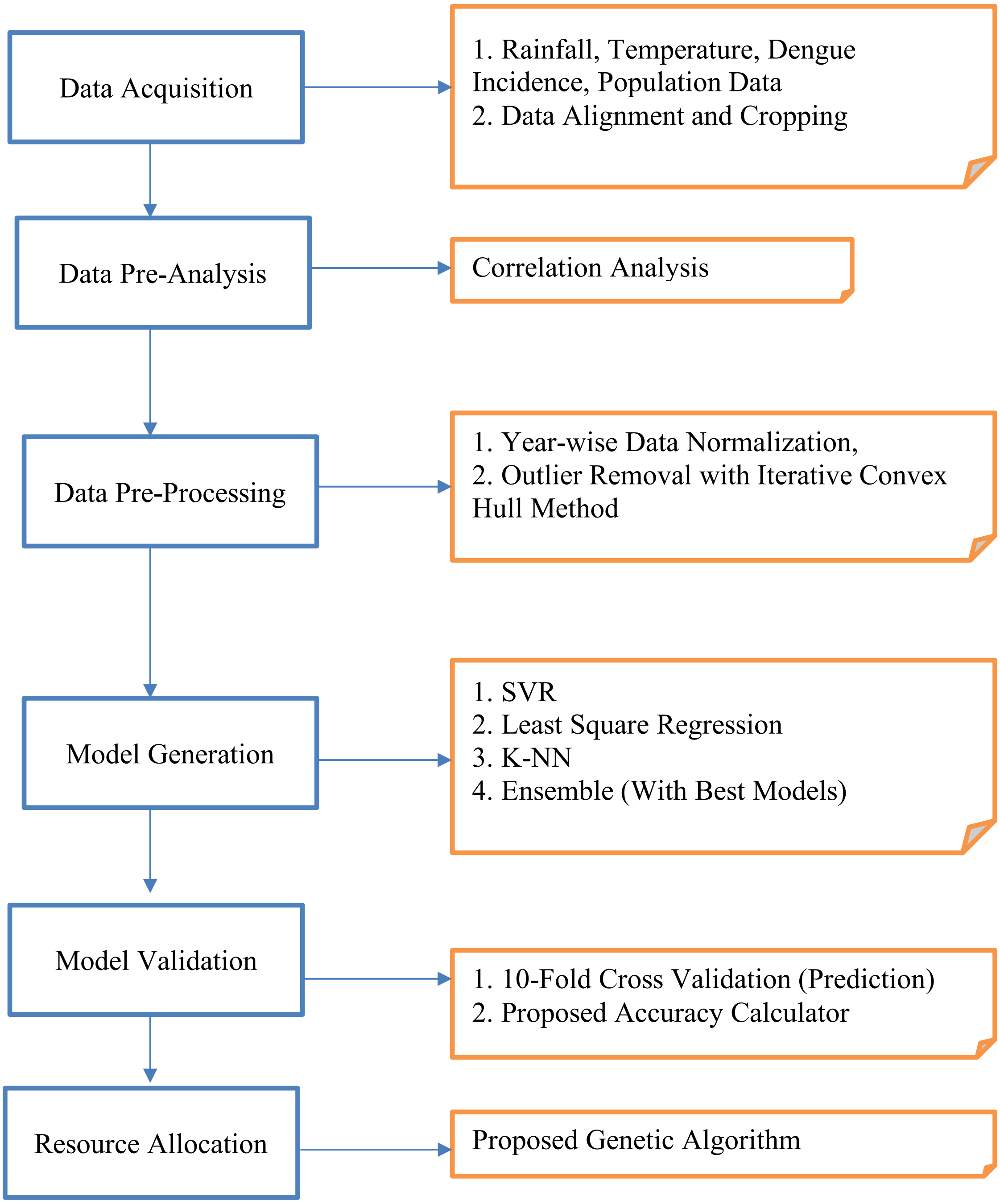


Figure 4.1 Proposed Work-Flow for Dengue Epidemic Mitigation

4.1 Data Gathering

The number of dengue cases reported depends on various factors such as rainfall, temperature, population density, waste management efficiency, land use, and water body management etc. In this study, rainfall, temperature, and population densities are considered in model generation. These factors are gathered from various sources based on the availability. The following sections explain my strategies of obtaining each factor.

4.1.1 Rainfall Data

The rainfall data was obtained from Global Rainfall Map in Near Real Time (GSMaP_NRT) distributed from the JAXA Global Rainfall Watch, which was developed based on activities of the GSMaP (Global Satellite Mapping of Precipitation) project. The GSMaP project is promoted for the study "Production of a high-precision, high-resolution global precipitation map using satellite data," that is sponsored by the Core Research for Evolutional Science and Technology (CREST) and it is a part of the Japan Science and Technology Agency (JST) [14]. The GSMaP_NRT repository provides hourly rain rate data in 0.1-degree resolution (10km at the equator). The repository divides the globe into 15 distinct regions as shown in Figure 4.2 and provides the rainfall data separately for each region as Comma Separated Values (CSV) files. The registered users get free access to the repository. The users can get data using an FTP client that is connected to the repository using credential provided by the repository management. Thailand is included in 02_AsiaSE area. Table 4.1 lists location specific information for each Asian region. For the model training, data from five consecutive years was used.

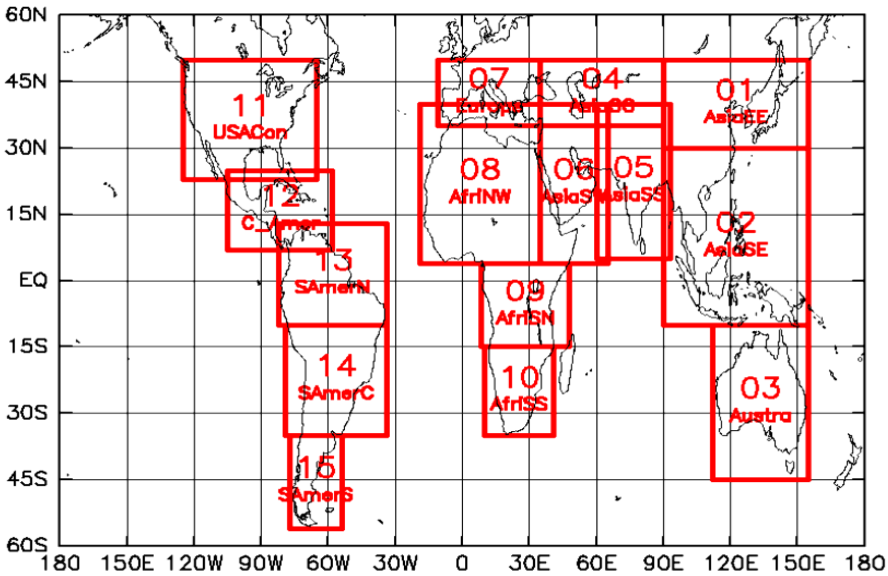


Figure 4.2 Definition of text areas of JAXA data repository for text data [14]

Table 4.1 GSMap text area declaration for Asian region [14]

Area name	**Lon (W)**	**Lon (E)**	**Lat (S)**	**Lat (N)**	**Description**
01_AsiaEE	90	155	30	50	East Asia
02_AsiaSE	90	155	-10	30	South East Asia
04_AsiaCC	35	90	35	50	Central Asia

05_AsiaSS	60	93	5	40	South Asia
06_AsiaSW	35	65	4	40	Arabian Peninsula and East Africa

4.1.2 Temperature Data

Temperature data was obtained from the Thai Meteorological Department (TMD) and the NASA Earth Observations data archive [69]. The average temperature value for each month for each district was used in the training. Time span of temperature data is five consecutive years.

4.1.3 Dengue Case Data

The Dengue case data for each district for five consecutive years were obtained from Sri Lanka and Thailand Epidemiology Units. Dengue case data is given in three groups which are Dengue Hemorrhagic Fever (DHF), Dengue Fever (DF) and Dengue Shock Syndrome (DSS). I combine all three categories to form a single entity and used in model training as dengue cases. The Dengue case data was obtained from the Department of Disease Control, Ministry of Health.

4.1.4 Population Data

Population data was obtained from the Department of Census from both countries.

4.2 Data Processing

Preprocessing of data is needed as the data sources provide data in different format and at different time and spatial resolutions.

4.2.1 Extracting Relevant Data and Alignment of Time Resolution

The GSMaP_NRT region 02_AsiaSE covers a larger area than Thailand geographical region (Figure 4.2). And the region 05_AsiaSS covers a larger area than Sri Lanka. This results in a large amount of non-related data being loaded into the spatial database making it heavy for fast computations. To reduce the data load overhead, only the rainfall data that falls inside Thailand and Sri Lankan geographical area were obtained by cropping the dataset using longitude and latitude. The non-relevant data was discarded. As the time resolution of rainfall data is one reading per hour, monthly rainfall data was computed from hourly data. This matches the time resolution of each factor before use in training process as temperature and population data recorded on monthly basis. Further, there are multiple observation points fallen in a single district as shown in Figure 4.3. The average accumulated value of all the points that fall in a district was taken as the monthly rainfall of that district. The unit of recording is mm per hour(mm/hr). Sample data file format for rainfall from the GSMaP_NRT is given in Table 4.2.

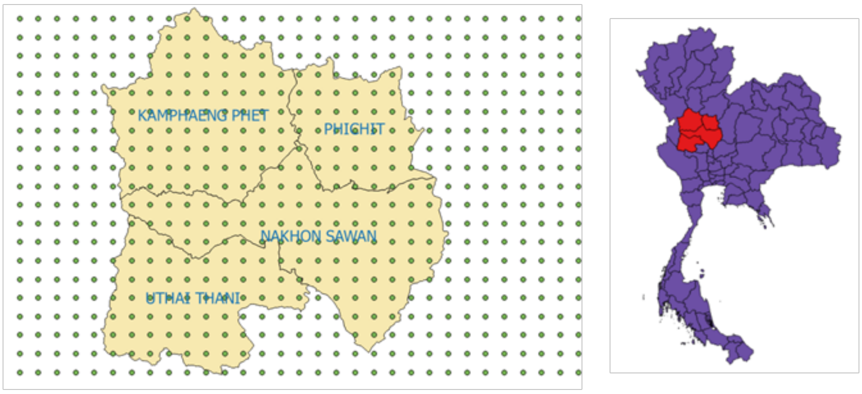


Figure 4.3 Rain fall data observation points and geographical boundaries of all four provinces

Table 4.2 Fragment of rainfall data text file from GSMap_NRT

Lat	**Lon**	**RainRate**
20.95	97.05	0.1
20.85	97.05	0.06
20.75	97.05	0.04
20.65	97.05	0.06

4.3 Pre-analysis of Data

The proposed model used in this study is the SVR [15] and Ensemble. The SVR is built on regression analysis. To get a better result from a regression analysis, there must be a positive correlation between explanatory variables (factors) and dependent variable (dengue incidence). As the primary model of prediction is SVR, this study needs a data analysis before moving forward with the SVR. A separate correlation analysis was conducted for each factor (rain, temperature and population) to determine the suitability of the regression analysis of the proposed factors. Correlation is a statistical relationship between those two sets of data which describes the strength of the relationship between those two data sets. If the correlation is low, there is a weak interdependency between those two sets. If the correlation is high (normally greater than 0.5 negative or positive), there is a considerable relationship between those two sets. Correlation of two data sets is computed as given in the equation below.

$$\rho_{X,Y} = \frac{E[(X-\mu X)(Y-\mu Y)]}{\sigma X \sigma Y} \qquad (1)$$

Where $\rho_{X,Y}$ is the correlation between datasets X and Y. E is the expected value operator. μX is the mean of data set X, μY is the mean of data set Y. σX and σY are standard deviation of data sets X and Y respectively. The correlation value is interpreted as shown in Table 4.3.

Table 4.3 Correlation values and their meanings

Correlation Value	Interpretation
-1	A perfect downhill (negative) linear relationship
-0.7	A strong downhill (negative) linear relationship
-0.5	A moderate downhill (negative) relationship
-0.3	A weak downhill (negative) linear relationship
0	No linear relationship
+0.3	A weak uphill (positive) linear relationship
+0.5	A moderate uphill (positive) relationship
+0.7	A strong uphill (positive) linear relationship
+1	A perfect uphill (positive) linear relationship

4.4 Pre-processing of Data

Pre-processing of data is needed as to eliminate human recording errors and machine introduced noises.

4.4.1 Year-Wise Data Normalization to Eliminate Year Specific Influences

Rainfall pattern for each district is a recurrent pattern that has little variation from the average rainfall data for the year. Dengue incidences are varying from year to year due to various reasons. Dengue case data has a constant and strong correlation to the rainfall data. The correlation of the rainfall data is specific to the reference year and does not hold for every year in general due to various other influencing factors. There may be a boost or a decline in the number of cases due to some other influences such as temperature, special environmental events such as flood and droughts, etc. The rainfall pattern along with dengue incidence for the district Amnat Chareon is given in the Figure 4.4 for six years starting from 2012. This variation is not cooperating well with machine learning tools especially with regression tools. Therefore, I use year-wise normalization to eliminate or to reduce the impact of the afore mentioned effects from the rainfall and dengue incidences. The normalized dengue incidence pattern along with the rainfall data for the district Amnat Chareon is given in the Figure 4.5.

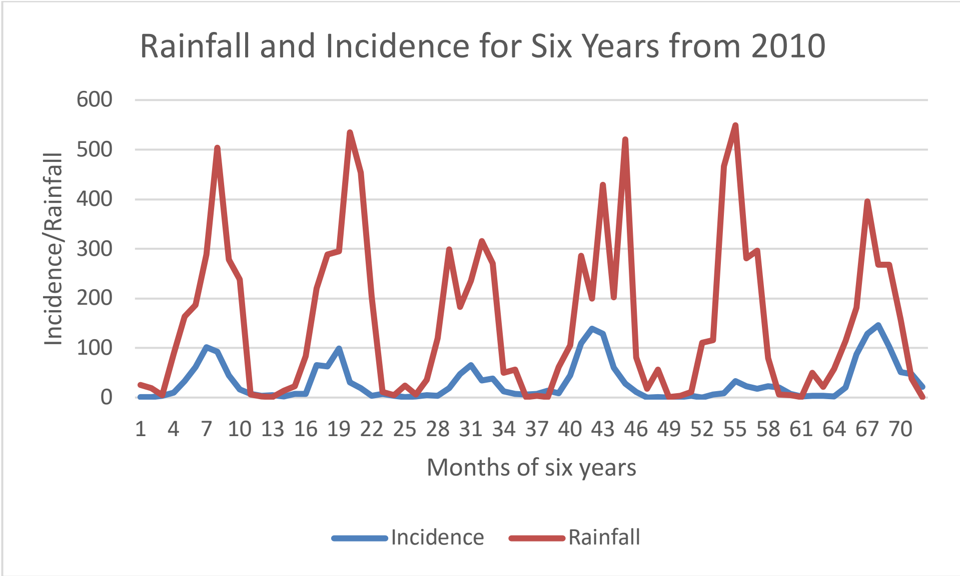


Figure 4.4 Monthly Rainfall and Incidence Data for Six Years from 2010

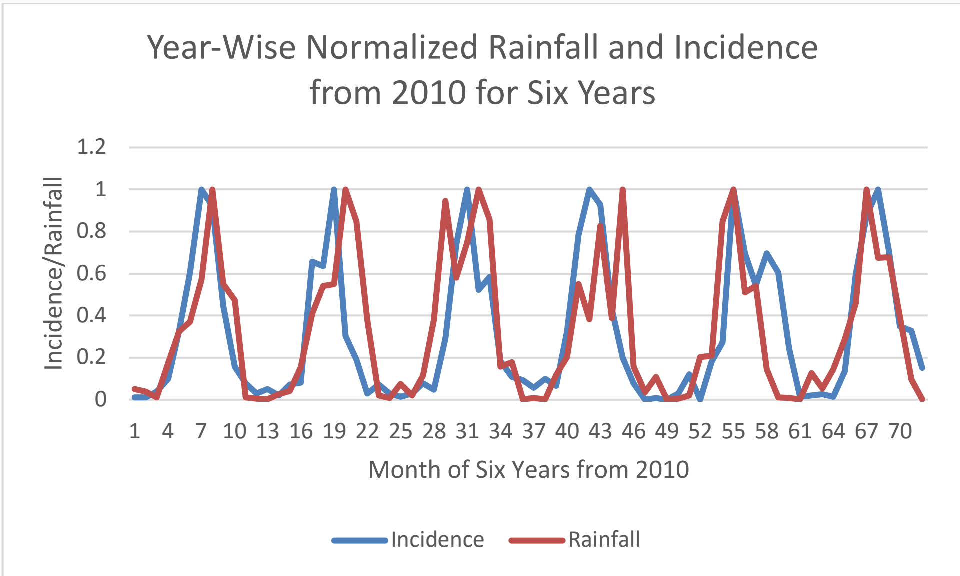


Figure 4.5 Normalized Monthly Rainfall and Incidence Data for Six Years from 2010

4.4.2 Outlier Removal

The Dengue case data may be reported to the authorities partially. If the data reported is partial, the relationship between influencing factors and dengue cases is not clearly seen. It is critical to identify these instances in advance and treat them properly to ensure the quality of the generated predictive model. These unusual instances are known as outliers. I proposed a method of outlier removal in which a convex hull is used to determine the outliers. This method uses the fact that the outliers (extreme points) lie further to actual points. As the correlation between influencing factors and the dengue cases reported is strong, the chance of having an observation that position further away from the cluster of points is less. This technique is shown in the Figure 4.6.

.

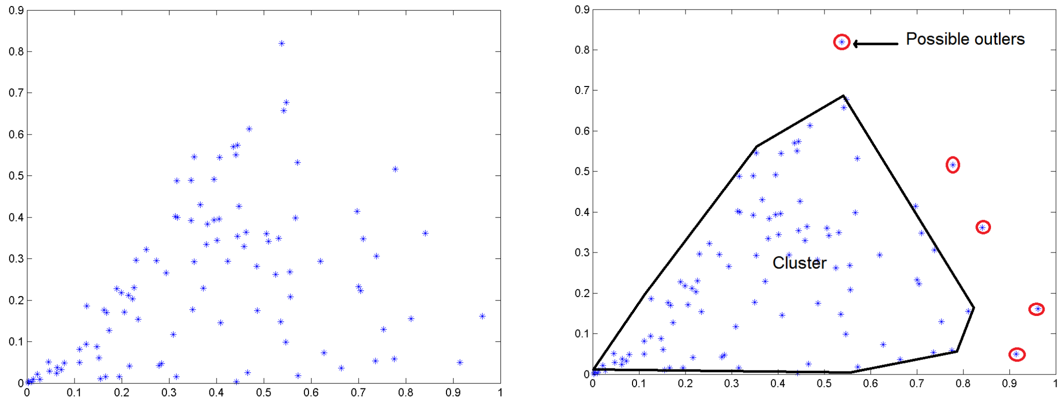


Figure 4.6 Outliers in Data Points

The proposed method operates in two stages which are identification of the presence of an outlier and removal of identified outlier. In the first stage, a convex hull is generated for the projection of each influencing factor to the dengue incidence in 2-dimensional Cartesian coordinate system. This stage is shown in the Figure 4.7 with and without outliers.

