Hybrid Application of AVHRR based Satellite Remote Sensing and ENSO Signals for Early Warning and Monitoring of Malaria in Asia and South America

BY

Mohammad Nizamuddin

A dissertation submitted for the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
The City University of New York

2010

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of philosophy.

_____	_____
Date	Leonid Roytman, Ph. D Chair of Examining Committee
_____	_____
Date	Mumtaz K. Kassir, Ph.D Executive Officer

Supervisory Committee:

Mohammed Ali, Ph.D
Alexander Gilerson, Ph.D
Nir Y. Krakauer, Ph.D

Outside reader:

Felix Kogan, Ph. D

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Hybrid Application of AVHRR based Satellite Remote Sensing and ENSO Signals for Early Warning and Monitoring of Malaria in Asia and South America

By

Mohammad Nizamuddin

Advisor: Professor Leonid Roytman

A better understanding of the relationship among satellite-observed vegetation health, climatic anomalies, ENSO events and malaria epidemics could help mitigate the worldwide increase in incidence of mosquito-transmitted diseases.

This thesis investigates association between Vegetation Health (condition) Index and malaria transmission over the last 15-years in eight regions of different ecosystems within four different nations in South Asia and South America. The Vegetation Health (condition) Index, derived from a combination of Advanced Very High Resolution Radiometer (AVHRR) based Normalized Difference Vegetation Index and 10-μm to 11-μm thermal radiances, was designed for monitoring moisture and thermal impacts on vegetation health.

This study attempts to identify the potential factors for malaria transmission for different climatic conditions. We demonstrate that thermal condition is more sensitive to malaria transmission with different seasonal malaria activities. The weekly VH indices were correlated with the epidemiological data. A good correlation was found between malaria cases and TCI and VCI one to two months earlier than the malaria transmission season. Two different malaria transmission seasons in some area have been also detected because of two major vectors of different seasonal behavior. Following the results of correlation analysis principal component regression (PCR)

method along with leave one out cross validation was used to construct a model to predict malaria as a function of the TCI and VCI.

Furthermore, I investigate 15-year association between monthly sea surface temperature (SST) anomalies in the tropical Pacific and Southern Oscillation Index (SOI) with Vegetation Health Indices for the some study regions. I also incorporate SST, SOI and Vegetation Health data into statistical model for long term forecasting malaria transmission. The overall results show that remote sensing is a valuable tool for anticipating malaria severity well in advance so that preventive measures can be taken.

FOREWORD

First of all, let me take the opportunity to express my deepest sense of gratitude to my professor Leonid Roytman for his invaluable patience, guidance and kindness to guide me all along. My heartiest thanks go to Dr Felix Kogan for helping me develop my expertise in the fascinating world of ecology, environment and meteorological satellites which galvanized me to work and to love being in the field. Both of them gave me a tremendous level of support and encouragement throughout my Ph.D. endeavor. I am also very grateful to the other members of the committee Professor Mohammed Ali, Professor Alexander Gilerson, and Professor Nir Y. Krakauer. Their constructive comments and suggestions have greatly improved the quality of my work.

I thank my family for all their motivational role, patience, and support. My mother, Jahanara Begum, who passed away in the middle of my Ph.D. program, was always instrumental to boost up my confidence and made me know the values she cared for. I am especially ever grateful to her for instilling into me the sense of ethics. My deepest sense of gratitude goes to my father, Dr. Abu Sufian, who has given me determination and to take pride in ones work. My brother Salahuddin with whom I was brought up and share our feelings since our childhood, encouraged me all the time to complete my degree soon and had extended moral support both in season and out of season.

Words are not enough to thank my wife Sabreen Imam duly. She smilingly shared all the pains and stresses equally with me throughout my Ph.D. journey. My daughter Noha's presence in my life added extra dimension and has always been a source of delight and inspiration. The stresses I went through during the period were overshadowed by her divine company. She deserves my special gratitude.

Table of Contents

List of Figures

List of Tables

CHAPTER ONE

1. Introduction

Malaria is the most deadly parasitic human infection, with an estimated 500 million clinical attacks worldwide resulting in more than a million deaths each year (Thomson et al. 2006). Human malaria also imposes drastic economic production losses (Sachs and Malaney 2002). Nearly half of the world's population, consisting of poor countries primarily, is at risk for malaria. This is a complex disease that requires an association of three factors – parasite, vector and host – to continue its life cycle. The increasing trend of environmental change is dramatically affecting malaria distribution patterns locally and regionally (Srivastava 2001). On the other hand anti-malarial drugs are too expensive to be of any use to low-income populations. Although many low-income countries have been using insecticides as an affordable way to control mosquito population and malaria, additional funding is still needed. Therefore malaria early warning models would be valuable for economically disadvantageous nations in targeting mosquito control efforts and aiding the government health management system to prepare in advance before the peak malaria season starts.

Inter-annual climate variability is an important determinant of malaria epidemic (Kovats 2000) . Specifically, climate drives both mosquito vector dynamics and parasite development rates. Therefore, skillful seasonal climate forecasts may provide early warnings for instead the changing rate of increased malaria risk. In most of the cases, the required information includes temperature, humidity and precipitation for forecasting malaria. Up-to-date ground–based information on meteorological conditions may not be available for some affected regions due to expensive and time-consuming collection process. Satellite remote sensing can be used as a potential alternative since it covers the entire land surface with good temporal and spatial

resolution at a relatively low cost. Correlations between environmental condition and remotely sensed radiances could be helpful for malaria early warning models. To develop a model for malaria prediction, statistical analysis can be performed once significant correlation between malaria incidence and satellite sensed surface conditions is identified. The model is to produce adequate and timely forecasting of malaria epidemics with the ultimate aim for reducing the suffering caused by the disease. A timely response is the most relevant factor in reducing the impact of an epidemic, because effective control measures can be undertaken as soon as the epidemic has been detected and before it spreads further.

1.1 Satellite Remote Sensing

The science and art of analyzing an object on the basis of the data obtained by a device which is not directly connected to that object is known as remote sensing (Lillesand and Keifer 1994). In order for a remote sensing to produce an effective analysis on an object, three things are necessary, namely, a source of energy, a target, and a sensor for recording the interactions of electromagnetic radiation (ER) with that object. Satellite remote sensing involves specially designed satellite sensors that receive the energy reflected or emitted from the object which facilitates an analysis of nature of the target.

1.1.1 Spectral Characteristics of Ecosystem

It is essential to know and understand the variations in vegetation pattern including its growth cycles relating to climatic conditions and alterations of biological phenomena to provide a full description of the geological, climatic, and physiographic properties or characteristics of a

certain location (Jones et al. 1996). Determining the physiographic characteristics of a specific location is possible through digital image algorithms and sensors from remote sensing data that can actually extract the critical vegetation information (Didan and Huete 2004).