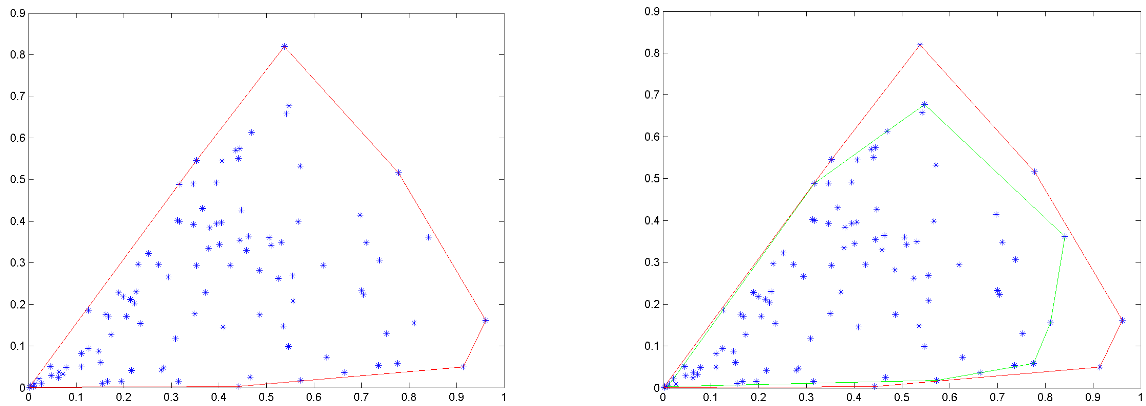


Figure 4.7 Outlier Removal Levels

4.4.2.1 Iterative Convex Hull Reduction to Remove Outliers

The outlier removal takes place as an iterative process. First, the convex hull of the two-dimensional dataset is generated. The area of the outer most convex hull is computed. In the next iteration, the points that formed the convex hull are removed from the original data set and regenerate the convex hull for the remaining points. Compute the difference in areas of the convex hull generated from the previous iteration to the current iteration. Iteration stops if the difference in areas is below a predetermined threshold value or preset number of maximum iterations reached. If the difference in area is significant, that indicates the presence of an outlier and hence need to determine the outlier point from the set of convex hull points.

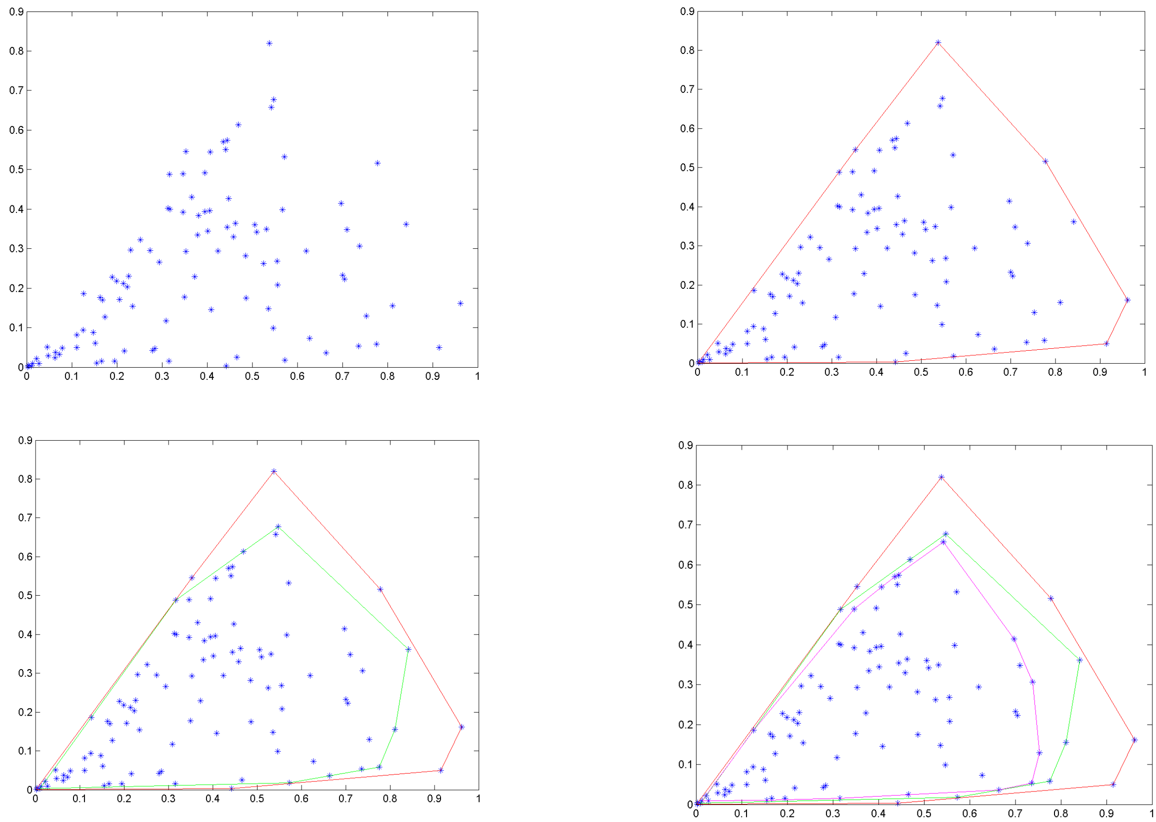


Figure 4.8 Three Levels of Outlier Removals

Let ΔA is the difference in area between two iterations.

$$\Delta A = |A(ConvexHull)_i - A(ConvexHull)_{i-1}|$$

4.4.2.1 Determination of Points to be Removed from Outer Convex Hull in the Presence of Outliers

Once the presence of the outliers detected, next step is to identify the exact outliers from the points of outer convex hull. This step is important as all the points that make outer convex hull are not outliers. I perform an exhaustive search to identify the outliers among the points on the outer convex hull. I compute the difference in surface area keeping one point of the convex hull out. All the points that generate a difference in area greater than or equal to a predetermine threshold value are classified as outliers. All the remaining points stay in the dataset.

4.5 Model Generation

The model generation is performed on the processed data sets that went through pre-analysis and pre-processing steps. The data set is divided into two parts and used in model generation and model validations steps. The model generation steps are given in the following sections. The SVR, LS and KNN are trained with the data sets.

4.5.1 Support Vector Regression

The behavior of each factor (rainfall, temperature, and population density) on dengue cases is spatially dependent. The effect of rainfall on dengue incidence for each district is different from district to district as emphasized in [5]. Hence, a separate analysis for each district was conducted and a separate model for each province was generated. Data for five years were combined for each district and fed into the model for training. The proposed arrangement can capture the spatial heterogeneity of each province and hence improve the performance of the prediction model.

The SVR model is based on the regression analysis. A regression analysis can estimate the relationship between two data sets (random variable) and fit a cure to the data sets (explanatory variable and dependent variable). This curve can then be used in prediction of unknown cases. The regression curve for this study has three explanatory variable, Rain R, Population P, and Temperature T. The regression model for this study is given in the equation below.

$$C_{di} = \beta_0 + \beta_1 P_i + \beta_2 R_i + \beta_3 T_i + \varepsilon \qquad (2)$$

Where Pi is the population in i^{th} region, Ri is the rainfall for i^{th} region and T_i is the temperature for i^{th} region. The error term is ε. C_{di} is the dengue cases for region *i*. Intercept is β_0 and it is a constant.

SVR improves the detection speed as it keeps only a subset of training data as support vectors in the model. The SVR uses the same principles as the Support Vector Machine (SVM) for classification, with only a few minor differences. SVR's output is a real number which makes it difficult to match target output on test dataset. A margin of tolerance (epsilon) is set in approximation to the SVM to address the problem associated with real numbers output. General construction of SVR is given in the following equations.

SVM regression is constructed by first mapping the input vector X into an m-dimensional feature space using a non-linear mapping function. The linear regression model is then constructed in this feature space. The linear model $f(x, \omega)$ is given by equation 3.

$$f(x, \omega) = \sum_{j=1}^{m} w_j g_j(x) + b \qquad (3)$$

Where $g_j(x), j = 1, \dots m$ denotes a set of nonlinear transformations and b is the "bias" term. The bias term can be dropped with the assumption of zero mean data set.ω is the normal vector.

The quality of estimation is measured by the loss function $L_\varepsilon\left(y, f(x,\omega)\right)$ given in equation 4. The loss function is computed as proposed in [16].

$$L_\varepsilon\left(y, f(x,\omega)\right) = \begin{cases} 0 & if \quad |\, y - f(x,\omega)| \leq \varepsilon \\ |y - f(x,\omega)| - \varepsilon & otherwise \end{cases} \qquad (4)$$

Then the empirical risk function is given in equation 5.

$$R_{emp}(\omega) = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(y, f(x,\omega)\right) \qquad (5)$$

The model generated by minimizing the $\|\omega\|^2$. This can be achieved by introducing (non-negative) slack variables ξ_i, ξ_i^* $i = 1, \dots n$ to measure the deviation of training samples outside ε-insensitive zone. Thus the SVM regression is formulated by minimization of the function given in equation 6.

$$\arg\min{}_{(\omega)} = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad (6)$$

Such that $\begin{cases} y_i - f(x_i,\omega) \leq \varepsilon + \xi_i^* \\ f(x_i,\omega) - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0, \; i = 1, \dots, n \end{cases}$

This optimization problem can be transformed into the dual problem and its solution is given by the equation in 7.

$$f(x) = \sum_{i=1}^{n_{sv}}(\alpha_i - \alpha_i^*)K(x_i, x) \qquad (7)$$

Such that: $0 \le \alpha_i^* \le C, \ 0 \le \alpha_i \le C$

Where $n_{sv}$is the number of Support Vectors (SVs) and the kernel function is given by the equation in 8.

$$K(x_i, x) = \sum_{j=1}^{m} g_j(x) g_j(x_i) \qquad (8)$$

The RBF was used as the kernel function and epsilon was set to 0.001. The cost parameter was kept at 100.

4.5.2 k-Nearest Neighbor Regression

The *k*-nearest neighbor (*k*-NN) Regression is a conventional non-parametric classifier [67]. The *k*-NN classifier calculates the distances between the testing point and points in the training data set. There are several distance calculation methods available. Euclidean distance is the most popular metric. k is an integer. The value selected for k has a direct impact of the predicted value and hence must be carefully selected for the problem domain. For this study, Euclidean distance was the distance metric and number of neighbors was three. The final predicted value is the average of k nearest neighbors.

To measure the distance between points *A* and *B* in a feature space, various distance functions have been used in the literature, in which the Euclidean distance function is the most widely used one. Let *A* and *B* are represented by feature vectors $A = (x_1, x_2, \ldots, x_m)$ and $B = (y_1, y_2, \ldots, y_m)$, where *m* is the dimensionality of the feature space. To calculate the distance between *A* and *B*, the normalized Euclidean metric is used by

$$dist(A,B) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \qquad (9)$$

4.5.3 Least Square Regression

Linear least squares regression is a widely used modeling method. It is also referred to as regression, linear regression or least squares. linear least squares regression can be used to fit the data with a function of the form

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \cdots \beta nxn \qquad (10)$$

in which

1. Each explanatory variable in the function is multiplied by a coefficient
2. There is one constant
3. All terms are summed to form the final function

Error minimization is done as the same way as SVR. The only difference is the way the model is generated. In SVR it uses support vectors to keep the model parameters.

4.6 Prediction

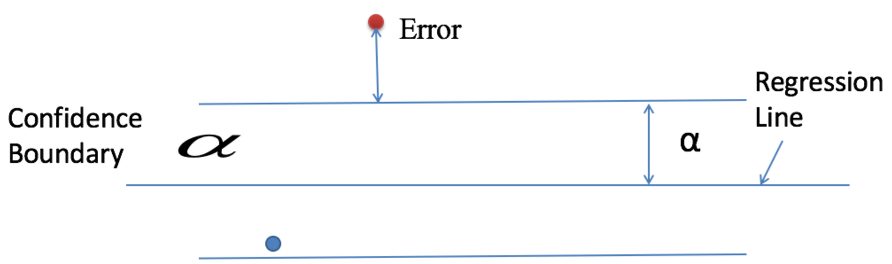
A vector of unseen data for rainfall, temperature and population data is fed into the trained model and estimated output is obtained from the SVR, KNN, and LS. This output is not a label as in SVM. It is a real number approximating the dengue incidence pertaining to the given scenario. The dengue incidence predicted show cases the severity of the condition that may occur if the given scenario appears in the future.

4.7 Model Validation

Conventional regression models are evaluated based on the Mean Square Error (MSE) of the cross validation (mostly 10-fold cross validation). MSE cannot capture the total picture of the behavior of the data set. Several outliers can affect the final outcome of the validation. Another problem of regression analysis is there is no way of computing the accuracy of the prediction with cross validation. Regression gives real values as estimates and there is theoretically infinite number of possibilities with a real number. This fact makes it impossible to compare against target value.

In this study a novel yet, simple accuracy calculation method was introduced. A positive confidence boundary parameter α was included in cross validation. If *|actual value – estimated value| > α*, we label the estimated value as a correct prediction and incorrect prediction otherwise. The proposed accuracy calculation is given in the

Figure 4.9.



$$\text{Class} = \begin{cases} 1 & \hat{y} - \alpha < y < \hat{y} + \alpha \\ 0 & otherwise \end{cases} \qquad Error = |\hat{y} - y| - \alpha$$

Figure 4.9 Accuracy Calculation of SVR

Accuracy is calculated according to the equation given below.

$$\mathrm{ACC} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \qquad (11)$$

4.7.1 Determination of the Degree of Fit of the Regression Model to the Dataset with Parameter Alpha (α).

The value of α is inversely proportional to the model accuracy. If the model generates a higher accuracy value for a lower value of α, the regression model fits well to the dataset. If the model accuracy is high only for a large value of α, *the* dataset is loosely correlated to the influencing factors. Higher accuracy for a smaller alpha value indicates that the dataset and the fitted regression is a best fit for the problem domain.

4.7.2 10—Fold Cross Validation

Cross-validation is a model validation technique for assessing how the results of an analysis will generalize to the underlying population of the extracted data set. It is used in predictions where one needs to evaluate the outcome of the model. In a prediction problem, a model is usually given a training dataset with known outcome values, and a dataset of unknown outcome data. The goal of cross-validation is to test the model's ability to predict new data that was not used in training stage of the model. In this study, I used 10-fold cross validation where dataset is divided into 10 subsets. One portion is kept as the testing data set and 9 portions are used in training. This process is repeated 10 times alternating the testing portion from the initial division. Final outcome of the model is obtained averaging the 10 outputs of the process.

4.8 Resource Allocation

A successful mitigation plan includes an effective resource allocation scheme. An effective resource allocation should eliminate wastage, over allocation and under allocation of resources while satisfying the demand of each facility for a given resource. Over allocating may result in run out of available resources before every demand is satisfied. Under allocating may result in failure of mitigation plan. Hence, an effective resource allocation is needed to ensure a proper distribution of available resources among facilities to satisfy their demands.

4.8.1 Problem Definition

Let F_1,F_2….F_n is the set of facilities demanding for resources R_1, R_2,……R_m with a demand for each resource W_{00}, W_{01},…….W_{nm}. with the constraint $W_{ij} \in \mathbb{R}\,[0, m]$. As the available resources may not be sufficient enough to fulfill the demand of each facility, the allocation algorithm must generate an optimum allocation that minimizes the penalty of allocating less than the demand. The final allocated scheme is represented as A_{00}, A_{01},…..A_{nm}. This problem can be modeled as a bi-partite matching between facilities and resources with a weight of each mapping as the demand. I represent this problem using a bi-partite grape as show in Figure 4.10.

4.8.2 Problem Representation

The problem of resource allocation with demand can be successfully represented with a bi-partite graph. The graph is constructed as resources and facilities as two sides and demand as weight of associations. The graphical representation can be seen in Figure 4.10. In the figure resources are represented with R_0…R_m, facilities are represented with F_0,…F_n

and corresponding demands are represented with W_{00},.....W_{nm}.

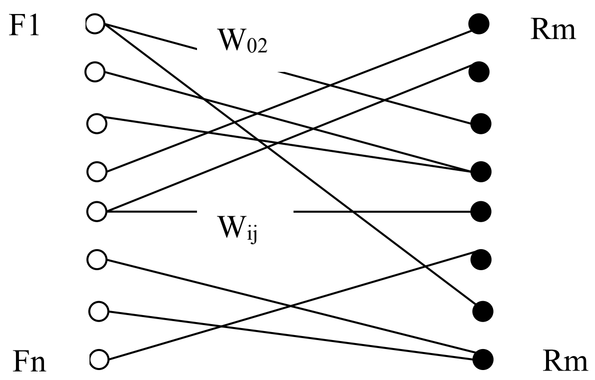


Figure 4.10 Bi-Partite Graph of Resource Mapping

4.8.3 Digital Representation

The resource allocation problem is represented as an adjacency matrix that is algorithm friendly. Resources are arranged along x axis and facilities are arranged along y axis. Weight (demand) of the matching is the corresponding element of the matrix.

Demand of resources

$$\begin{array}{cc} & \text{R1} \;.\; \ldots \quad \ldots \quad \ldots \;\; \text{Rm} \\ \begin{array}{c}\text{F1}\\ .\\ .\\ \text{Fn}\end{array} & \begin{bmatrix} W_{00} & W_{01} & W_{0m} \\ \vdots & \vdots & \vdots \\ W_{n0} & W_{n1} & W_{nm} \end{bmatrix} \end{array}$$

Allocation of resources

$$\begin{array}{cc} & \text{R1} \;.\; \ldots \quad \ldots \quad \ldots \;\; \text{Rm} \\ \begin{array}{c}\text{F1}\\ .\\ .\\ \text{Fn}\end{array} & \begin{bmatrix} A_{00} & A_{01} & A_{0m} \\ \vdots & \vdots & \vdots \\ A_{n0} & A_{n1} & A_{nm} \end{bmatrix} \end{array}$$

4.8.4 Finding the Optimum Solution for Resource Allocation Problem

The best allocation is to deliver each facility resources they demanded. The scarcity of resources prevents us from delivering the demanded amount of resources for all the facilities. It is required to find the optimum resource allocation that will minimize the impact of being delivered below the demanding amount. It is required to determine an effective method of calculating the impact of resource allocation on each facility. I suggest calculating a penalty score for each facility based on its important properties. The severity of allocating fewer resources than demanded is depended on the population that needed to be served and surface area of the serving region. The penalty score is calculated according to the equation given below.

$$Penalty\ Score = \left|W_{ij}(Allocated) - W_{ij}(Demand)\right| * Population_i * area_i * risk_i$$

The goal of the minimization algorithm is to minimize the total penalty score which is given in the equation below.

$$Total\ Penalty\ Score = \sum_{i=0,j=0}^{i=n,j=m} \left|W_{ij}(Allocated) - W_{ij}(Demand)\right| * Population_i * area_i * risk_i$$

4.8.5 Weight Adjustments

The best approach is to use exhaustive search to look for the best match for weight for allocated matrix. The number of candidates in the exhaustive search space is increased exponentially with the number of facilities and number of available resource categories.

The exhaustive search is then become resource intensive and is not an ideal candidate to find weights of the allocated matrix. I propose to use Genetic Algorithm approach with several modifications to match the problem domain.

Genetic Algorithm can be successfully applied in searching for a best fit from large number of candidate solutions. GA performs best when the available number of candidates is high. State of the art GA is defined as follows.

Genetic Algorithm is an optimization technique built on the principle of Genetics and Natural Selection. GA comes under the category machine learning. Given a set of possible solutions to the given problem, GA tries to select the best match based on the criteria set in advance. The GA selects a subset of population that meets the given criteria. This subset undergoes two steps, namely crossover and mutation, producing new children similar to natural process of selecting best fit candidates. This process is repeated over several iterations. Each candidate solution is assigned a fitness value and the fitter individuals are given a higher chance to mate and yield more “fitter” individuals.

GAs is randomized in nature when selecting crossover points and mutation points. The randomness in crossover and mutation is not applicable in my problem domain. I introduce a variant of crossover and mutation such that the resulting off springs are compliance with the requirements of the resource allocation.

I propose several major modifications to the standard GA for achieving optimum results for the resource allocation problem. These modifications are generalized and can be applied to any resource allocation problem. The modifications are listed below.

1. Introducing chromosome representation for resource allocation
2. Change the way the standard GA generates the initial population.
3. Chane the way the standard GA does crossover operation
4. Change the way the standard GA does mutation
    a. Introduction of resource lock chromosome
    b. Introducing sliding mutation scheme to accelerate the convergence
5. Change the iteration of the GA to accommodate sliding mutation scheme

All the above stages are detailed in the following sections.

4.8.6 GA Representation of the Problem

The problem of allocating set of resources among set of facilities must be represented in a single string of genes called a chromosome. I select decimal numbers in each location of gene as oppose to binary digits in standard GA. The binary chromosome and the proposed chromosome are given in the Figure 4.11 below.

Standard binary chromosome

1	1	0	0	0	1	0	1	1	0

Proposed chromosome

Gene Boundary for Facility

W11	W12	..	W1m	..	..	Wn1	Wn2	..	Wnm

Figure 4.11 The Modified Chromosome for Genetic Algorithm

As shown in the figure, resources for a given facility must be placed in adjacent genes of the chromosome. Gene boundary is used here after throughout the document as to represent the set of resources for a given facility.

Resource allocation requirements that must be embedded in GA are given below.

- Total number of allocated resources must match the total number of available resources when performing mutations.
- The Gene Boundary must not be broken apart when performing crossovers.

4.8.7 Proposed Population Generation Procedure

As opposed to the standard GA I use a customized population generation technique. In the proposed scheme, a constraint is set forth and eliminates chromosomes that violate the constraint. All the resulting chromosomes allocate resources that are under the limits of availability of resources. The constraint is modeled as shown in equation below. The number of facilities is n and number of resources is m.

$$Allocated\ Amount\ of\ Resource(i) = \sum_{j=1}^{n} A_{ji} : 1 \leq i \leq m$$

Constraint -> $Allocated\ Amount\ of\ Resource\ (i)$ $\leq$ Available amount of resources (i)

4.8.8 Proposed Crossover Operation

Crossovers are performed such that the gene boundaries are preserved. A gene boundary is treated as a single entity in crossover and hence the problem reduced to binary crossover in standard GA. A sample crossover is shown in Figure 4.12.

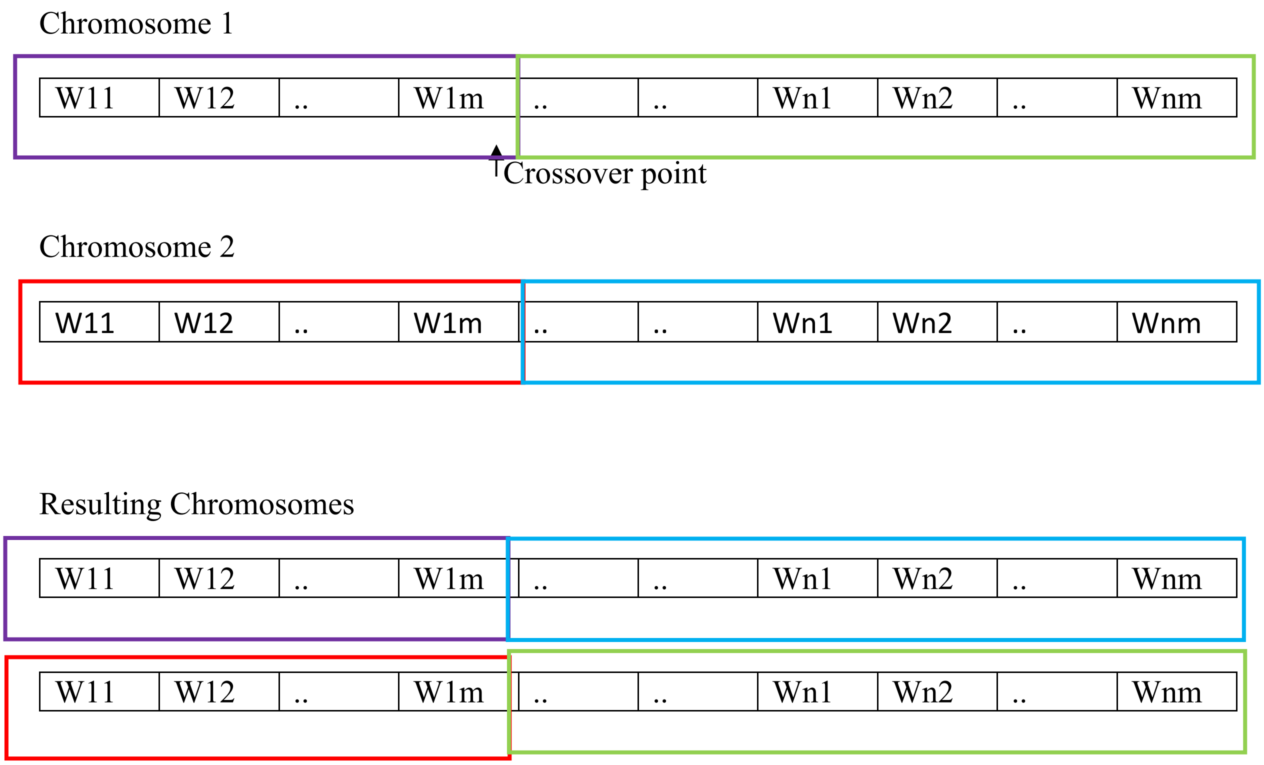


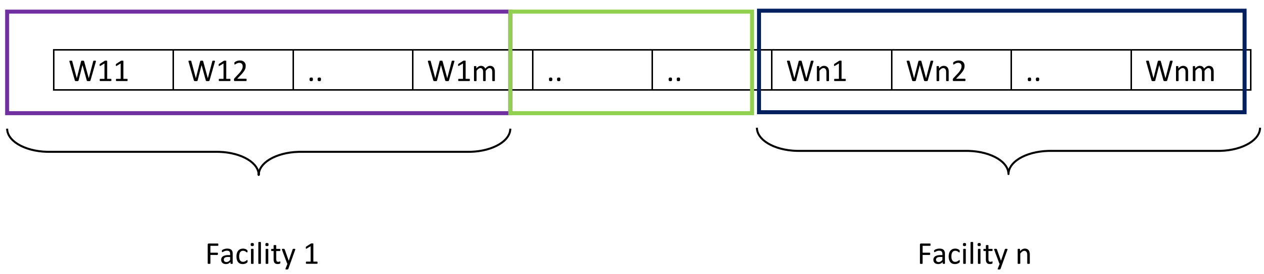
Figure 4.12 Proposed Cross-Over Operation

4.8.9 Proposed Mutation Operation

The proposed mutation operation selects number of mutation points based on the mutation rate as like standard GA. It is required to make sure that the resulting chromosome is adhering to the requirements set forth by the problem domain as opposed to standard GA. In standard GA it is simply select random points based on the mutation rate and flip the selected bit. In the proposed mutation operation, if the resulting chromosome demands more resources than available amount of resources after mutation it is not a good fit. A random mutation is not suitable in the proposed problem domain. A constraint must be set in mutation operation to consider the total amount of allocation for a given resource must not exceed the total amount of availability of the resource.

I randomly pick a resource type from the list of available resources. Next, set of facilities is randomly picked to apply the mutation. A resource exchange operation takes place between selected facilities for randomly selected resource type. This will make sure that the allocated amount of resources is not exceeding the available amount of resources. The proposed mutation operation is shown in Figure 4.13. Let the mutation operation randomly picked facility 1 and facility n for resource exchanging. All the other facilities remain unchanged and resource type 1(W_{i1}) was randomly selected for mutation.

Chromosome 1



Mutation -> $W_{11} = W_{11} – a$

Mutation -> $W_{n1} = W_{n1} + a$

Figure 4.13 The Proposed Mutation Operation

4.8.10 Fitness function

The fitness function is modeled as a function of risk associated with the given facility based on the location of the facility, population count of the facility, amount of area to be served by the facility. The proposed method of fitness calculation will make sure that resources are allocated appropriately. Final fitness value is computed for all the facilities (from facility Fi=0 to n) for all the resources (from resource Rj=0 to m) together as shown in the equation below.

$$Final\ risk\ value = \sum_{i=0}^{n}\sum_{j=0}^{m} |R_j(allocated) - R_j(Demanded)|_i * Area * Risk * population$$

Where, the risk is obtained from the ensemble model for each district. The higher the predicted dengue incidence for a given district the higher the risk is. The predicted value is obtained for a future feature set containing future rainfall, future temperature and population.

4.8.11 Proposed Concept of Locked Genes

In the real-world settings, the possibility of having more of a certain resource than the total demand from all the facilities is not rare. This will bring up a concern that the amount of time we spend on mutating on that resource is a waste of computational time. We can improve the performance of the GA by eliminating these resources in the iterations. I proposed a concept of locked genes in to the standard GA in which genes that has more resources than requested are locked. The GA does not mutate the locked genes. There will be a in-memory locked chromosome as a reference for GA to use in each iteration. Sample lock chromosome is given in Figure 4.14. The sample chromosome is generated for two facilities requesting five resources. The first resource is abundant and hence needed to be locked. Other four resources are not abundant and hence needed to be allocated efficiently via GA.

1	0	0	0	0	1	0	0	0	0

Figure 4.14 The Lock Chromosome for two facilities with five resources. First resource is abundant and hence locked.

4.8.12 Proposed Sliding Mutation Scheme

The standard GA uses random mutation and crossover operations. This will inherently produce random children and hence the fitness of the generated child will be varying. The algorithm selects chromosomes randomly to apply the two operations. This will lead to lower the fitness of the selected best fit chromosome as well. To reduce the fluctuation of fitness, I introduce a new method of mutation at each iteration. The proposed method will make sure that the next generation will be generated based on a sliding mutation scheme where the amount of mutation applied to a given gene is proportional to the risk associated with the chromosome. This method will mutate chromosomes with an amount that proportional to the risk associated with the chromosome. This will drive each chromosome gradually towards the best fit. The cross over operation will ensure that the GA will not end up in a local minimum. The experimental results proved that the proposed scheme perform very well and resulted in a higher bet fit value than standard GA in all the trials that performed. The proposed sliding mutation is given in equation below.

$$\mathrm{D}istance(Allocated) = \sum_{i=1}^{n} |Target(i) - Allocated(i)|$$

$$Remaining\ Percentage = \frac{\sum_{i=1}^{n} Target(i) - Distance(Allocated)}{\sum_{i=1}^{n} Target(i)} * 100$$

$$Amount\ of\ Mutaton(i) = random\big(0, (Allocated(i) * Remainig\ Percentage)\big) \text{ for i} = 1 \ldots \mathrm{n}$$

The Figure 4.15 shows the pseudo codes of the proposed GA with modifications and standard GA. Added steps are bolded in the proposed GA.

Algorithm 1: Standard GA

1: k ← 0;
2: PK ← InitPopulation(n) {Generating Initial Population with n individuals}
3: Compute fitness(i) for each i ∈ Pk; {Evaluate everyone in the population Pk: }
4: While not termination **do**
5: sort(Pk) {sorting the population in descending order}
6: Select (1 − ChurnEntropy) × n members of Pk and insert into Pk+1; { Select Subset of PK:}
7: Select ChurnEntropy × n members of Pk; pair them up; produce offspring; insert the offspring into Pk+1; { Crossover:}
8: Select MutationEntropy × n members of Pk+1; invert a randomly-selected bit in each; {Mutate}
9: Compute fitness(i) for each i ∈ Pk; { Evaluate Pk+1:}
10: k ← k + 1;
11: end while
12: return the fittest individual from Pk;

Algorithm 2: Proposed GA

1: k ← 0;
2: PK ← InitPopulation(n, **resourcesMap**) {Generating Initial Population with n individuals and constraints}
3: Compute fitness(I, facilityInfo) for each i ∈ Pk; {Evaluate everyone in the population Pk: }
4: While not termination **do**
5: sort(Pk) {sorting the population in descending order}
6: Select (1 − ChurnEntropy) × n members of Pk and insert into Pk+1; { Select Subset of PK:}
7: Select ChurnEntropy × n members of Pk; **generate crossover with constraints**; insert the offspring into Pk+1; { Crossover:}
8: Select MutationEntropy × n members of Pk+1; **perform constrained and sliding mutation based on risk**; {Mutate}
applyRiskBasedPenalty for each individual; {applying over and under allocation penalty}
9: Compute fitness(i) for each i ∈ Pk; { Evaluate Pk+1:}
10: k ← k + 1;
11: end while
12: return the fittest individual from Pk;

Figure 4.15 Standard and Proposed GA for resource allocation

4.8.13 Time and Space Complexity Analysis of the Proposed GA

4.8.13.1 Space Complexity

Space complexity is a constant for the proposed GA. Let L be the number of genes in the chromosomes, N is the number of chromosomes in the population, and m is the number of generations it produces. As it replaces the existing least fit chromosomes with newly generated best fit chromosomes resulting in the same number of individuals in the population, the population will not grow along with the iterations. Proposed GA uses two times the space required to store the generated population as the same way it does in the standard GA.

Hence, the space complexity of the proposed GA is O(2*L*N) and it is a constant complexity.

4.8.13.2 Time Complexity

It is not an easy task to formulate an exact equation for the calculation of time complexity of the GA. GA uses evolutionary algorithm making it hard to come up with an exact complexity value. But it is possible to formulate an estimated complexity for the GA. In this section, I formulate a complexity estimator for the proposed GA.

Let G_k is the k^{th} generation, N is the number of individuals in each generation, L is the length of the chromosome, O(crossover) is the complexity of the crossover operation, O(fitness) is the complexity of the fitness function which depends on the implementation of the fitness function and O(mutation) is the complexity of the mutation operation. The proposed GA can adapt fitness function and mutation scheme according to the problem domain. And pm and pc are the entropy of mutation and crossover operations. In this study, O(fitness), O(crossover) and O(mutation) are O(L) operations with is the time complexity is a constant.

Hence the total time complexity is the summation of computations required by the fitness function, crossover operation and mutation operation. At each iteration there will be sorting operation performed to get the best fit chromosome. Hence, it is adding to the computation complexity.

Let's consider G_k generation,

Number of fitness calculations (F_k) = NL

Number of crossover operations (C_k) = pc* N

Number of mutation operations (M_k) = pm*N*L

Hence, the number of total computations needed at k^{th} iteration

= $F_k + C_k + M_k$

Time complexity O(GA) = O($\sum_{k=1}^{m}(\mathrm{Fk} + \mathrm{Ck} + \mathrm{Mk})$

O(GA) = O(Nm)

The proposed GA has a constant time complexity as the standard GA does.

CHAPTER 5

RESULTS:

5.1 Pre-analysis of Data

I conducted correlation analysis of the dataset used in this study. It is vital to have a strong correlation between dependent and independent variables to build a better SVR model. The initial correlation analysis did not show a strong correlation between rainfall and dengue incidence. The literature strongly pointed out that there is a strong correlation between rainfall and dengue incidence. This finding strongly suggested to have a preprocessing of data to eliminate noise in dataset which is the main influencing factor to have a very low correlation value. And also, I had to find out that is there any other factors affecting the correlation between rainfall and dengue incidence. Therefore, I performed several data preprocessing steps to eliminate external influences on rainfall and dengue incidence. The initial correlation analysis is given in the Figure 5.1. The correlation for the global model containing all 76 districts is 0.523.

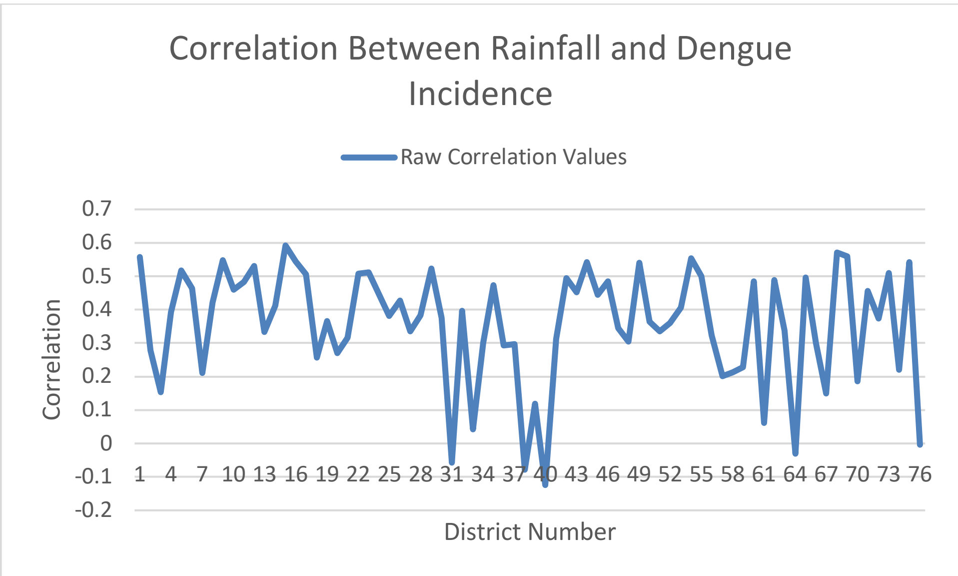


Figure 5.1 Correlation between Rainfall and Dengue Incidence for Raw Data

5.2 Preprocessing of Data

5.2.1 Data Normalization

Initial correlation analysis did not produce a desirable result. There were several external influencing factors contributing to the low correlation values. The main reason for that to happen was the effect of year specific event happened during the period of data collection. I used a dataset which combines multiple years' worth dengue incidences and weather data. There is a possibility of having a special event impacts such as flooding on reported data and hence the effect can be elevated or declined. To minimize the effect of special event in individual year, I normalize the data set year wise. The improvement is visible and reflected well in the correlation analysis of the dataset before and after normalizing. The complete correlation analysis for all districts with and without data normalization is given in Figure 5.2.