The earth surface has four major characteristics, namely, the difference of the living vegetation from the dead vegetation, chlorophyll pigments red light absorption, water low wavelength reflection, and linear wavelength demonstration of the soil reflectance (Curran 1984, Jensen 2000, Liang 2004). Healthy vegetation reflects greater infrared light than the stressed or dead vegetation because of the ability of leaf internal tissues to scatter light. On the other hand the healthy vegetation absorbs more visible light than stressed or dead vegetation because of its ability for pigments to absorb light (Curran 1984, Jensen 2000, Liang 2004). In addition, reflectance helps to infer vegetation differences from the values and results gathered from spectral measurements (Tucker et al. 1983, Tucker et al. 1985).

1.1.2 Spectral Properties of Vegetation

The interaction of visible and near infrared sunlight with land surface involves plants and leaves with varying spectral properties. So, it is important to understand inner structure of leaf, which controls the leaf's spectral characteristics.

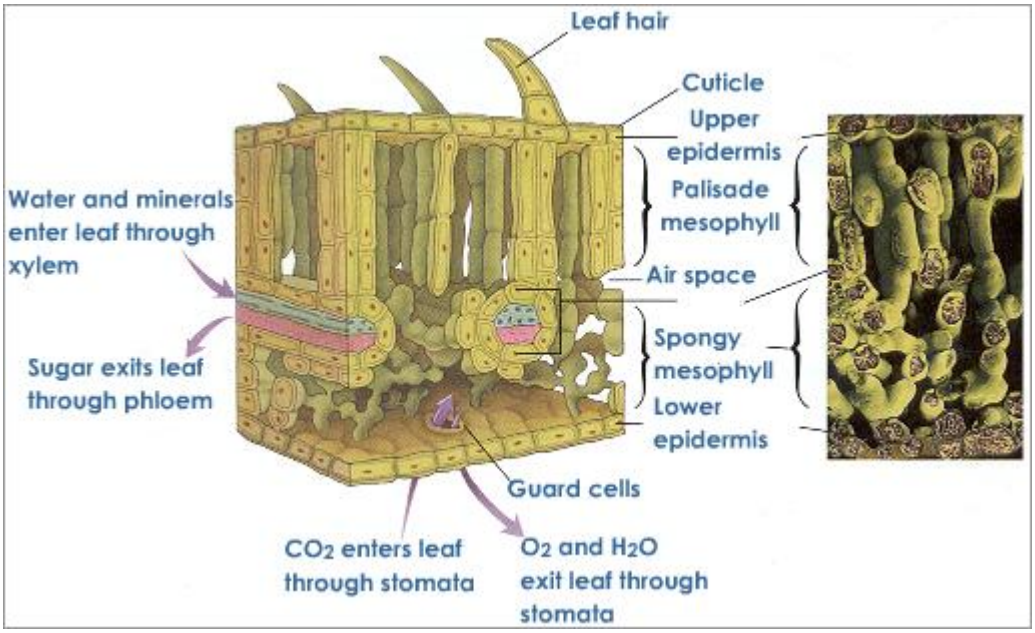


Figure 1.1: Cross Sectional Diagram of a Typical Leaf

The Figure 1.1 shows the typical cross section of a leaf (http://image.tutorvista.com/content/nutrition/cross-section-of-leaf.jpeg). The cuticle is the translucent layer that covers the lower as well as the upper surface of the leaf. The stomates of are the minute openings located within the upper epidermis that helps balance the leaf moisture content (Campbell et al. 2004).

Chlorophyll absorbs 70 to 90% in the blue as well as the red parts of the spectrum, with an absorption minimum around 0.55 μm that contributes to the green appearance of the leaf. The mesophyll tissue is responsible for scattering the NIR radiation that passes through the leaf's upper epidermis and causes the leaf to absorb only very little of the NIR radiation, because 60% of the spectrum wavelength is already scattered outside the plant's leaf (Jensen 2000).

Campbell et al. (2004) mention that as the plant matures, the absorption spectrum location also changes. Gates et al. (1965) first reported the chlorophyll red-shift absorption edge. Collins

(1978) later confirmed the chlorophyll absorption red shift. An important assumption for the plant to go thorough stress is lack of plant moisture which can easily be detected from visible and NIR parts of the spectrum. Any damage made on the mesophyll tissue or the leaf's internal structure causes a direct decrease of its NIR reflectivity (Myneni et al. 1998).

Figure 1.2 shows the major factors that control the leaf reflectance (spectral analysis of vegetation, www.forestry.ubc.ca)