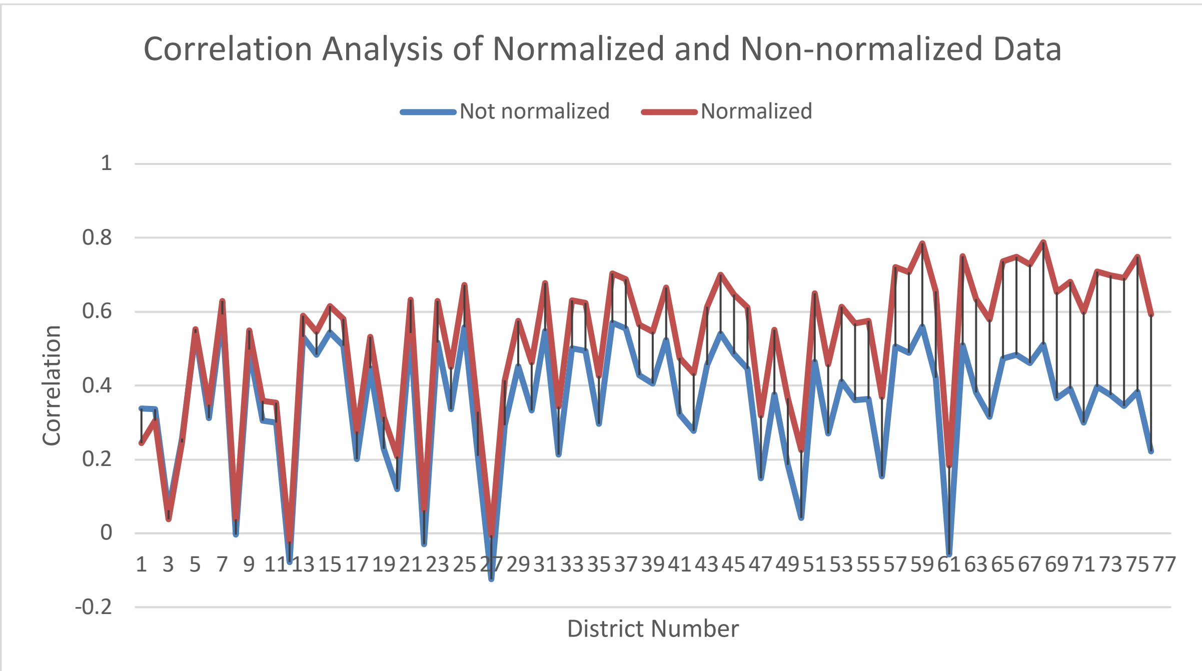
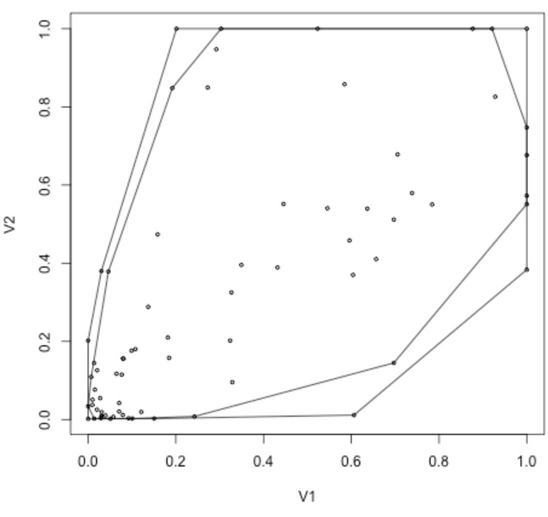


Figure 5.2 Correlation comparison with and without Normalization

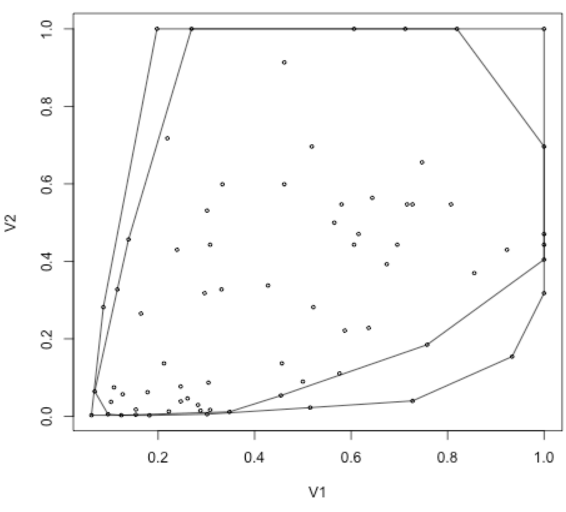
5.2.2 Outlier Removal with Convex Hull Iterative Approach

I suggested to use an iterative outlier removal method. The experimental analysis revealed that each level of outlier removal process increased the correlation between rainfall and dengue incidence. The concept behind the approach is to eliminate outliers and hence increase the correlation. The following set of figures show three levels of outlier removal process and their corresponding correlation values. Figure 5.3 shows set of districts undergone outlier removal and Figure 5.4 shows correlation values for each district in each level of removals.

Amnat Charoen



Ang Thong



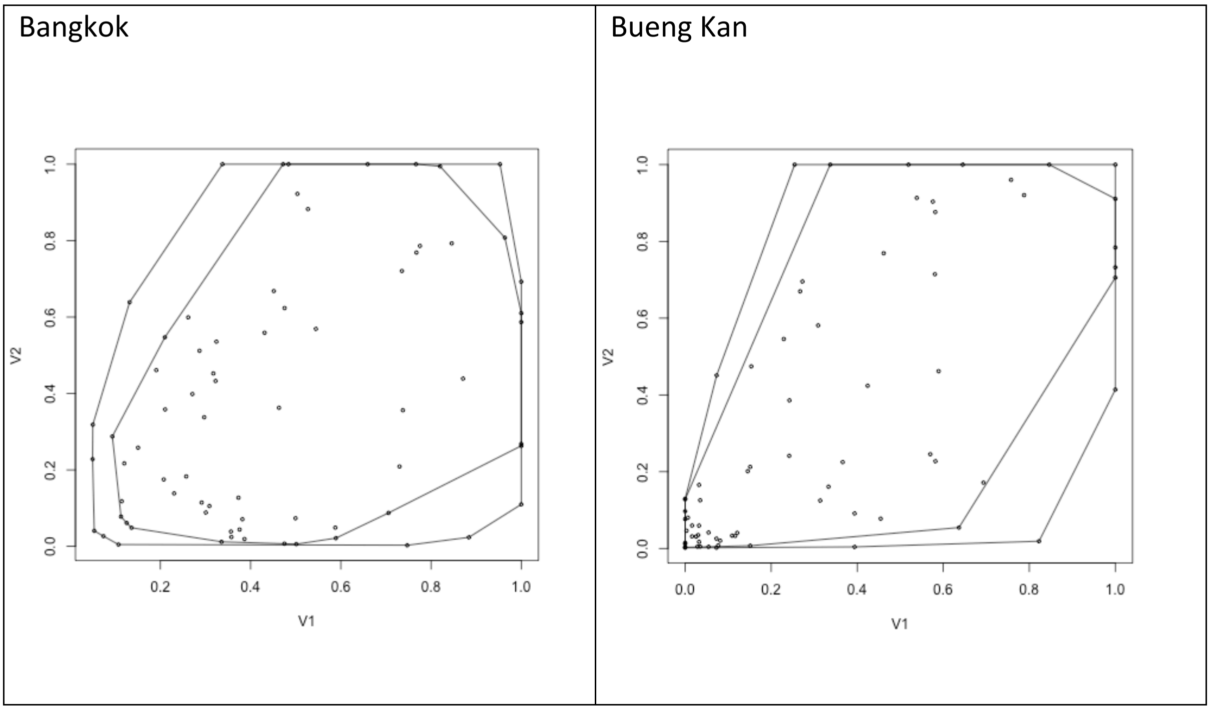


Figure 5.3 Multi Level Outlier Removal with Convex Hulls (v1-rainfall, v2-dengue incidence)

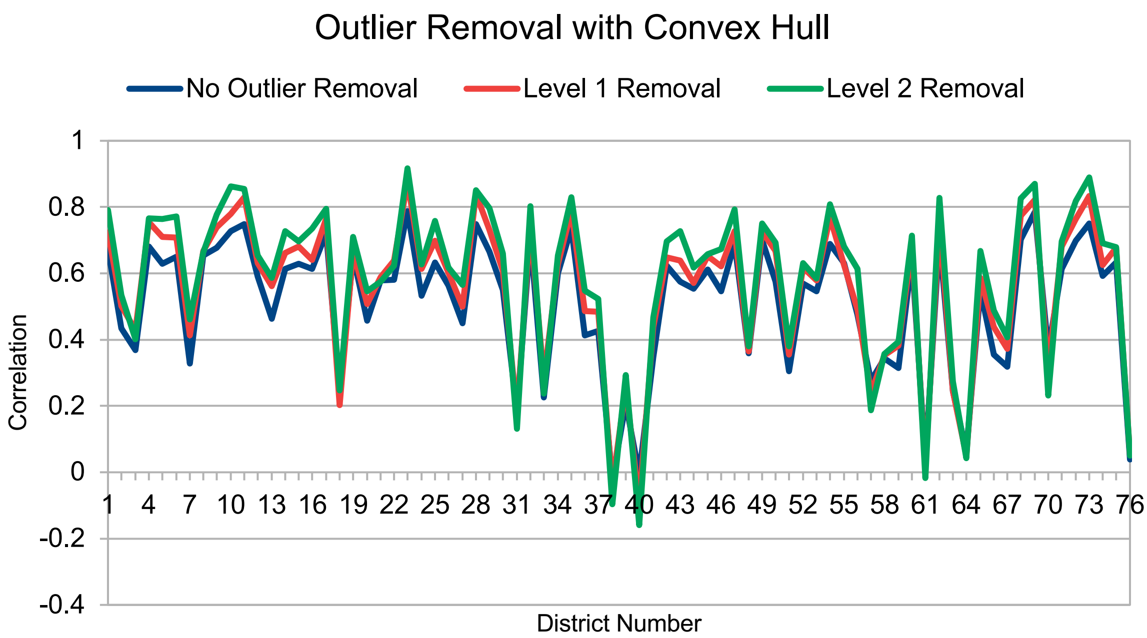


Figure 5.4 Correlation for all Districts at each Outlier Removal Level

It is clear from the Figure 5.4 that correlation between rainfall and dengue incidence is increased with the outlier removal. For a handful of districts, the correlation is not improving as expected. There can be other influencing factors or reporting errors for those districts. For most of the districts there is a good or considerable improvement which is helping SVR to perform better.

The remainder of the results are organized as follows. Spatial non-stationarity behavior of the dengue epidemic is studied in Sri Lanka and results are presented. The generation of prediction models was done for Thailand and results are presented. As per the fact that there is no real data available for resources, a thorough testing was conducted on synthetic data for resource allocations with the proposed Genetic Algorithm approach.

I conducted a thorough analysis I Sri Lanka which gave a clear understanding on how the dengue epidemic is geographically varying. The study conducted in Sri Lanka revealed the usage of district specific prediction models will outperform global prediction models. The results are listed in the following section.

5.3 Results of GWR and Least Square Analysis of Dengue Epidemics in Sri Lanka

5.3.1 Least Square Analysis

In OLS results VIF values shows whether the predictor variables are multicollinear. VIF < 10 means the variables are not multicollinear. In this study, rainfall and population density was used as explanatory variables. VIF value determined these two factors are not correlated and hence can be used in regression analysis together. Every explanatory variable used is unique and contributing to the variation in dengue incidence. OLS regression result also shows that the Adjusted R-Squared value is 0.332054 for the year 2014. This indicates the model built with a combination of population density and rainfall data explains 33.2% of the variation in dengue incidences. According to the OLS regression results, all explanatory variables (rainfall and population density) are statistically significant but the value for

Jarque-Bera statistics is also significant. Significance in Jarque-Bera statistics indicates the model is biased and hence undesirable. Also, the Koenker test is statistically significant (P value < 0.01 for both rainfall and population). This implies non-stationary relationship between the dependent and some or all of the explanatory variables. That reveals the explanatory variables (rainfall and population density) behave differently in different spatial regions.

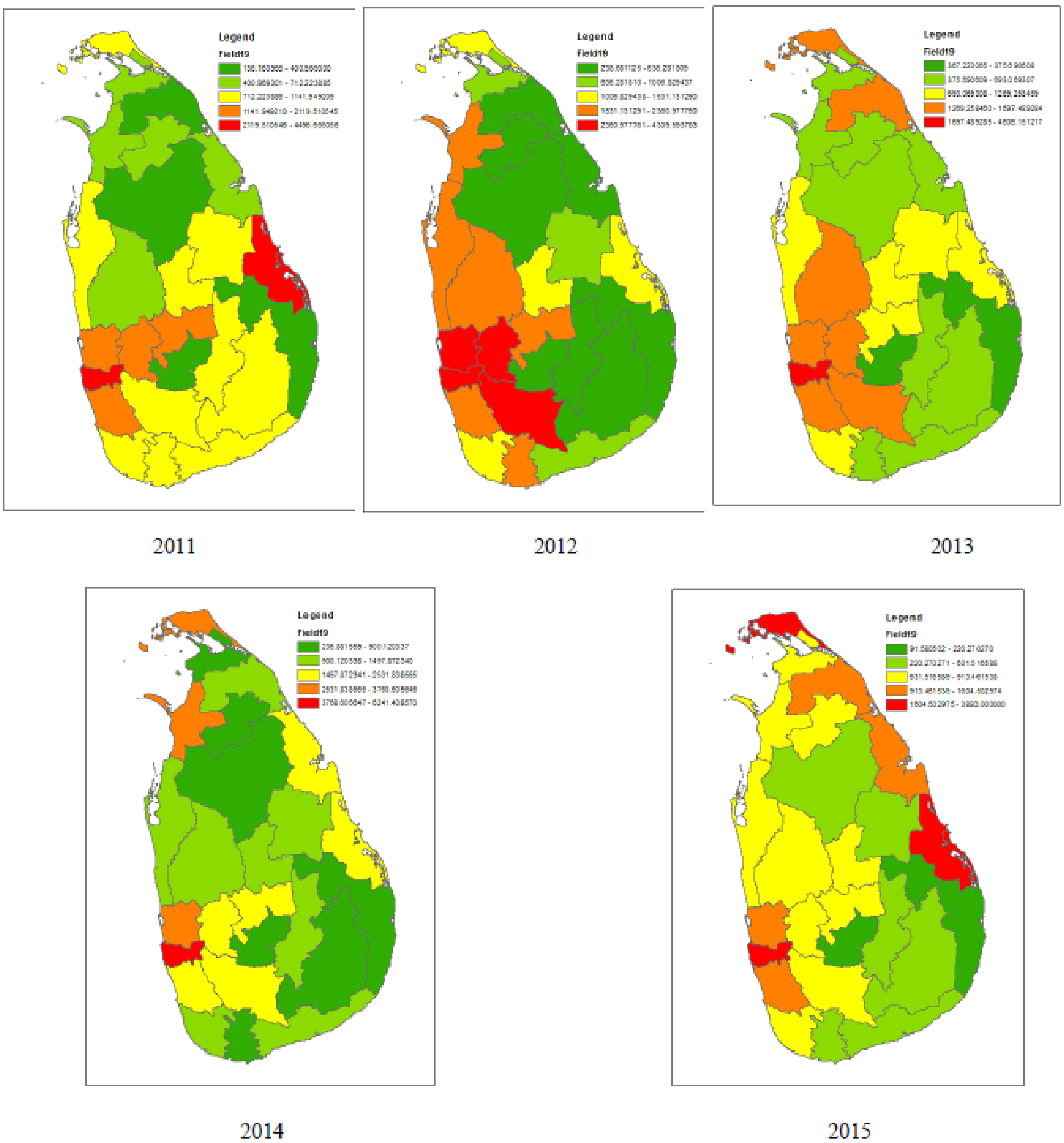


Figure 5.5 The spatial distribution of dengue incidence from 2011 to 2015 in Sri Lanka.

5.3.2 GWR analysis

For GWR, adaptive kernel was used as kernel type and AIC was selected as bandwidth. In GWR AIC value determines the performance of the model. AIC can be used to compare two different models generated with regression analyses. AIC value for OLS is greater than in GWR. Hence GWR is a better analysis tool for dengue incidence with rainfall and population density as explanatory variables.

The GWR model results show that the Adjusted R-Squared values is 0.5632 (R2= 0.621). This indicates the model generated with population density and rainfall as explanatory variables can explain 56.3% of the variance in dengue incidences in 2014. These results also reveal that there are other variables besides population density and rainfall data that has stronger relationships with dengue incidence. These variables are not included in the model. But the model cannot provide a clue of those variables so that they cannot be included in the model. They have to be identified by experimenting with various candidate explanatory variables.

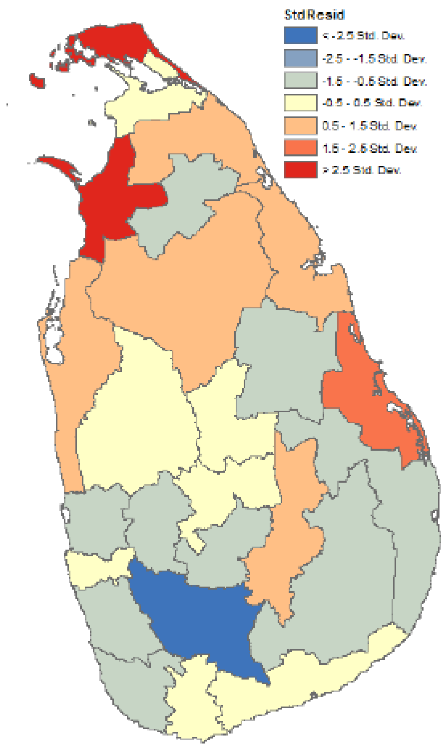


Figure 5.6 The GWR standard residual map for dengue incidence with rainfall and population density for the year 2014.

The standard residual map for the model developed for dengue incidence is shown in the Figure 5.6. The red areas indicate under predictions where the actual number of dengue cases is higher than the model predicted values. The blue areas indicate over predictions where actual dengue cases are lower than predicted values. Random locations of red or blue areas indicate the model performs fairly well. Red or blue clustered areas indicate under/over prediction of the model and hence the performance of the model is poor. Spatial clustering of over/under prediction indicates missing one or more key explanatory variables in the model. The standard residual map in Figure 5.6 shows clustered over and under predicted areas.

It is required to find how well each explanatory variable predicts the dengue incidence for each administrative region. It is revealed from previous sections that there are no global explanatory variables that hold consistent relationship across administrative regions. An analysis was conducted to reveal the variation in strength of explanatory variables in each administrative region in explaining the relationship between the variable and dengue incidence. Results of the analysis are shown in Figure 5.7 (a) and (b). Figure 5.7 (a) provides the spatial distribution of regression coefficients for rainfall and Figure 5.7 (b) provides the same for population density. Lighter colors represent lower coefficients and darker colors represent higher coefficients.

Mapping these coefficients shows the relationship between each explanatory variable and the dependent variable that how they change across the study area. The darker areas in figures indicate the explanatory variables, rainfall and population density, are strong predictors of the dengue incidence, whereas, the lighter areas are locations where they are comparatively weak.

GWR regression results show that relationship of incidence with rainfall and population density is spatially varying across districts of Sri Lanka. Figure 5.7 (a) shows that spatial distribution of regression coefficient of population density is a strong predictor in eastern coastal areas in Trincomalee district, and a weak predictor in Mannar. Figure 5.7 (b) also shows that spatial distribution of regression coefficient of rainfall is a strong predictor in northern areas including Mannar and in eastern coast it is a weak predictor. There is an inverse effect of rainfall and population density on dengue incidence. When rain becomes a strong predictor in some areas, population density is a weak predictor and vice versa. It is very important to understand this variation for making local policies to mitigate dengue. GWR model can also be used to predict values of dependent variables for locations within the study area with unseen explanatory variables values. This will give an estimate of the dependent variable (dengue incidence) using the regression model generated.

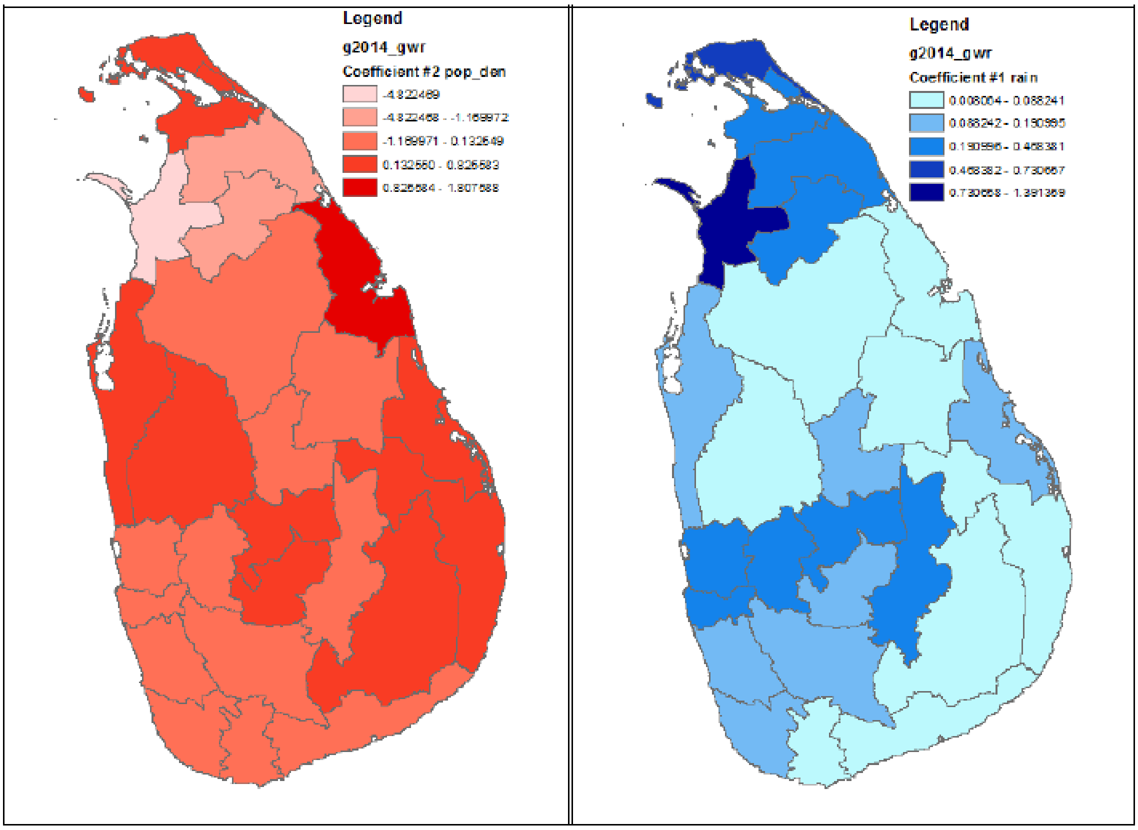
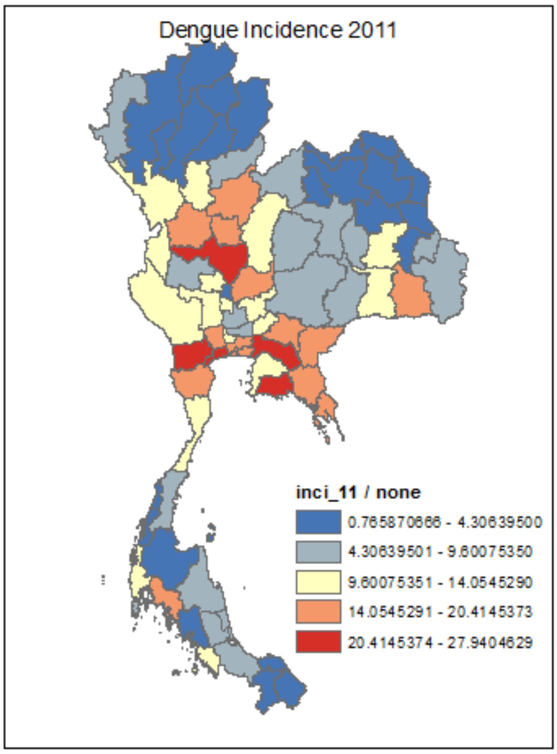


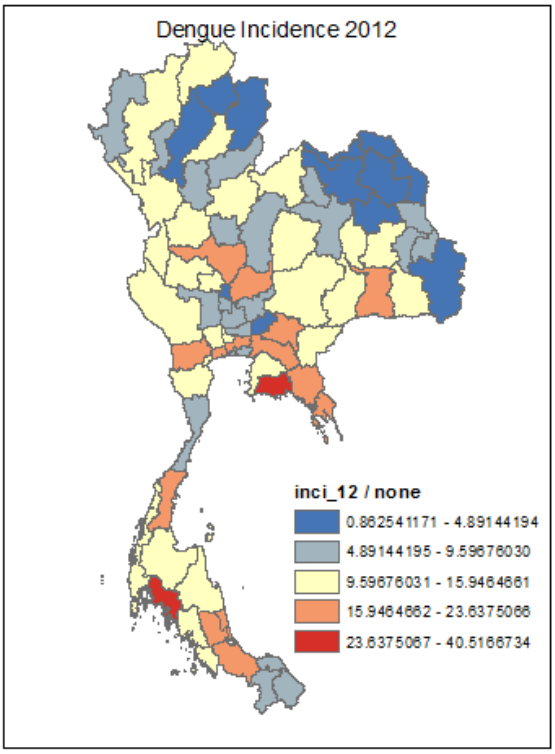
Figure 5.7 The spatial distribution of regression coefficients for (a) population density (b) rainfall.

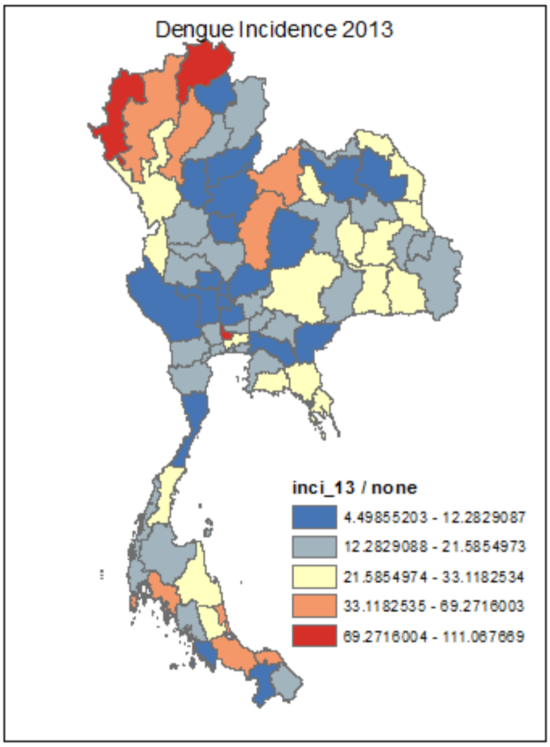
5.4 Generating Prediction Models for Dengue Epidemic in Thailand

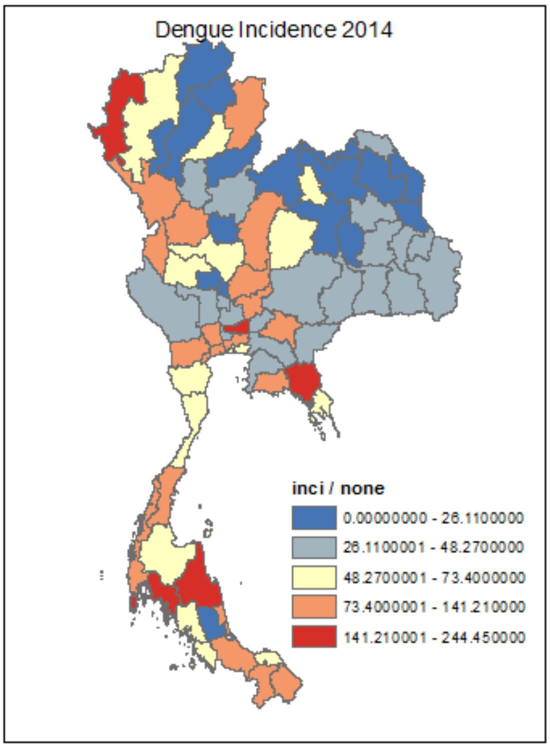
The prediction models for Thailand is generated and evaluated. The results are given in the following sections.

5.4.1 Dengue Incidence in Thailand


Dengue Incidence 2011
inci_11 / none
0.765870666 - 4.30639500
4.30639501 - 9.60075350
9.60075351 - 14.0545290
14.0545291 - 20.4145373
20.4145374 - 27.9404629


Dengue Incidence 2012
inci_12 / none
0.862541171 - 4.89144194
4.89144195 - 9.59676030
9.59676031 - 15.9464661
15.9464662 - 23.6375066
23.6375067 - 40.5166734


Dengue Incidence 2013
inci_13 / none
4.49855203 - 12.2829087
12.2829088 - 21.5854973
21.5854974 - 33.1182534
33.1182535 - 69.2716003
69.2716004 - 111.067669


Dengue Incidence 2014
inci / none
0.00000000 - 26.1100000
26.1100001 - 48.2700000
48.2700001 - 73.4000000
73.4000001 - 141.210000
141.210001 - 244.450000

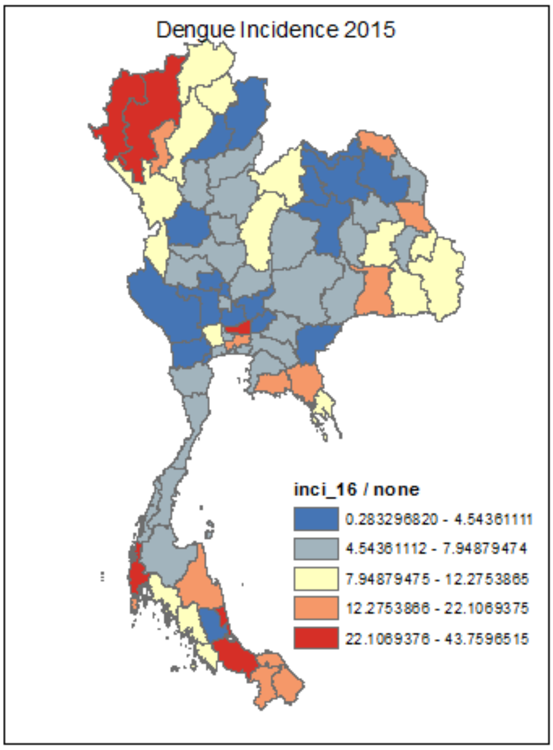


Figure 5.8 Dengue Incidence Map of Thailand from 2011 to 2015

Multiple phases of testing were conducted on the same data set with multiple levels of noise removals. A separate SVR model was trained for each district. Each model was validated using 10-fold cross validation. SVR models are assessed using MSE and proposed accuracy calculation method. Finally, multiple classifiers are combined and generated a micro ensemble to eliminate bias in each classifier for a given district.

5.4.2 Prediction Results for Global Model

Global model contains data from all 76 districts. Each data point is treated the same way as all the other and geographical variations re not considered. There is a single SVR is trained on the entire data set. The plot of rainfall vs. dengue incidence is depicted in Figure 5.9. The results for various scenarios are listed below.

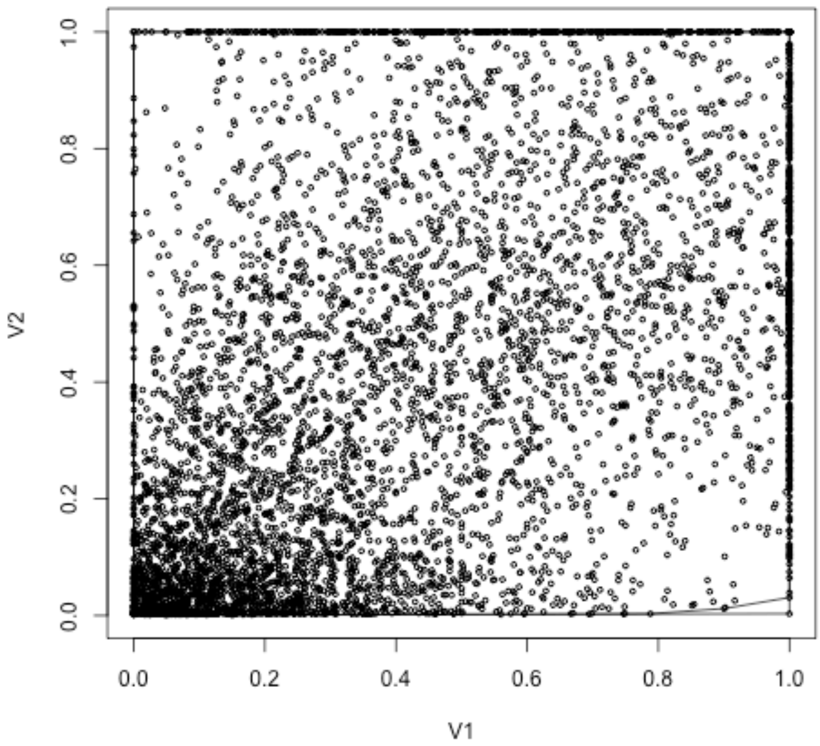


Figure 5.9 Plot of Rainfall vs Dengue Incidence for the Global Dataset (v1-rainfall, v2-dengue incidence)

Correlation of dengue incidence to rainfall for global dataset is 0.523. As per the Figure 5.9, it is clear that the global model has many outliers that make it to have lower correlation coefficient. It is hard to determine an optimal value for the level of outlier removals. Correlation coefficient for 3 levels of outlier removal is given in the Table 5.1. SVR model validation results from 10-fold cross validation with no outlier removal are given in the Table 5.2.

Table 5.1 Correlation Coefficient for Three Levels of Outlier Removals on Global Dataset

Level of outlier removal	Correlation Coefficient
Level 0	0.523
Level 1	0.524
Level 2	0.525

As per the table, it is clear that the increase in correlation coefficient in each level of outlier removal is very small. It is advisable to have many levels of outlier removals to get a higher coefficient for correlation. There is no means by which the optimum number of outlier removals is estimated. The result obtained from the global model using 10-fold cross validation is listed in Table 5.2. Note that the accuracy is calculated based on the method proposed in this study.

Table 5.2 Results of 10-fold Cross Validation of SVR on Global Dataset

Fold	**Accuracy**	**MSE**
1	63.25	0.19322216
2	56.85	0.20835374
3	59.23	0.20811646
4	55.94	0.21453015
5	63.07	0.19367864
6	59.41	0.20107433
7	58.68	0.20458908
8	60.51	0.2005876
9	60.14	0.20600364
10	59.04	0.20545753
Average	59.61	0.20356133

5.4.3 Prediction Results for Local Models

Each district is considered as a separate entity and having unique characteristics. Hence, the models generated for individual district is specific to that particular district. The results of local models trained for 76 districts are given below.

There are three models generated for the same data set and validated based on 10-fold cross validation. The best models are then used in ensemble generation and achieve better output from the ensemble model. The ensemble model then be used in prediction and risk estimation. The models are generated based on SVR, Least Square and K-NN tools. Performance of each model in each outlier removal level along with the performance of ensemble is given in Figure 5.10, Figure 5.11 and Figure 5.12 for level of outlier removals 0, 1 and 2 respectively. It is very clear from the figures that the accuracy improvement with each level of outlier removal.