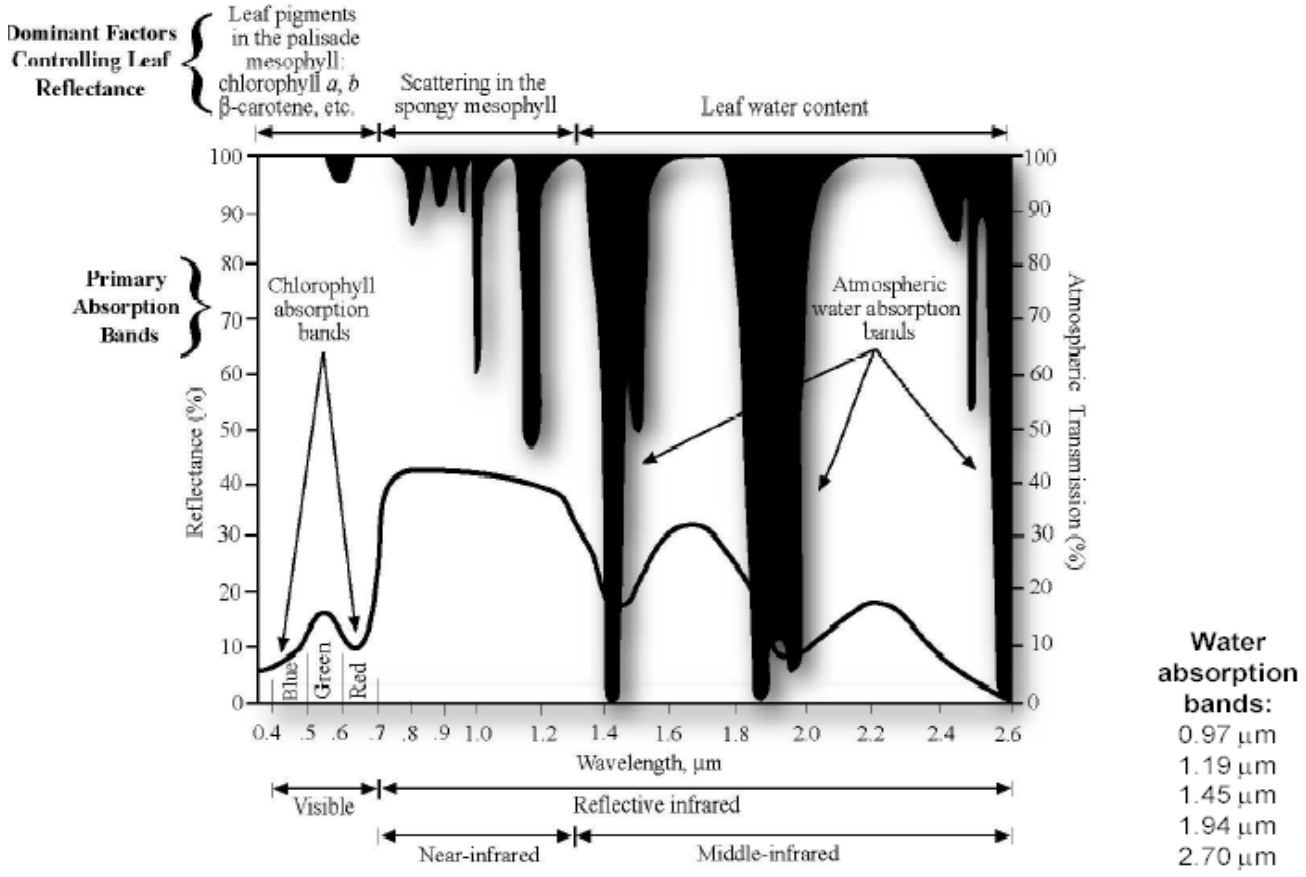


Figure 1.2: Factors Controlling Leaf Reflectance

Researchers have developed models for surface temperature Ts extracted from data gathered through thermal remote sensing. The algorithm of the emissivity estimation on land, part of the land-surface temperature calculation, has been reviewed by Prata et al. (1995).

1.2 Environmental Satellites

1.2.1 Meteorological Satellites

Meteorology is an interdisciplinary study that focuses on the atmosphere with the help of observational earth sciences and theoretical physics. It is considered a branch of observational earth sciences with an aim for providing an exact knowledge of the atmosphere. Images obtained from satellite are invaluable source of information for the purpose of operational forecasting. These images are used for a number of purposes including as (a) an analytical tool, especially for regions with limited data availability like the tropics; (b) direct aid to short period forecasts (6-12 hours ahead of cloud, rainfall, floods etc.); (c) input to numerical weather prediction models (NWP) for defining initial conditions; and (d) monitoring the model forecast. Meteorological satellites operate as an observing platform with appropriate sensors on board. These satellites transmit valuable information in the form of images and sounding to the base stations located on the earth's surface.

The Figure 1.3 shows the metsats locations, which includes research satellites and operational geostationary and the polar-orbiting satellites.

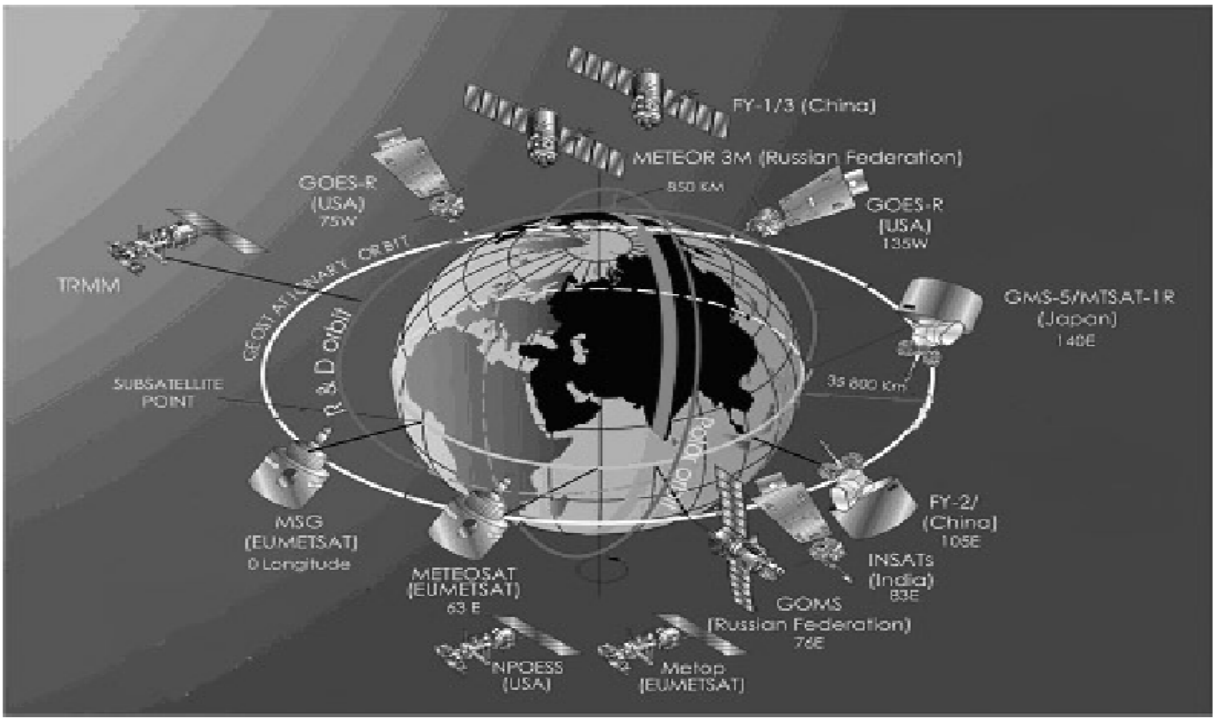


Figure 1.3: Global Networks of Geostationary and Polar-Orbiting Satellites (http://www.wmo.int/index-en.html space-based global observation systems)

The polar orbiting satellites orbit from pole to pole at a height of 860 km. These satellites revolve around the earth at a sun synchronous orbit. This makes it possible for the satellite to scan the earth in an entirely different strip in each orbit. The polar orbiting satellite always crosses the earth's equator at a fixed estimated local time that is relative to the sun. The satellite can complete one orbit at about 102 minutes. The swaths have a distance of approximately 2600 km wide, complete about 14 orbits a day, and provide coverage at least twice in a 24-hour cycle (Conway 1997). Example of polar orbiting satellites includes NOAA, IRS, ERS-1 and ERS-2, TRMM (low inclination), DMSP, Oceansat-1 etc.

1.2.3 NOAA Satellites

The NOAA (National Oceanic and Atmospheric Administration) and the NESDIS (National Environmental Satellite Data and Information Service) are two agencies that operate the environmental satellites. The NOAA is responsible for the GOES and the polar orbiting satellites. In this study we use polar orbiting satellite which is appropriate for capturing long term data, moisture profiles, vertical temperature, and forecasting. At present, NOAA is operating five polar orbiters. The Advanced Very High Resolution Radiometer (AVHRR) provides four to six–band multi spectral data from the NOAA polar-orbiting satellite series. With fairly continuous global coverage since June 1979, the AVHRR measures clouds over the ocean and land in 5 visible and near IR bands. A new series of polar orbiters includes improved sensors which had started with the launch of NOAA-15 in May 1998 and NOAA-16 on September 21, 2000. The newest, NOAA-17, was launched on June 24, 2002. NOAA-l2, NOAA-14 and NOAA-15 all continue transmitting data as standby satellites (NOAA KLM user's guide http://www2.ncdc.noaa.gov/docs/klm and http://science.nasa.gov/missions/noaa). The important characteristics of NOAA satellites for AVHRR sensors are the following.

Spatial Resolution

1.1 km at nadir

8 km at 55 degree view angle

Temporal Resolution

9 day repeat cycle for orbit

Giving 2 images per day (one in day, one at night)

Sun synchronous orbit at 830 km altitude

Spectral Resolution

Bandwidth (microns)

Channel 1: 0.580 - 0.68 (RED)

Channel 2: 0.725- 1.00 (NIR)

Channel 3: 3.550 — 3.93 (MIR)

Channel4: 10.300-11.30 (TIR)

Channel5: 11.500-12.50 (TIR)

Radiometric Resolution 10 bit quantization

Calibration Details

Bands 3, 4, 5: on board calibration

Bands 1 and 2: deep space count only

1.2.3 Data Acquisition

The AVHRR data are collected in two different formats; High Resolution Picture Transmission (HRPT) and Local Area Coverage (LAC). Full resolution image of HRPT data are transmitted to a ground station as they are collected (Kidwell 1997). LAC are full resolution data that are recorded on an onboard tape for subsequent transmission during a station overpass. NOAA samples the HRPT data to an approximate range of 5 km pixels so as to achieve global coverage by means of onboard storage facility. The specific format of HRPT data as sampled by NOAA is commonly known as the global area coverage (GAC). GAC data are derived from four out of every five samples along the scan line that are used to compute one average value. The data from only every third scan line are processed, yielding 1.1 km by 4 km resolution at the sub-point. Sampling those observations and mapping them into a regular grid the following two global–mapped AVHRR datasets have been collected, with more reduction of the data volume by

temporal compositing that also eliminates cloud contamination (Gutman et al. 1999): NASA Global Inventory Monitoring and Modeling Studies (GIMMS) and NOAA global vegetation index (GVI). In my research, I use third generation Global Vegetation Index (GVI) weekly composite data which are produced operationally by NOAA/NESDIS.

The use of long term data sets for monitoring biophysical phenomena has been instrumental to emphasize the need for instrument calibration (Kogan et al. 2001, Simoniello et al. 2004). The calibration process of a satellite radiometer is somewhat complex, which requires identifying the transfer function between digital counts from the instrument and the scene input spectral radiance for all of its spectral channels. Recognizing its importance, the AVHRR instruments carried on each successive NOAA mission were all calibrated against a standard to obtain a linear relationship between digital numbers and spectral radiance (Kidwell 1997). This specific calibration is known as the pre-launch calibration and has been adopted by NOAA. An important assumption underlying the pre-launch calibration is that both gain and offset of the instrument were stable over time. A number of studies, however, revealed that the assumption might not hold true as the actual calibration values were not stable over time and differed from the published pre-flight values (Price 1988, Kaufmann and Zhou 2000). However, once the satellite is in space, the onboard facility to determine the gain of the solar (visible) channels (one and two) is no longer available.

The inter-annual and intra-annual noises due to changing illumination and viewing conditions, degradation of sensor quality over time, satellite navigation and orbital drift, atmospheric and surface conditions, methods of data acquisition, and communication and random errors (Los et al., 1994, Gutman 1991) make it necessary for post launch calibration of VIS and NIR following Rao and Chen (1995, 1999) for GVI-based radiances. This post lunch calibration includes also

calculation of NDVI and converting the channel 4 (CH4) radiance to brightness temperature (BT), with the latter being corrected for non-linear behavior of the sensor (Kidwell 1997). As a result, the long-term noise and shifts between the satellites in radiances and NDVI were reduced substantially.

1.2.4 Vegetation Index

A Vegetation Index (VI) is developed by some combination of remote sensing bands and expressed in unit-free numbers. The VI is normally linked to the amount of vegetation in a given image pixel. More specifically, there is supposed to be a relationship between the vegetation index and plant biophysical parameters that holds for the most ecosystem so as to facilitate both validation and calibration (Jiang et al. 2006, Xiang et al. 2003). It is important to note that the algorithm for vegetation index should be normalized to avoid internal (for example canopy background) and external (for example sun angle) effects.

Some examples of vegetation index are:

Normalized Difference Vegetation Index (NDVI): This index has been discussed by Kriegler et al. (1969): NDVI = (NIR-red)/(NIR+red). This is the most common VI and has a range of -1 to 1. The visible and the near-infrared light reflected by vegetation are used to calculate the NDVI. While healthy vegetation absorbs most of the visible light that hits it, unhealthy or sparse vegetation reflects most of the visible light. As a result, healthy vegetation has for higher NDVI than unhealthy vegetation.

The percentage vegetation index or PVI was first used by Crippen (1990) as a replacement for NDVI. Crippen explained that as the sum of NIR and red was the important use of the red band

in NDVI, PVI was sufficient as the latter captures all of these elements in its formula as given by PVI = NIR / (NIR + red). Unlike the range of -1 to 1 for NDVI, PVI can assume any value between zero and unity.