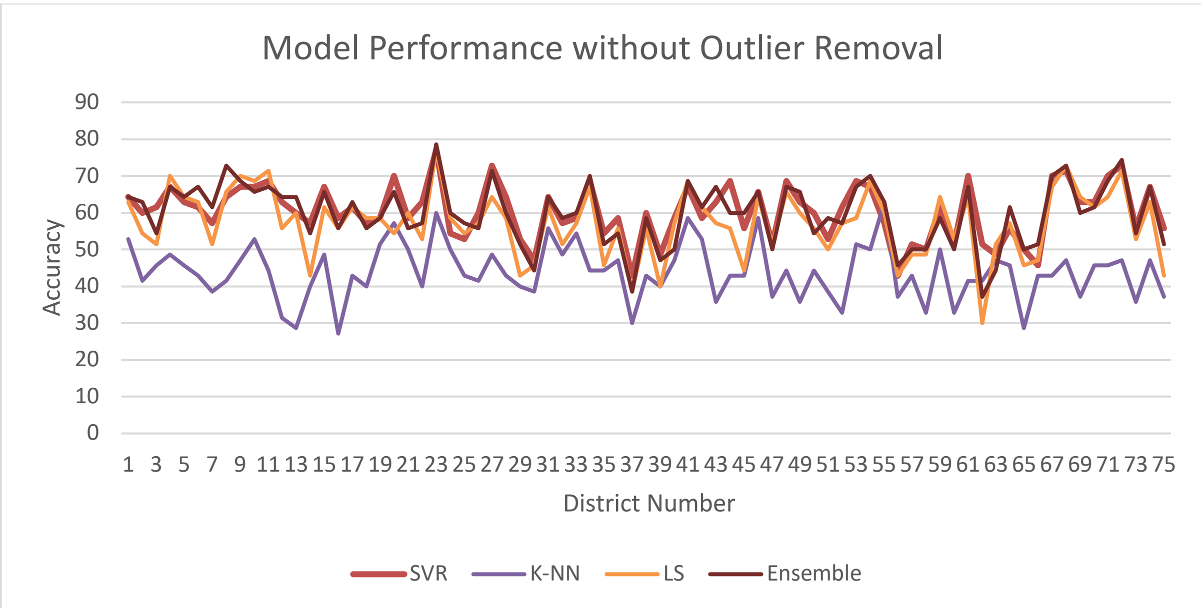


Figure 5.10 The model performance without outlier removal

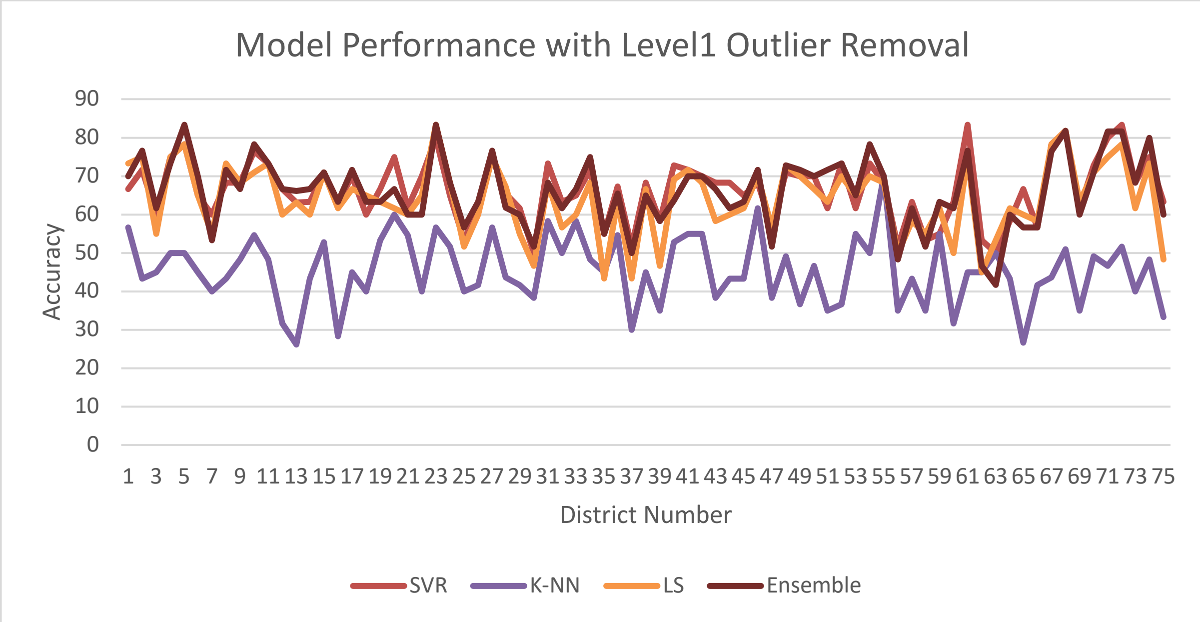


Figure 5.11 The model performance with level1 outlier removal

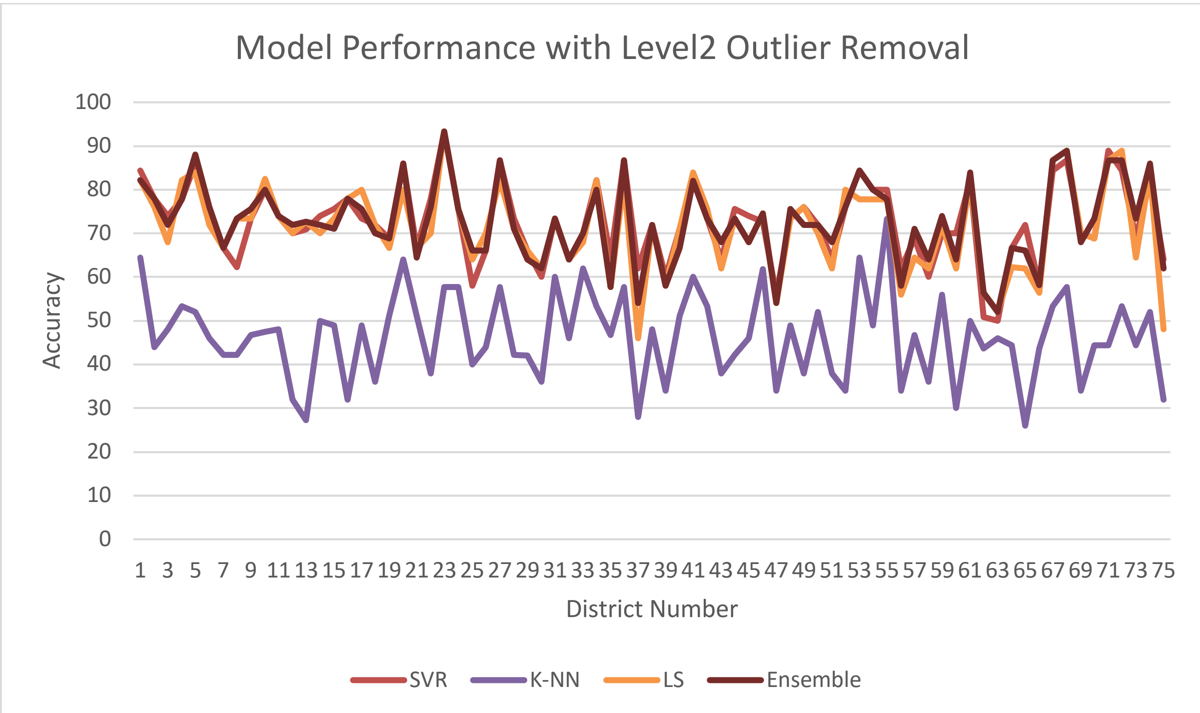


Figure 5.12 The model performance with level2 outlier removal

5.5 Resource Allocation

I performed resource allocation on four different synthetic datasets confirming to the setting listed in the Table 5.3. I listed the complete setup for the Trial 1 and list only best fit value graph and the allocation results for the remaining trials. All the other results are found under appendix A.

Table 5.3 Experimental Trial Setup

Trial No	Number of facilities	Number of Resources	Facility Properties
1	10	10	3
2	50	5	3
3	100	10	3
4	500	10	3

Performance Comparison of GA with random and sliding mutation with lock chromosome

5.5.1 Trial 1

Table 5.4 Facility Information (High risk facility is highlighted)

Population	Risk	Area
293,931	4,435	259
424,873	2,722	619
300,904	1,930	743
222,133	4,372	195
185,651	2,426	328
257,318	2,368	354
453,762	2,720	522
141,650	1,956	493
402,820	1,555	128
259,771	1,944	840

Table 5.5 Resource Availability

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1,304	7,441	7,297	980	6,276	9,691	5,735	6,476	2,754	146

Table 5.6 Requested Resources from each facility

Resource / Facility	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
F1	43	369	362	49	208	481	190	214	91	5
F2	66	564	553	74	317	734	290	327	139	7
F3	55	467	458	61	262	608	240	271	115	6
F4	207	1,770	1,735	233	995	2,305	909	1027	437	23
F5	102	875	858	115	492	1,140	450	508	216	11
F6	207	1,770	1,735	233	995	2,305	909	1,027	437	23
F7	129	1,108	1,087	146	623	1,444	570	643	273	14
F8	164	1,400	1,373	184	787	1,823	719	812	345	18
F9	189	1,614	1,583	213	908	2,102	829	936	398	21
F10	143	1,225	1,201	161	689	1,595	629	711	302	16

Table 5.7 Lock Chromosome

0	0	0	0	1	0	1	1	1	1

Performance of the GA for random, no-sliding with constraints and sliding with constraints GAs. There are 10 facilities with four properties. Each facility is requesting 10 different resources. The performance of each category is shown in Figure 5.13.

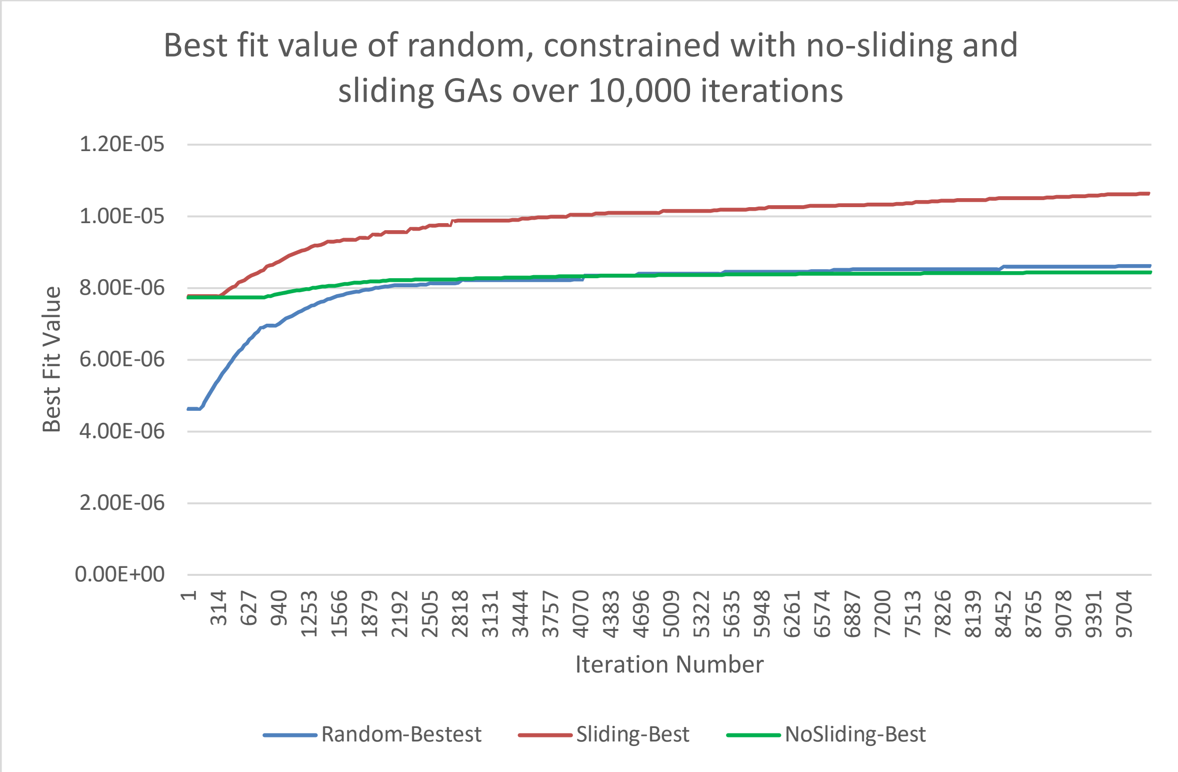
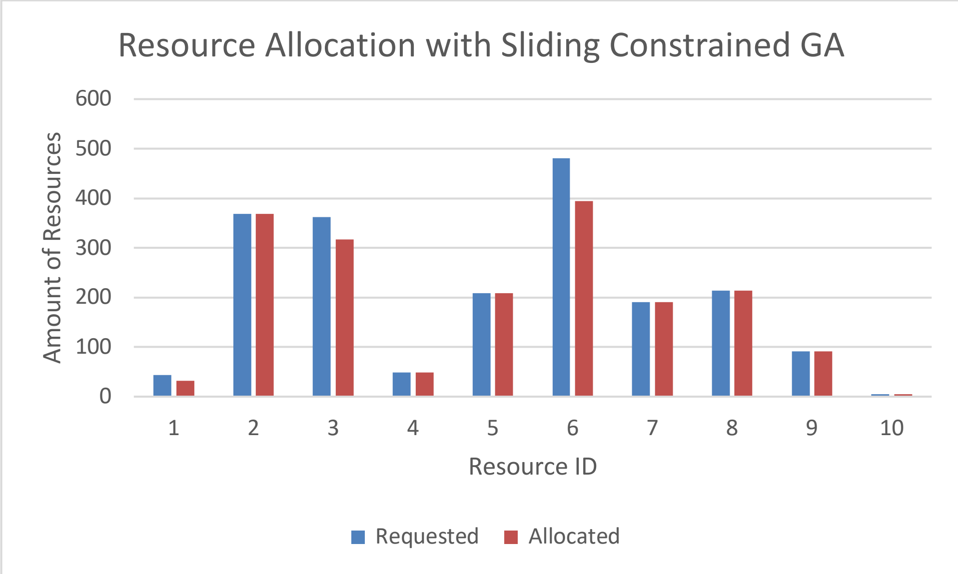


Figure 5.13 The performance of proposed GA for 10 facilities requesting 10 resources

Resource Allocation Results of Trial 1 for high risk facility and low risk facility are given in Figure 5.14 and Figure 5.15.



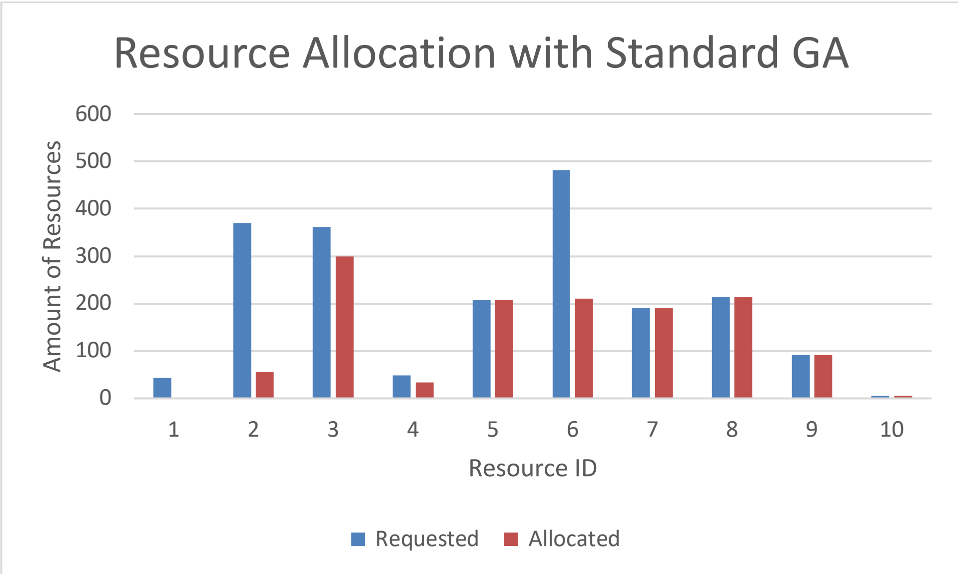
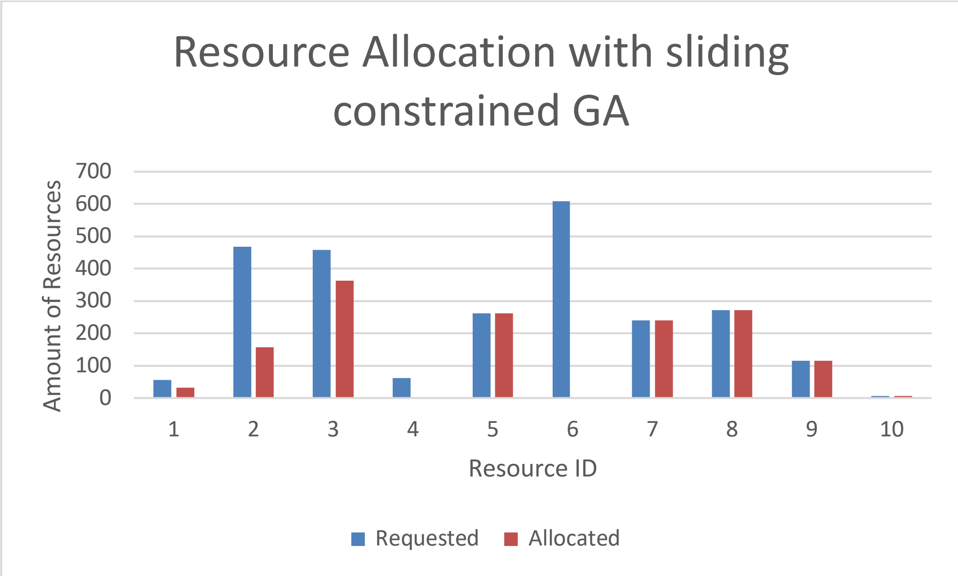


Figure 5.14 High risk facility



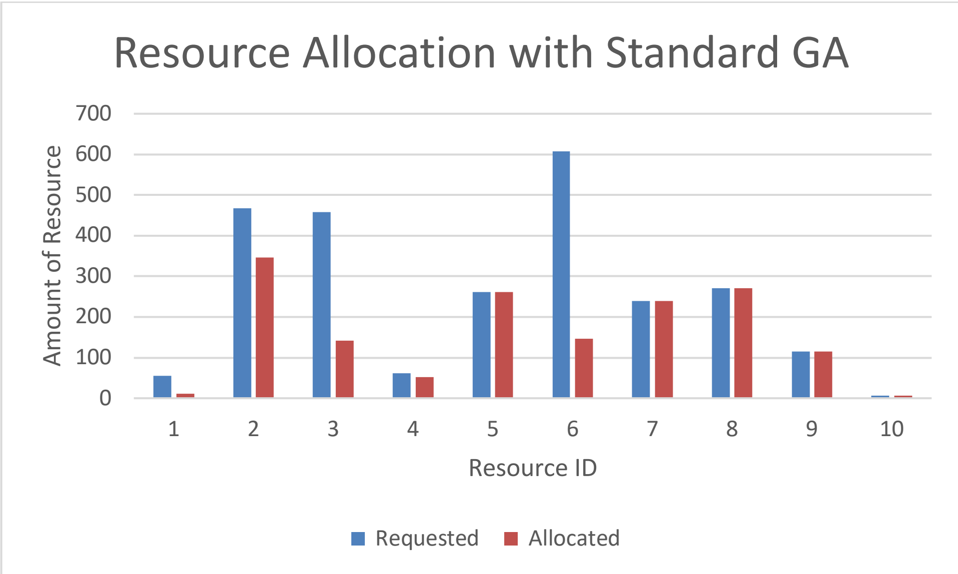


Figure 5.15 Lowest risk facility

5.5.2 Trial 2

The performance of the GA for random, no-sliding with constraints and sliding with constraints Gas for 50 facilities with four properties was tested. Each facility is requesting 5 resources. The performance of the trial 2 is given in Figure 5.16.

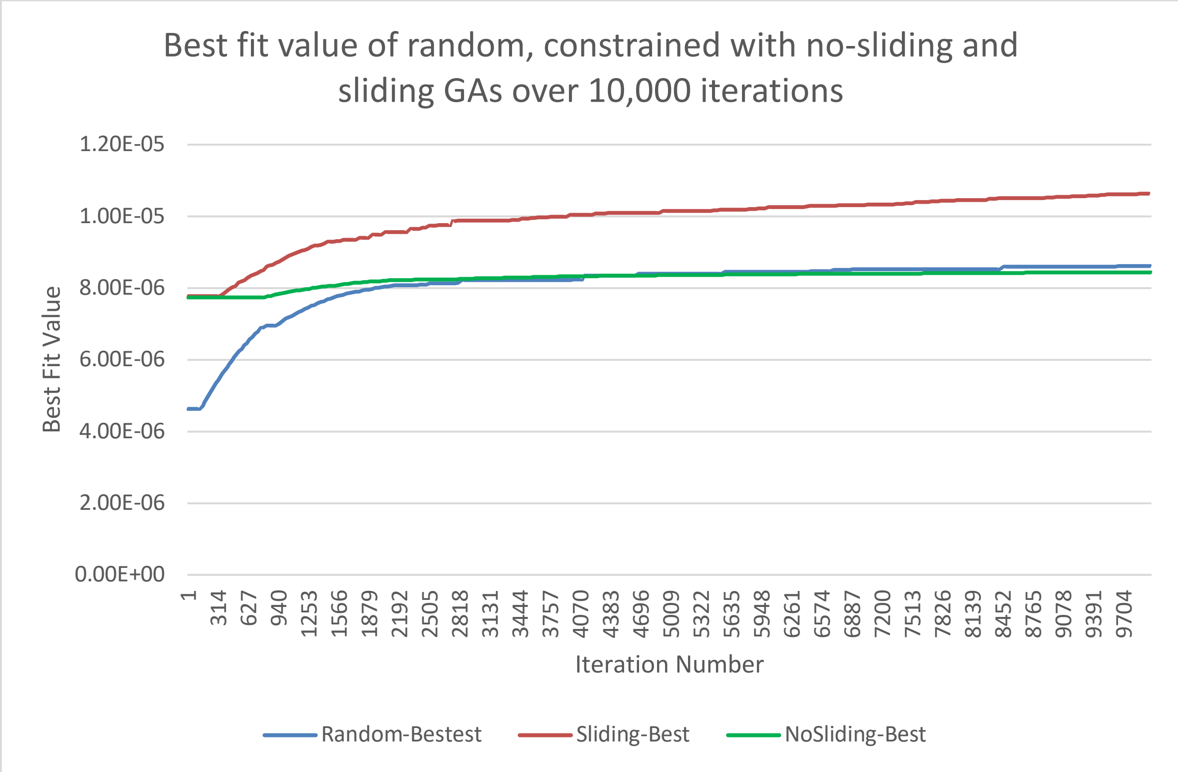
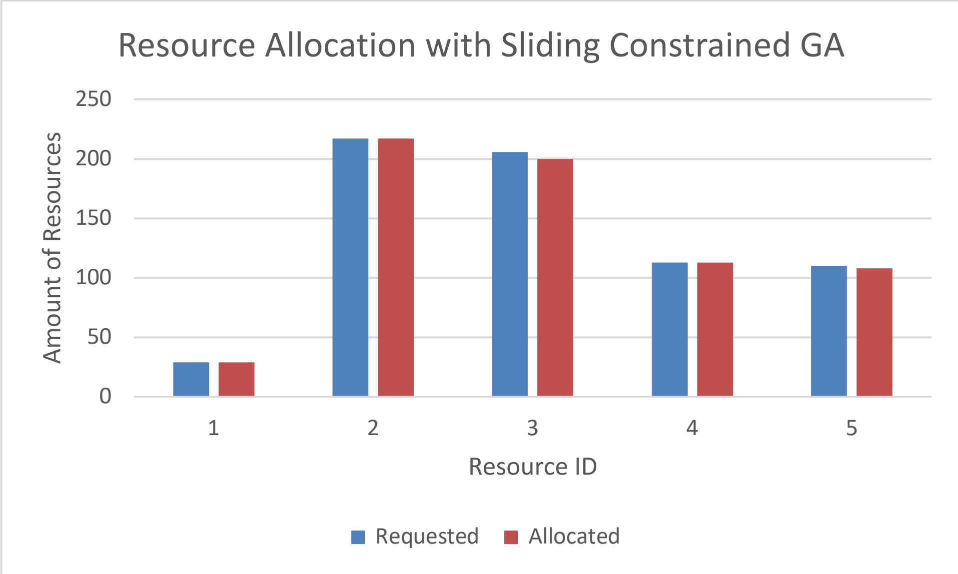


Figure 5.16 Performance of the proposed GA for 50 facilities and 5 resources

The high-risk facility is the facility with id 27. The following graphs (Figure 5.17) show the resource allocation for the facility 27 with random GA and sliding with constrained GA.