1.3 Entomological – climate Interactions of Malaria Transmission

The focus of this thesis is to understand the likely influence of climate change on vector production and malaria transmission of different ecosystem and how to apply satellite data for accurately forecasting malaria. Three different types of epidemics have been generally defined by scientists (Thomson et. al 2000, Greena et al. 2002), which are caused by a disturbance of the equilibrium between host, parasite and vector. Those epidemics are caused by;

a. Meteorological conditions (create temporary epidemics)

b. Landscape changes or colonization of sparsely populated areas (create a new equilibrium level of endemicity)

c. Interruptions in measures that were controlling malaria

In order to clearly understand the influence of entomological–climate interactions on malaria transmission the life cycles of parasite and vector are important to investigate. The range of temperature, precipitation, humidity, for malaria proliferation shall be discussed in this section. Malaria is a complex disease that requires an association of three factors – parasite, vector and host – to continue its life cycle. Epidemic malaria is associated with seasonally warm, semi-arid, highland and forest hill areas. There have been immense efforts to correlate malaria and the environmental condition as the latter influences development of both parasite and vector. In spite of several surveys of vector distribution from South Asia (India and

Bangladesh in our study area) and South America (Brazil and Colombia) large areas still remain unexplored, specifically deep–forested areas where manual surveys are very tedious due to difficult accessibility. Despite limitations in our ecological understanding, we can say that inter-annual climate variability is an important determinant of the epidemic where the climate drives both mosquito vector dynamics and parasite development rates. Hence skillful seasonal climate forecasts may provide early warning of changes of risk in epidemic-prone regions. The following section explains the environmental conditions and life cycles for malaria parasites and vectors and its proliferation.

1.3.1 Malaria as a Disease

Malaria in humans is caused by four species of *Protozoan* parasites of the genus *Plasmodium: P. falciparum, P. vivax, P. ovale,* and *P. malariae.* Although *P. vivax* is responsible for most malaria infections in the world, the most severe form of malaria is caused by *P. falciparum.*

1.3.2 Geographic Distribution of Malaria

It is worth noting that half of the world's population (Fig. 3.3) is exposed to the risk of malaria and in 2008 alone, about 243 million cases of malaria were reported leading to approximately 863,000 deaths (World Malaria Report 2009). Several hundred million new malaria infections are reported each year causing about a million deaths mostly in children under the age of five (WHO 1999). Malaria has a highly non-linear response to temperature (Patz et al. 1998), indicating that other things remaining constant, a slight increase in temperature could have large effect on malaria incidence.

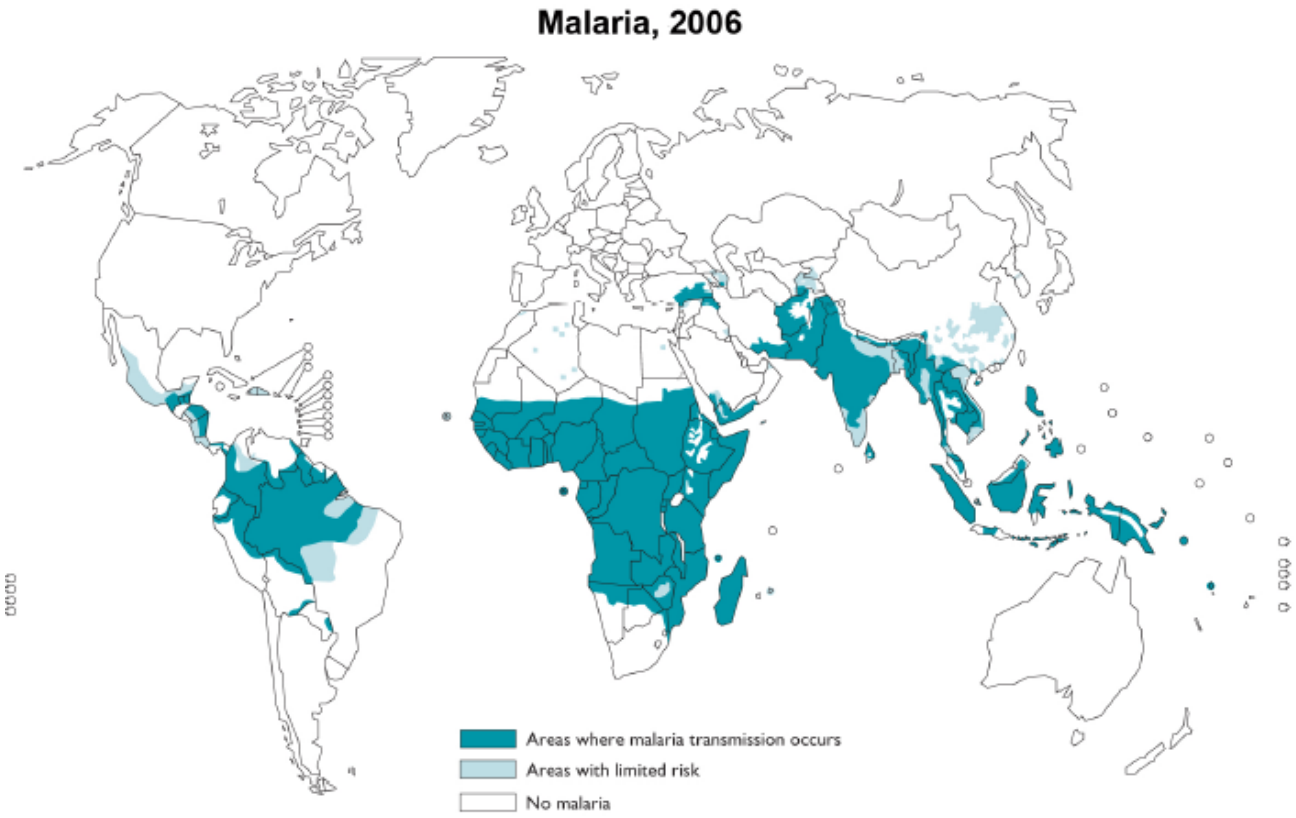


Figure 1.4: Distribution of Malaria around the World (WHO 2007)

1.3.3 Life Cycle of Malaria Parasite

The epidemiology of malaria constitutes man as host, four species of *Protozoan* (*Plasmodium vivax, Plasmodium falciparum, Plasmodium malariae* and *Plasmodium ovale*) as parasite and *anopheline* mosquitoes as vectors. Environmental conditions play an important role in the transmission dynamics of malaria, as the parasite has to pass its development a part of its cycle in

the mosquito. The three main climatic factors that affect malaria transmission and distribution are temperature, precipitation and relative humidity. Climate predicts, to a large degree, the natural distribution of malaria. The logic behind the possible increase or decrease in malaria transmission in response of climate change is based on the interplay between temperatures and developmental cycles in *Anopheline* mosquitoes. It is imperative to understand the intricacies of the biology of *Anopheline* mosquitoes and the lifecycle of the malaria parasite.

The life cycle of the malaria parasite is complex. The process – three phases in the mosquito and two in the human host – has been divided by scientists into nearly a dozen separate steps. The parasite is transmitted to humans by the sporozoite forms in the saliva of infected female mosquitoes of the genus *Anopheles*. Soon after entering the human host, the sporozoites invade liver cells, where during the next 5 to 15 days they develop into schizonts. Each schizont contains 10,000 to 30,000 “daughter” parasites called metozoites, which are released and invade the red blood cells. Once inside the red blood cells, each merozoite matures into a schizont containing 8 to 32 new merozoites (Sinha 2005). The red blood cell eventually ruptures and releases the merozoites, which are then free to invade additional red blood cells. The rupturing of red blood cells is associated with fever and signals the clinical onset of malaria (Sinha 2005).

Some merozoites differentiate into sexual forms, namely gametocytes, which are ingested by a mosquito during its next blood meal. Once in the mosquito, the sexual forms leave the blood cells, male and female gametes fuse to form a zygote. Over the next 12 to 48 hours, the zygote becomes highly motile and is called “ookinete”. The ookinete penetrates the wall of the insect’s stomach and becomes an oocyst. Over the next week, depending on the parasite species and ambient temperature, the oocyst enlarges forming more than 10,000 sporozoites (Sinha 2005).

When the oocyst ruptures, the sporozoites migrate to the mosquito salivary glands, from where they may be injected back into the human host, thus completing the cycle. Figure 1.5 shows different stages of malaria parasite life cycle with brief discussion.

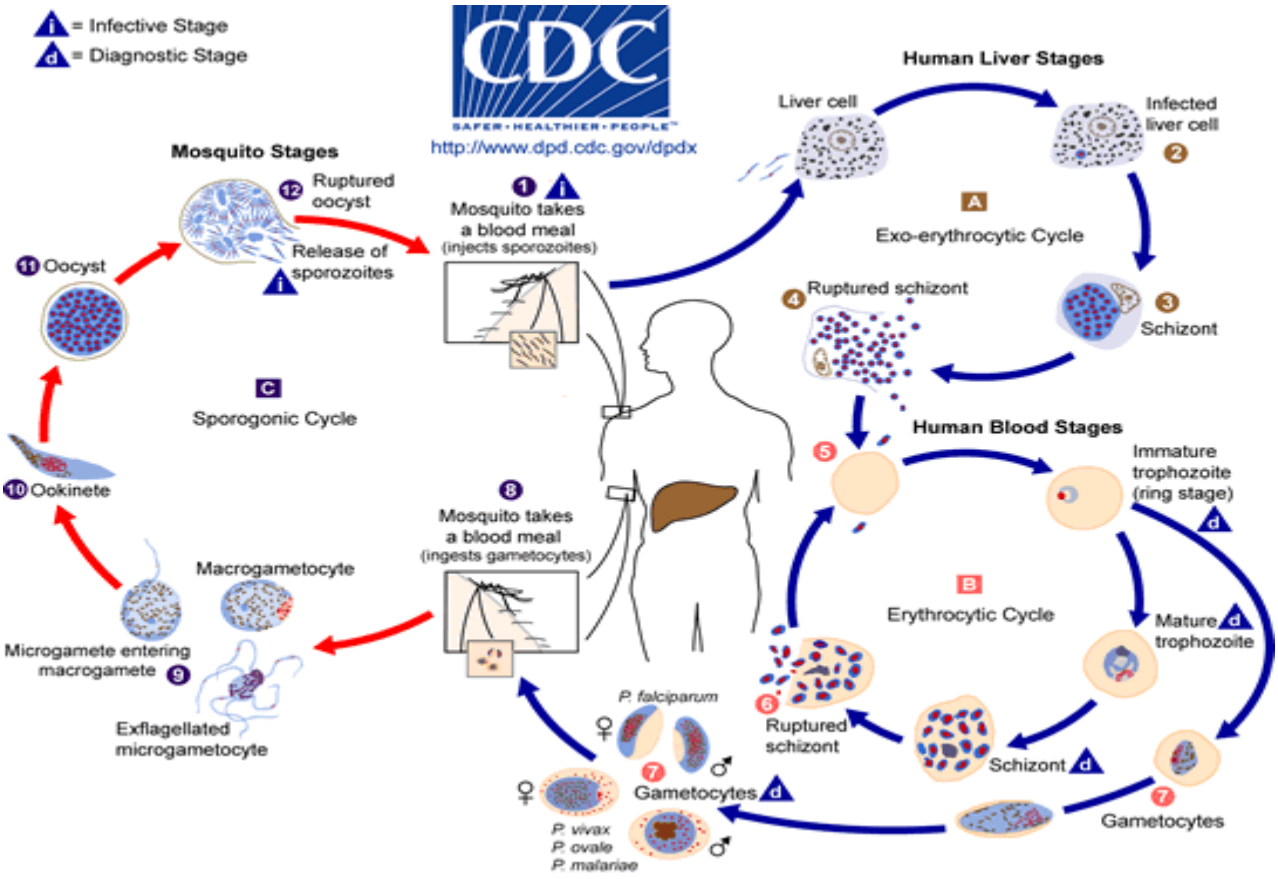


Figure 1.5: Malaria Parasite Life Cycle (CDC)

The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female *Anopheles* mosquito inoculates sporozoites into the human host ❶. Sporozoites infect liver cells ❷and mature into schizonts ❸, which rupture and release merozoites ❹. (Of note, in *P. vivax* and *P. ovale* a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years later.) After this initial replication in the liver (exo-erythrocytic schizogony Ⓐ), the parasites undergo asexual multiplication in the

erythrocytes (erythrocytic schizogony B). Merozoites infect red blood cells 5. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites 6. Some parasites differentiate into sexual erythrocytic stages (gametocytes)7. Blood stage parasites are responsible for the clinical manifestations of the disease.

The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an *Anopheles* mosquito during a blood meal 8. The parasites' multiplication in the mosquito is known as the sporogonic cycle C. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes 9. The zygotes in turn become motile and elongated (ookinetes) 10which invade the midgut wall of the mosquito where they develop into oocysts 11. The oocysts grow, rupture, and release sporozoites 12, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites 1into a new human host perpetuates the malaria life cycle (Centers for Disease Control and Prevention (CDC)).

1.3.4 The Mosquito

Only a sub-group of 50 to 60 species out of the more than 2,500 known species of mosquitoes worldwide belonging to the genus *Anopheles* are capable of transmitting malaria. Female *Anopheline* require blood meals to reproduce (Figure 1.6 a). Some *Anopheline* species are indiscriminant feeders; others prefer to feed on animals (zoophilic) or humans (anthropophilitc).