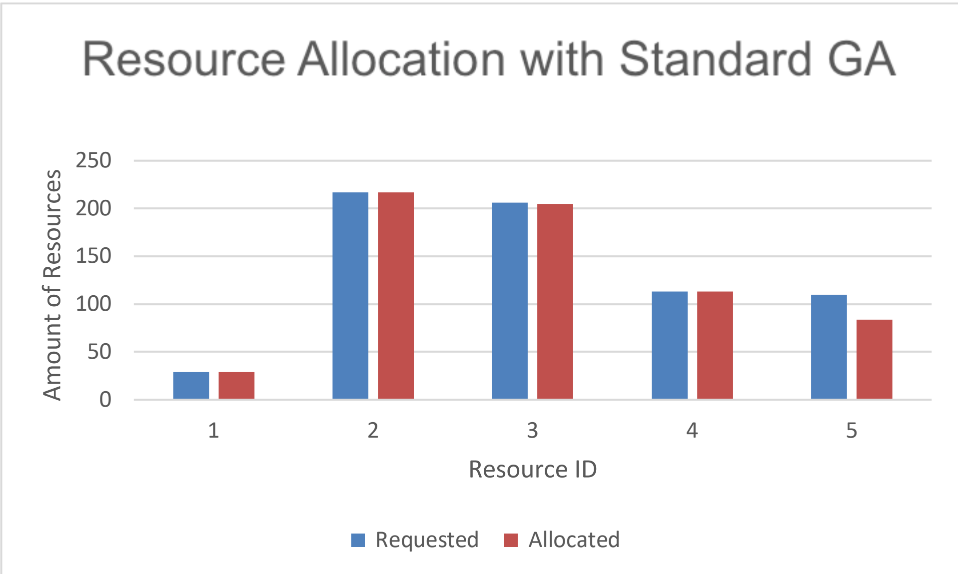


Figure 5.17 The resource allocation for the high-risk facility of Trail 2

5.5.3 Trial 3

The performance of the GA for random, no-sliding with constraints and sliding with constraints GA for 100 facilities with 4 properties that request for 10 resources was tested. Performance is given in Figure 5.18

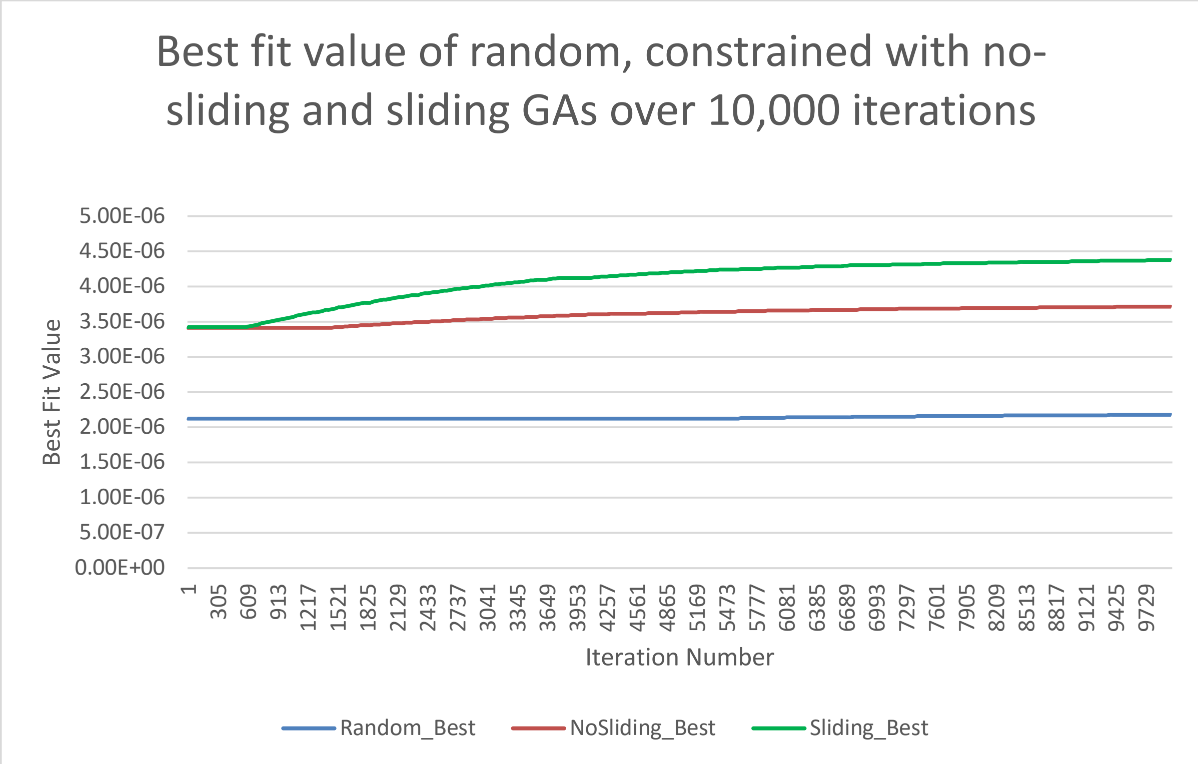


Figure 5.18 The performance of the proposed GA for 100 facilities with 10 resources

The high-risk facility is the facility with id 49. The following graphs (Figure 5.19) shows the resource allocation for the facility 49 with random GA and sliding with constrained GA.

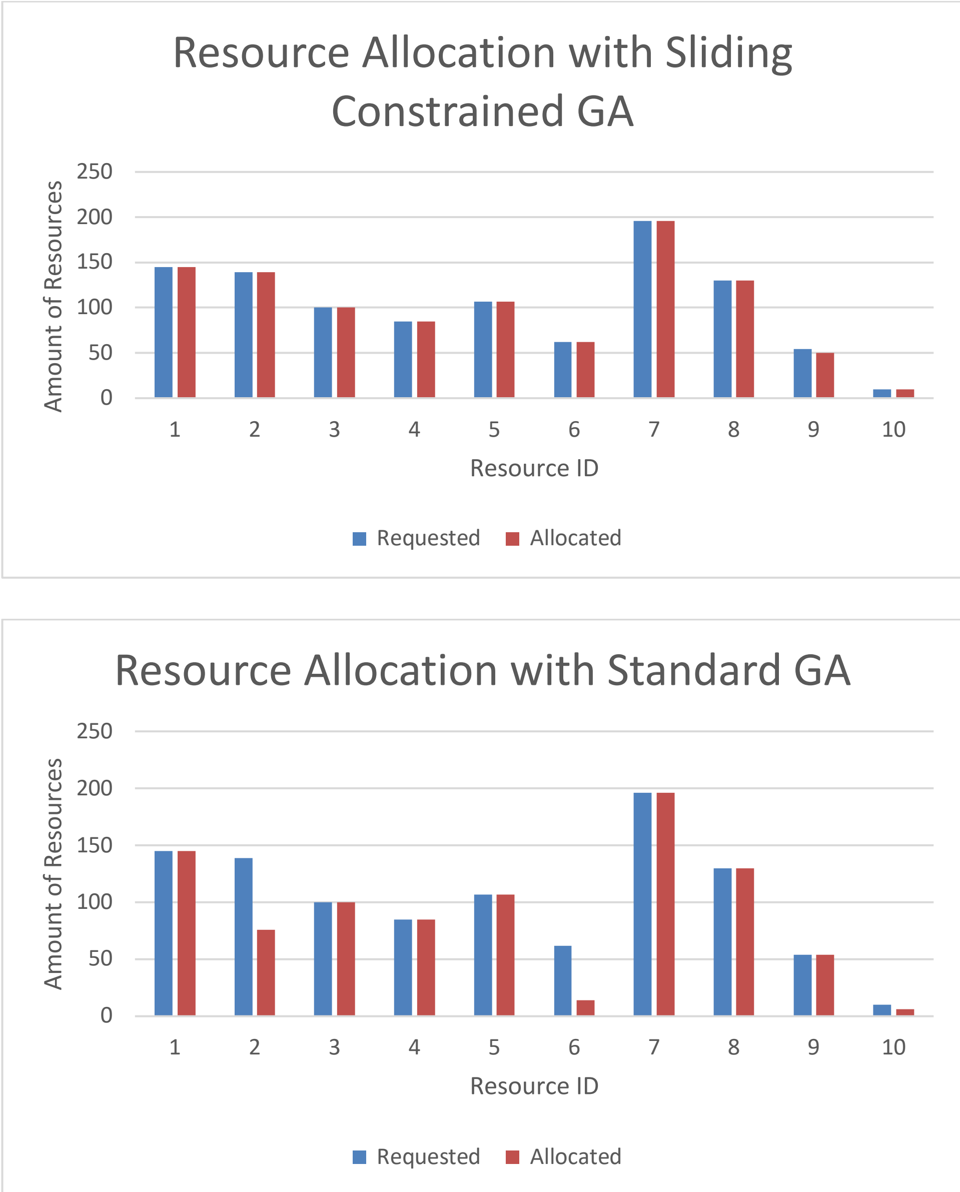


Figure 5.19 The resource allocation for the Trail 3

5.5.4 Trial 4

The performance of the GA for random, no-sliding with constraints and sliding with constraints GA for 500 facilities with 4 properties that request for 10 resources was tested. Performance is given in Figure 5.20.

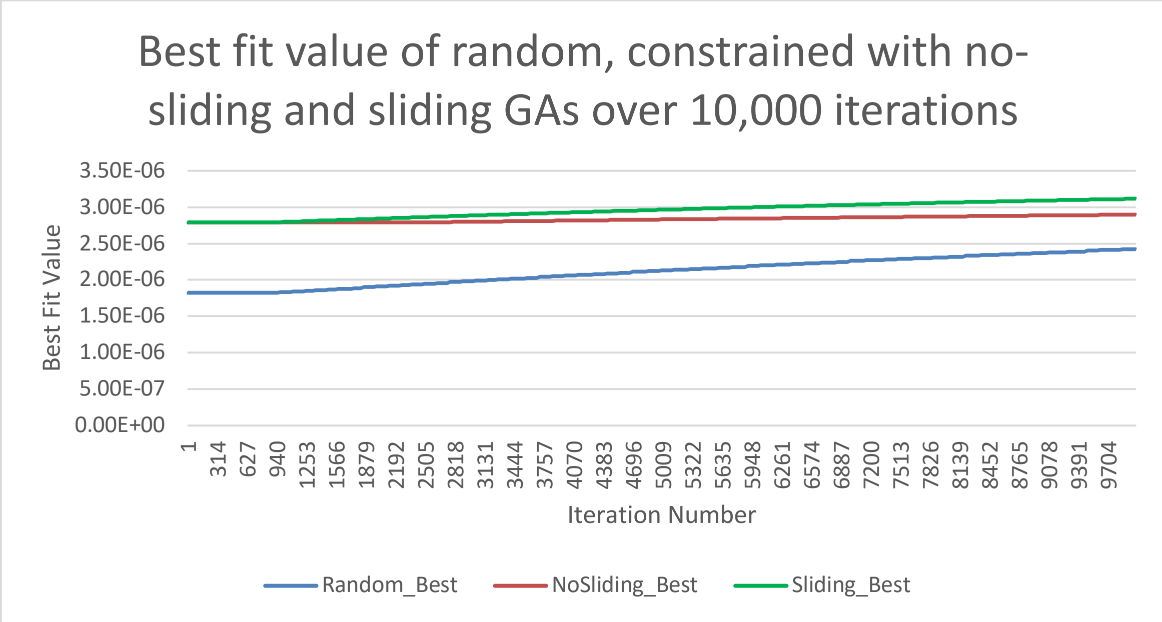
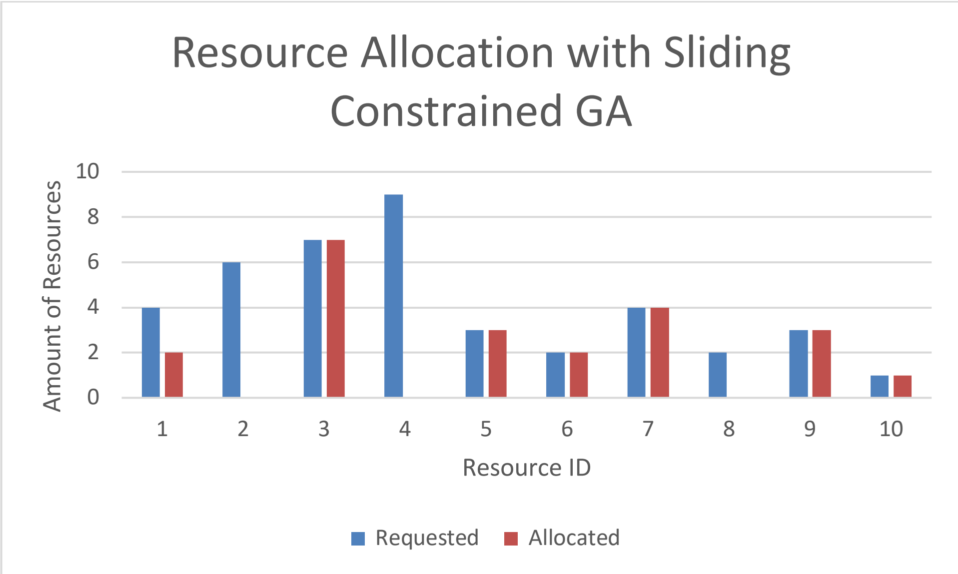


Figure 5.20 The performance of the proposed GA for 500 facilities with 10 resources

The high-risk facility is the facility with id 377. The following graphs (Figure 5.21) show the resource allocation for the facility 377with random GA and sliding with constrained GA.



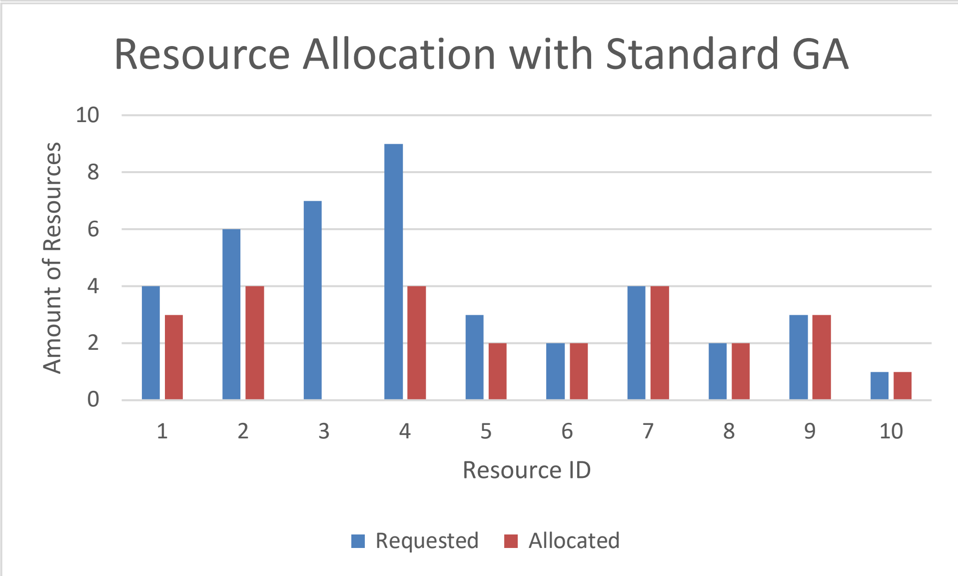
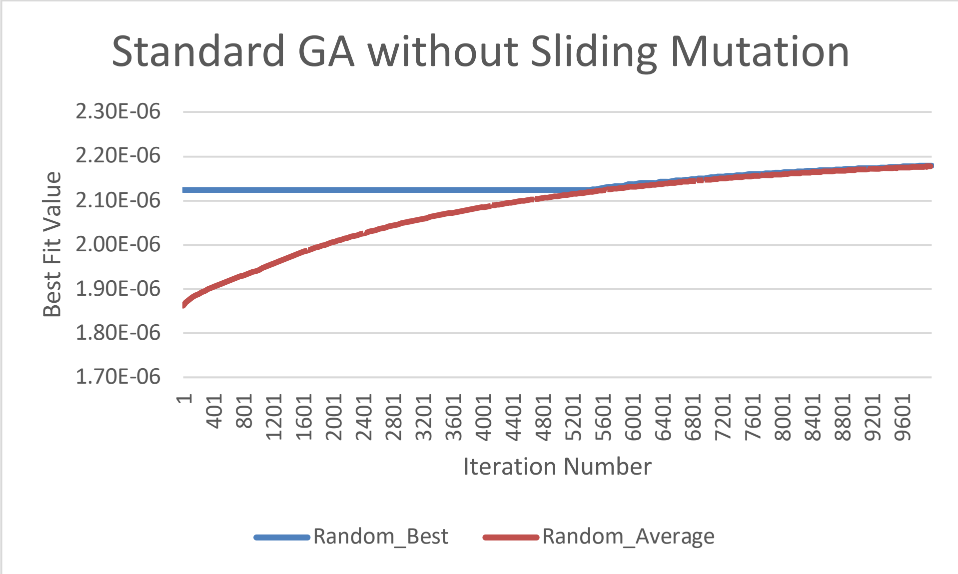


Figure 5.21 The resource allocation for the Trail 4

5.6 Comparison of Proposed GA with Sliding Mutation against Standard GA with Random Allocation and Mutation

The standard GA with random population generation and mutation is always starting with a lower fitness value. The variation in fitness value among each individual is very high giving a very lower value of average fitness value compared to best fit value. In contrast, the proposed GA is always starting with a higher fitness value for both best fit values and average fitness value. The difference between best fitness and average fitness is small for the proposed GA as the constrained based population generation always produces offspring that are closer to the target chromosome. These observations are clearly shown in Figure 5.22.