The mosquito undergoes four stages of growth: egg, larva, pupa, and adult (imago), Adult females mate once and store the sperm. The female may deposit a total of 200 to 1,000 eggs in three or more batches (Sinha 2005). Actual egg production is dependent on blood consumption. After hatching, *Anopheline* larvae lie along the water-air interface, where it is thought that they

feed on organisms along the surface film. Adult mosquitoes develop from the pupil stage within 2 to 4 days. An adult mosquito will emerge from the egg stage in 7 to 20 days, depending on the species of mosquito and environmental conditions.

Female *Anopheline* mosquitoes can survive at least a month under favorable conditions of high humidity and moderate temperatures. That is sufficient time for them to take a blood meal, for the parasite to develop, and for the mosquito to take another blood meal and thus transmit the parasite to a second human host (Sinha 2005).

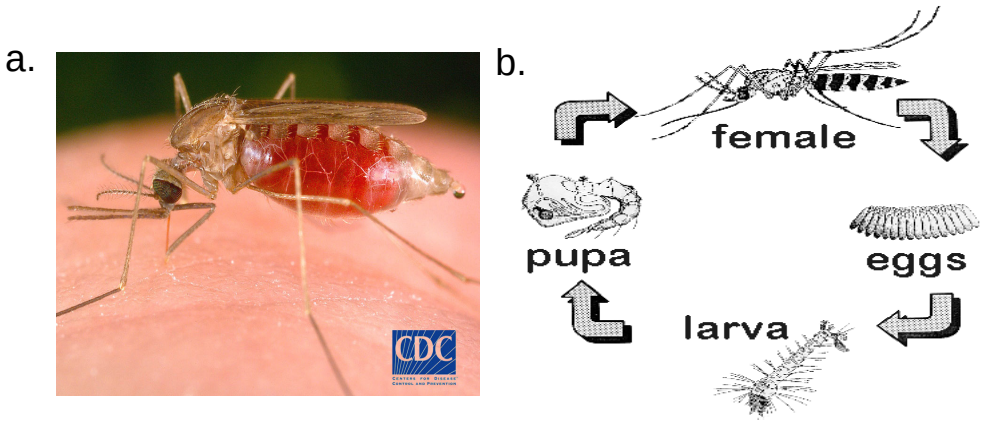


Figure 1.6 : a) *Anophilis Dirus* Mosquitoes b) Typical Life Cycle of Mosquitoes(CDC)

1.3.4 Malaria Incidence Climate Parameter and Their Duration

Malaria transmission is normally affected by two sets of factors. The first set of factors has influence on malaria transmission due to short term variation in their characteristics. This includes mostly weather related fluctuation such as temperature, precipitation, and relative humidity. On the other hand, the second set of factors affecting the malaria transmission from a longer term perspective, include such environmental components as ecosystem, desertification, sea level rise, changing vegetation, and agricultural practices. In the long run, malaria transmission also depends on some characteristics relating to human population such as migration, spread of drug resistance, change in immune status, and spread of pathogens into new areas. Environmental conditions are keys to understanding the Malaria transmission because the parasite, the mosquito, has to develop and progress through a cycle to transmit Malaria. Here, I focus on the short–term component only.

When an infected mosquito bites a healthy person, it transfers sporozoites to him/her, which after entering the hepatic cells and the red blood cells develop into male and female gametes within two weeks. The average incubation period of *Plasmodium vivax* and *Plasmodium falciparum*, the major parasites in India is 15 (12–17) and 12 (9–14) days respectively (Ramesh et al. 2008). The parasite cannot develop further in human beings. The gametes are taken up by female *anopheline* mosquitoes along with blood meal, which develop into sporozoites. The development of a parasite into a mosquito is called extrinsic incubation period or sporogony, the duration of which depends on environmental temperatures (Ramesh et al. 2008). Temperature affects the duration of different stages of a mosquito's lifecycle: blood feeding rate; gonotrophic cycle (physiological process consisting of digestion of blood-meal and development of ovaries); and longevity. At increased temperatures, the rate of digestion of blood-meal increases, which in turn accelerates

ovarian development, egg laying, reduction in gonotrophic cycle and a greater frequency of feeding on hosts, thereby is increasing the probability of transmission (Ramesh et al. 2008). A reduction in the duration of the gonotrophic cycle would make the vectors bite more frequently, thereby increasing the probability of malaria transmission. It takes about 10 days for a mosquito egg to reach the adult stage of an *Anopheline* mosquito at an optimum temperature of 28 °C. At lower temperature, the duration gets prolonged while at increased temperature the duration is reduced. However, at more than 40 °C, mortality occurs in adult mosquitoes. Reduction in the duration of gonotrophic cycle and in the extrinsic incubation period of malaria parasite is related to increased rate of transmission. These two entomological variables are especially sensitive to changes in environmental temperature, being reduced with increments in temperature. The minimum temperature required for development of *Plasmodium vivax* parasite in *anopheline* mosquitoes ranges from 14.5– 16.5 °C, while for *Plasmodium falciparum* it ranges from 16.5–19 °C, However, the best conditions for development of the malaria parasite are 20–30 °C temperature and 60% relative humidity (RH) (Ramesh et al. 2008). Conversely, temperatures below 16 ºC, or higher than 30ºC have a negative impact on the growth of the mosquitoes (Yang and Ferreira 2000). The life span of mosquitoes is also determined by the relative humidity. High relative humidity lengthens mosquitoes lifespan. The Mosquito lifespan is a key factor for the malaria parasite to develop in the vector. Therefore, high relative humidity allows the parasite to complete the necessary life cycle and transmit the infection to many people at a geometric progression (Mouchet et al. 1998).

It is interesting to mention that as *Anopheline* mosquitoes require just the right amount of precipitation they mostly breed in water habitats (Bi et al. 2003), though little is known about the biology of this aquatic phase (Oaks et al. 1991). However, it is known that different mosquitoes

within the broad category of *Anopheline* breed on different types of water (Nagpal and Sharma 1995). Although mosquitoes need water to grow, excessive rainfall, or rainfall accompanied by storm conditions can flush away breeding larvae. Amount of rainfall is also critical in malaria transmission as it affects relative humidity and modifies temperature as well as determining where and how many mosquitoes breeding can take place (Pampana 1969).