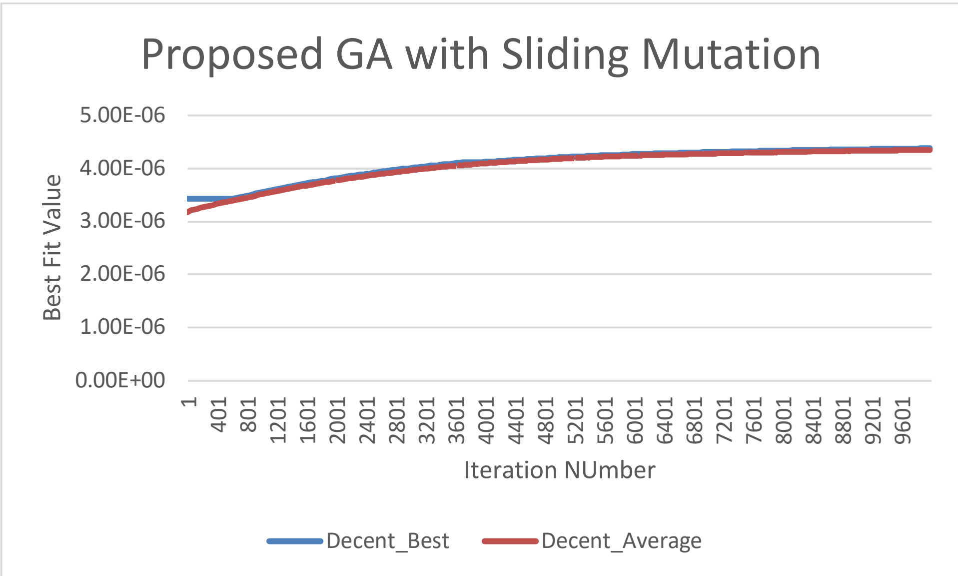


Figure 5.22 Comparison of standard GA and the proposed GA

CHAPTER 6

CONCLUSION

A computational approach for dengue epidemic prediction and mitigation based on the rainfall, temperature and population density was proposed. The required data was collected from various remote sensing data repositories including NASA Earth Observations (NEO). A set of pre-processing steps is applied on the acquired data sets to ensure the quality of the prediction.

The proposed framework introduced a year-wise data normalization was introduced to eliminate the year specific shifts of dengue incidence as the first step of the pre-processing steps. These shifts occur due to special environment event such as flood or droughts. The technique of year-wise data normalization increased the correlation between the rainfall and the dengue incidence. The increase in correlation is above 0.3 for some districts. The increase in coefficient is visible in 73 districts out of 76 districts. There is an at least 0.1 increase in correlation for more than 70% of the districts after the year-wise normalization was applied. There is no increase in the coefficient for the district Saraburi.

Further, a novel method of outlier removal as the second step of the data preprocessing steps. The prediction results of the data sets were increased after the noise removal applied to the original data set. The noises were removed from the data sets with the proposed outlier removal process using iterative convex hull. This technique increased the correlation of the rainfall and the dengue incidence for 73 out of 76 districts. Pathum Thani, Phangnga, and Saraburi districts have very low correlation coefficients and the reason for low correlations is unknown. The correlation coefficients are not increasing even for the 3rd level of outlier removal for those districts. There is an at least 0.1 increase in the coefficient for more than 50% of the districts after the 3rd level of noise removal was applied. In some district, the increase in correlation coefficient is more than 0.2 and it is a significant improvement for a successful model generation.

A novel method of accuracy calculation of regressions was also introduced. The standard methods of evaluating regressions are MSE and adjusted R^2. These measures are affected by the extreme outliers. The proposed method can be used to evaluate the standard regressions with accuracy.

The preprocessing steps increased the correlation coefficients by 0.2 on average. This caused the prediction model to perform better and produced higher accuracies compared to the accuracies generated from the original data sets. The average amount of increase in the accuracy after the preprocessing steps is about 20%.

The behavior of the disease was modeled using an ensemble of regressors to overcome the geographical dependency of the dengue disease. The proposed regressor ensemble was used to predict the dengue epidemic and identified the high-risk areas. This information is then used in the proposed resource allocation method to efficiently allocate limited resources. The proposed resource allocation methodology was modeled based on the state-of-the-art Genetic Algorithm with major modifications to the algorithm. The mutation, crossover and initial population generations are modified to introduce constraints and to match the problem domain. In addition, the proposed GA introduced sliding mutation scheme in which the amount of mutation applied to a gene is determined based on the risk and the current allocation for that gene. This scheme ensures the fast convergence of the GA search. The time and the space complexity of the standard GA remains the same with the proposed GA. The proposed data cleaning, problem modeling, and resource allocation methods are promising and proved by the generated results.

6.1 Future Directions

The proposed research study created the foundation for an automated dengue mitigation system to be built on. I propose a web-based system to manage every aspect of the dengue mitigation from data cleaning to resource allocation. The proposed algorithms will be hosted in a web server and respond to various user queries. In addition, the web-based system will facilitate to enter the dengue incidence data for a given district. Then the system will automatically integrate the new data into the existing algorithm and updates the models to accommodate new information This will ensure that the response planes generated based on the recommendations provided by the web-system are up to date. The web-based system will also be capable of generating risk maps for a given climate and population parameters (prediction). And also, the system will facilitate to run the resource allocation based on the predicted risk map. The resources and the demand of each facility can be entered into the web system through the user interfaces provided.

The mobile devices are getting popular in recent years. The users favor mobile apps over desktop apps as it is portable and accessible at any time. I propose to develop a mobile app (iOS and Android to make it widely available) to acquire recent data from the general public. This will eliminate the need for waiting for government officials to update their data portals. There will be a mechanism to verify the user reported data at the web server. Once verified, the user entered data will be added to the models and start generating risk maps based on the latest data.

BIBLIOGRAPHY

[1]. Sirisena PDNN., Noordeen F. "Evolution of dengue in Sri Lanka—changes in the virus, vector, and climate". International Journal of Infectious Diseases, 19 (2014) 6–12

[2]. Neill II MO, Mikler AR, and Schneider T. An Extensible Software Architecture to Facilitate Disaster Response Planning. 2011.

[3]. Vitarana T, Jayakuru WS. Historical account of dengue haemorrhagic fever in Sri Lanka. WHO/SEARO Dengue Bulletin 1997; 21:117–8.

[4]. Kanakaratne N, Wahala MP, Messer WB, Tissera HA, Shahani A, Abeysinghe N, et al. Severe dengue epidemics in Sri Lanka 2003–2006. Emerg Infect Dis 2009; 15:192–9

[5]. Kusumawathie PH, Fernando WP. Breeding habitats of Aedes aegypti Linnaeus and Ae. albopictus Skuse in a dengue transmission area in Kandy, Sri Lanka. The Ceylon Journal of Medical Science 2003; 46:51–60

[6]. Kusumawathiea PH, Yapabandarab AM, Jayasooriyaa GA, Walisingheca C. Effectiveness of net covers on water storage tanks for the control of dengue vectors in Sri Lanka. J Vector Borne Dis 2009; 46:160–3

[7]. McMichael AJ, Woodruff RE, Hales S. Climate change and human health: present and future risks. Lancet 2006; 367:859–69

[8]. Dorji T, Yoon IK, Holmes EC, Wangchuk S, Tobgay T, Nisalak A, et al. Diversity and origin of dengue virus serotypes 1, 2, and 3, Bhutan. Emerg Infect Dis 2009; 15:1630–2.

[9]. Department of Health Service, Sri Lanka. Annual health bulletin of Sri Lanka. Colombo, Sri Lanka: Department of Health Services; 2002.

[10]. Briet OJ, Galappaththy GN, Konradsen F, Amerasinghe PH, Amerasinghe FP. Maps of the Sri Lanka malaria situation preceding the tsunami and key aspects to be considered in the emergency phase and beyond. Malaria J 2005; 4:8.

[11]. Ministry of Health, Sri Lanka. Surveillance report on dengue fever/dengue haemorrhagic fever 2007. Epidemiological Bulletin 2008; 49:1–20.

[12]. Ministry of Healthcare and Nutrition of Sri Lanka. Monthly distribution of suspected dengue cases from 2004 to 2010 by District in Sri Lanka. Ministry of Healthcare and Nutrition of Sri Lanka, 2011. Available at: http://www.epid.- gov.lk/Den- gue_updates.htm (Accessed on 20th of August 2013).

[13]. Martha A, Yuzo A. Male–female differences in the number of reported incident dengue fever cases in six Asian countries. Western Pac Surveill Response J 2011;2

[14]. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res 2010; 85:328.

[15]. Cavrini F, Gaiban P, Pierro AM, Rossini G, Landini MP, et al. Chikungunya, an emerging and spreading arthropod-borne viral disease. J Infect Dis 2009; 3:744– 52.

[16]. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli AC, et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet 2007; 370:1840–6.

[17]. Scott TW, Morrison AC. Vector dynamics and transmission of dengue virus: implications for dengue surveillance and prevention strategies: vector dynamics and dengue prevention. Curr Top Microbiol Immunol 2010; 338:115.

[18]. Shope RE. Global climate change and infectious diseases. Environ Health Perspect 1991; 96:171–4.

[19]. Ward MA, Burgess NR. Aedes albopictus: a new disease vector for Europe. J R Army Med Corps 1993; 139:109–11.

[20]. Kalra NL, Kaul SM, Rastogi RM. Prevalence of Aedes aegypti and Aedes albopictus vectors of DF/DHF in North, North East and Central India. Dengue Bulletin 1997; 21:84–92.

[21]. Chun L, Telisinghe PU, Hossain MM, Ramasamy R. Vaccine development against dengue and shigellosis and implications for control of the two diseases in Brunei Darussalam. Brunei Darussalam Journal of Health 2007; 2:60–71.

[22]. Ooi EE, Goh KT, Gubler DJ. Dengue prevention and 35 years of vector control in Singapore. Emerg Infect Dis 2006; 12:887–93

[23]. C.W. Lian, C.M. Seng and W.Y. Chai, “Spatial, Environmental and Entomological Risk Factors Analysis on a Rural Dengue Outbreak in Lundu District in Sarawak, Malaysia”, Tropical Biomedicine, vol. 23(1), pp. 85-96, 2006.

[24]. S. B. Seng, A. K. Chong and, A. Moore. “Geostatistical Modelling, Analysis and Mapping of Epidemiology of Dengue Fever in Johor State, Malaysia”, The 17th Annual Colloquium of the Spatial Information Research Centre University of Otago, Dunedin, New Zealand, 2005.

[25]. N. Rachata, P. Charoenkwan, T. Yooyativong, K., Chamnongthai C. Lursinsap and K. Higuchi, “Automatic Prediction System of Dengue Haemorrhagic-Fever Outbreak Risk by Using Entropy and Artificial Neural Network”. International Symposium on Communications and Information Technologies (ISCIT 2008), pp. 210-214, 2008.

[26]. T. Van Gestel, J.A.K. Suykens, D.E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, and J. Vandewalle, Financial time series prediction using least squares support vector machines within the evidence framework, IEEE Trans. On Neural Networks, vol. 12. Issue 4, pp. 809-821, (2001).

[27]. D. Mckay and C. Fyfe, Probability prediction using support vector machines, In Proceedings of Int. Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, vol. 1, pp. 189-192, (2000).

[28]. A. Fan and M. Palaniswami, Selecting bankruptcy predictors using a support vector machine approach, vol. 6, pp. 354-359, (2000).

[29]. F. Tay and L.J. Cao, Modified support vector machines in financial time series forecasting, Neurocomputing, Sept., (2001).

[30]. Soemsap, T.; Wongthanavasu, S.; Satimai, W., "Forecasting number of dengue patients using cellular automata model," Electrical Engineering Congress (iEECON), 2014 International, vol., no., pp.1,4, 19-21 March 2014

[31]. E. Ahmed, H. N. Agiza and S. Z. Hassan, "On Modelling Hepatitis B Transmission Using Celular Automata, " Journal of Statistical Physics, vol. 92, No.3/4, pp.707-712, 1998.

[32]. Hani M. Aburas, B. Gultekin Cetiner and Murat Sari., "Dengue confirmed-cases prediction: A neural network model, " Expert Systems with Applications, vol. 37, 2010, pp.4256-4260.

[33]. L. H. A. Monteiro, D. N. Oliveira. and J. G.Chaui-Berlinck, "The effect of spatial scale on predicting time series: A study on epidemiological system identification, " Mathematical Problems in Engineering, vol. 2009, 2009, pp.I-1 O.

[34]. S. Hoya WhiteS., A. Martm del Rey. and G.A. Rodnguez Sanchez, "Modeling epidemics using cellular automata, " Applied Mathematics and Computation, vol. 186, pp.193-202, I March 2007.

[35]. S. A. Billings and Yingxu Yang, "Identification of Probabilistic Cellular Automata, "IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics vol. 33, pp. 225-236, April 2003.

[36]. Centers for Disease Control and Prevention. Receiving, Distributing, and Dispensing Strategic National Stockpile Assets: A Guide for Preparedness, Version 10.02, 2006.

[37]. E. K. Lee, S. Maheshwary, J. Mason, and W. Glisson. Large-scale dispensing for emergency response to bioterrorism and infectious-disease outbreak. Interfaces, 36(6):591:607, Nov. 2006.

[38]. T. Schneider, A. R. Mikler, and M. ONeill II. Computational Tools for Evaluating Bioemergency Contingency Plans. In Proceedings of the 2009 International Conference on Disaster Management, 2009.

[39]. G. Pezzino and S. Fellow. Guide for Planning the Use of Scarce Resources During a Public Health Emergency in Kansas. 3936(September), 2009.

[40]. J. Timbie, J. Ringel, D. Fox, and D. Waxman. Allocation of Scarce Resources During Mass Casualty Events. (207), 2012.

[41]. J. G. Hodge and M. Ll. The Ethics of Allocation and Preparedness in Public Health Emergencies Responses 10 Core Principles of Public Health Emergency. pages 1{2, 2012.

[42]. ESRI. Disaster preparedness exercise uses GIS, 2006.

[43]. E. K. Lee, C.-H. Chen, F. Pietz, and B. Benecke. Modeling and Optimizing the Public-Health Infrastructure for Emergency Response. Interfaces, 39(5):476{490, Oct. 2009.

[44]. C. A. Bradley, H. Rolka, D. Walker, and J. Loonsk. Biosense: implementation of a national early event detection and situational awareness system. MMWR Morb Mortal Wkly Rep, 54(Suppl):11{19, 2005.

[45]. M. Coombes. De_ning locality boundaries with synthetic data. Environment and Plan-ning A, 32(8):1499{1518, 2000.

[46]. T. Jimenez, A. R. Mikler, and C. Tiwari. A Novel Space Partitioning Algorithm to Improve Current Practices in Facility Placement. IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans: a publication of the IEEE Systems, Man, and Cybernetics Society, 42(5):1194{1205, Sept. 2012.

[47]. A. Marn. The discrete facility location problem with balanced allocation of customers. European Journal of Operational Research, 210(1):27{38, 2011.

[48]. C. S. Revelle, H. A. Eiselt, and M. S. Daskin. A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research, 184(3):817{848, 2008.

[49]. S. H. R. Pasandideh, S. T. A. Niaki, and V. Hajipour. A multi-objective facility location model with batch arrivals: two parameter-tuned meta-heuristic algorithms. Journal of Intelligent Manufacturing, 24(2):331{348, 2013.

[50]. J. G. Carlsson and F. Jia. Continuous facility location with backbone network costs. Transportation Science, forthcoming. Available online at http://menet. umn. edu/ jgc/, accessed on June 6:2013, 2013.

[51]. T. K�u_c�ukdeniz, A. Baray, K. Ecerkale, and S_. Esnaf. Integrated use of fuzzy c-means and convex programming for capacitated multi-facility location problem. Expert Systems with Applications, 39(4):4306{4314, 2012.

[52]. J. Redondo, J. Fern_andez, J. _Alvarez, A. Arrondoa, and P. Ortigosa. Approximating the pareto-front of continuous bi-objective problems: Application to a competitive facility location problem. In Management Intelligent Systems, pages 207{216. Springer, 2012.

[53]. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL et.al. The global distribution and burden of dengue. Nature; 496:504-507.

[54]. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis. 2012;6: e1760. doi: 10.1371/journal.pntd.0001760.

[55]. Fotheringham, A. S.; Charlton, M. E.; Brunsdon, C. (1998). "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis". Environment and Planning A 30 (11): 1905–1927. doi:10.1068/a301905

[56]. A. Smola and B. Schölkopf. A Tutorial on Support Vector Regression. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK, 1998.

[57]. Cortes, Corinna, Vapnik, Vladimir N., "Support-Vector Networks," Machine Learning, vol. 20(3), pp. 273 – 297, 1995.

[58]. Anno S, Imaoka K, Tadono T, Igarashi T, Sivaganesh S, et al. (2014) Assessing the Temporal and Spatial Dynamics of the Dengue Epidemic in Northern Sri Lanka using Remote Sensing Data, GIS and Statistical Analysis. J Geophys Remote Sensing 3:135 doi: 10.4172/2169-0049.1000135

[59]. Rupasinghe, C.S.; Gamage, D.S.; de Alwis, C.; Mufthas, M.R.M.; Dabarera, R. “Using adaptive fuzzy systems for controlling dengueepidemic in Sri Lanka “. 5th International Conference on Information and Automation for Sustainability (ICIAFs)”, 2010.

[60]. Brunsdon, C., Fotheringham, A.S. and Charlton. M.E. (1996) “Geographically weighted regression: a method for exploring spatial non-stationarity’, Geographical Analysis, vol 28, no 4, pp 281-98.

[61]. Vaidya, A., Bravo-Salgado, Angel D., Mikler, A. R. “Modeling Climate-dependent Population Dynamics of Mosquitoes to Guide Public Health Policies”, Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2014, pp 380-389.

[62]. Mitchell, Melanie (1996). An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press. ISBN 9780585030944.

[63]. Hawe, GI, Coates, G, Wilson, DT, Crouch, RS. Agent-based simulation of emergency response to plan the allocation of resources for a hypothetical two-site major incident. Eng Appl Artif Intell. 2015; 46:336-345. http://dx.doi.org/10.1016/j.engappai.2015.06.023.

[64]. Petrovic N, Alderson DL, Carlson JM (2012) Dynamic Resource Allocation in Disaster Response: Tradeoffs in Wildfire Suppression. PLoS ONE 7: e33285

[65]. Fiedrich F, Gehbauer F, Rickers U. Optimized resource allocation for emergency response after earthquake disasters. Safety Science 2000;35(1e3):41e57.

[66]. Arora H, Raghu TS, Vinze A (2010) Resource allocation for demand surge mitigation during disaster response. Decis Support Syst 50(1):304–315

[67]. Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inf Theory. 1967;13(1):21–27. doi: 10.1109/TIT.1967.1053964.

[68]. Limkittikul K, Brett J, L'Azou M. Epidemiological Trends of Dengue Disease in Thailand (2000–2011): A Systematic Literature Review. Halstead SB, ed. PLoS Neglected Tropical Diseases. 2014;8(11): e3241. doi: 10.1371/journal.pntd.0003241.

[69]. https://neo.sci.gsfc.nasa.gov/ (Accessed December 2016)

[70]. WHO. Geneva: World Health Organization; 2009. Dengue: guidelines for diagnosis, treatment, prevention and control – New ed.

[71]. WHO. Geneva: WHO; 1997. Dengue haemorrhagic fever. Diagnosis, treatment, prevention and control; pp. 12–23.

[72]. Ooi E-E, Gubler DJ. Dengue virus-mosquito interactions. In: Hanley KA, Weaver SC, editors. Frontiers in dengue virus research. Norfolk, UK: Caister Academic Press; 2010. pp. 143–56.