1.4 Literature Review

In the existing empirical literature a significant link between climatic anomalies and malaria transmission has been identified in different regions of the world. Rainfall is recognized as one of the major factors influencing malaria transmission since it is responsible for creating sufficient surface water for mosquito breeding sites. Temperature is another factor which plays an important role in malaria transmission through its ability to influence the development rate of mosquito's larvae and the survival rate of mosquitoes. Furthermore, at warmer temperatures, the frequency of feeding and digesting of an adult female mosquito increases and the *Plasmodium* parasite develops faster within the female mosquitoes. Relative humidity is also considered as requirement for malaria transmission within a certain range (60% to 80%) (Anne et al. 2007). Recent research suggests that rainfall and temperature estimates can provide useful early warning information relevant to epidemics (Teklehaimanot et al. 2004a and Teklehaimanot et al. 2004b). Consequently, rainfall and temperature are among the essential elements to develop an integrated malaria early warning system for sub-Saharan Africa, as outlined by the World Health Organization, (WHO 2003). Similarly, using multi-model ensemble technique Thomson et al. (2006) forecast malaria based on seasonal climate.

In the past 15 years many scientists have utilized remote sensing techniques to forecast malaria. For example, to assist the malaria vector control in western Kenya highlands, Panchromatic aerial photos, Ikonos and Landsat Thematic Mapper 7 satellite images were collected by Emmanuel et al. (2006). Supervised classification of land-use and land-cover and visual identification of aquatic habitats were conducted to find probability of aquatic habitats and habitats with *Anopheles* larvae which later were modeled based on the digital elevation model and land-use types.

The NOAA and METEOSAT weather satellites have been used to predict changes in malaria transmission dynamics. Pietro et al. (2007) used cluster classification and Spearman's and Pearson's rank correlation to demonstrate that NDVI anomalies were correlated with malaria anomalies mostly in the semi arid north of Eritrea and along the coast of northern Red Sea. Advantage of using NDVI for malaria stratification and malaria early warning over rainfall is that it is indicative of soil moisture and is thus capable of finding areas where additional rainfall may have a significant impact on malaria transmission.

Alberto et al. (2007) examined the relationship between monthly notifications of malaria cases and the environmental conditions in Burundi highlands with specific focus on the Karuzi province. The study used time series of monthly malaria cases from local health facilities, data from rain and temperature records, and the normalized difference vegetation index (NDVI) and applied autoregressive integrated moving average (ARIMA) technique to model the relationship among the variables under investigation. Alberto et al. (2007) found that the variables were highly correlated. In particular, positive association was established among the number of *Anopheles* mosquitoes and rainfall, increased transmission and temperature, and vegetation and malaria seasonality.

Jean Gaudart (2009) introduced an alternative malaria transmission model for Bancoumana, Mali. For the purpose of using remote sensing data in his model, he collected the 15-day composite vegetation index (NDVI) from satellite imagery series (NOAA) from July 1981 to December 2006. Using the ARIMA framework, Jean Gaudart (2009) measured the statistical association between NDVI and incidence of *P.falciparum* infection in Bancoumana. In this particular study, a rise in the incidence of parasitaemia was estimated by an NDVI value obtained from the receiver operating characteristics (ROC) analysis.

Thomson et al. (1997) found a positive relationship between rainfall and NDVI value and vector abundance. In addition, Simon et al. (1998) used multi-temporal meteorological satellite image data to conduct a study to predict the malaria seasons in Kenya. Thus far, the application of remote sensing techniques to malaria control has focused on identifying mosquito habitats. In three study-sites, Simon et al. (1998) calculated NDVI and adjusted coefficients of environmental variables. The strongest correlations were found between malaria admissions and NDVI. The high correlation between epidemiological data of malaria and satellite-based Vegetation Health (VH) indices found in another study (Rahman et al. 2006) is valuable for researching malaria incidence in Bangladesh. In this regard, it is important to note that the VCI (Vegetation Condition Index) and TCI (Temperature Condition Index) measure moisture and thermal conditions, respectively. These indices are closely related to the weather parameters that are critical to mosquito breeding and Malaria epidemiology. For the purpose of his study, data were collected from NOAA (the NOAA-AVHRR) afternoon polar orbiting satellites. The Global Vegetation Index (GVI) has a spatial resolution of 4 km and is sampled to 16 km, with a temporal resolution of seven days. The results show that there is strong association between the numbers of malaria cases and cooler weather conditions, particularly, when TCI>60. On the

contrary, frequency of human malaria is lower when vegetation is under stress, specifically, when TCI<40 .

There exists a growing empirical literature that documents significant relationship between malaria epidemics and El Nino Southern Oscillation (ENSO). For example, Barrera et al. (1999) finds that the El Nino events in Venezuela were followed by a rise in malaria mortality and morbidity of as much 36.5%. On the other hand, in the south–western highlands of Uganda, incidence of malaria grew manifold as a result of the heavy rains associated with the 1997/8 El Nino (Kilian et al. 1998). In addition, Poveda et al. (2001) shows that malaria epidemics in Columbia were linked to ENSO cycles.

1.5 Goals and Objectives

Detailed dynamic models of malaria epidemiology have been developed extensively and have made important contributions to understanding the transmission of malaria. However, most of them have not been able to describe overall transmission dynamics. Even though mosquitoes and vertebrate host dynamics are generally included, most models do not include environment changes and climate patterns that could affect diseases incidence in both spatial and temporal scales. Moisture and thermal condition of a particular region are important factors for malaria transmission. In the past, most of the research for forecasting malaria using satellite data has used NDVI as predictive variable which is not indicative of moisture and thermal condition. Moreover, vegetation health indices invented by Dr. Felix Kogan (1995) requires long time climatology. Rahman et al. (2006) has used these indices only for Bangladesh. There is no model for the strength of relationship among VCI, TCI, climate change and malaria dynamics across

arid, semi arid and tropical regions. The epidemiological malaria scenario varies across states and regions. Therefore, the data for a specific region are not necessarily representative of the situation in other regions. The link between moisture condition, thermal condition and temperature, humidity, rainfall of the particular ecosystem is investigated in this work.

In this research I incorporate malaria incidence, satellite data and ENSO signals into a statistical model (principal component regression) to predict and monitor malaria transmission of four regions of India, four regions of South America and one region of Bangladesh. This thesis explores the 15-year correlation between regional malaria transmission and the vegetation health (condition) index. The vegetation health (condition) index, derived from a combination of the Advanced Very High-Resolution Radiometer based normalized difference vegetation index and 10-μm to 11-μm thermal radiances, was designed to monitor moisture and thermal impacts on vegetation health.

In light of the discussion above, this study will use both vegetation indices and 10.3 μm to 11.3 μm brightness temperature and aim to achieve the following objectives: (a) explore sensitivity of regional ecosystems condition to malaria events for the period of available satellite records; (b) identify some features in the impact of ecosystems in malaria transmission; (c) estimate the similarities and differences in the impacts of weather on different malaria epidemic areas in different regions of the world; (d) explore sensitivity of vegetation indexes to ENSO events, and their impact on malaria and (e) develop robust statistical model for forecasting malaria.

In this research, we incorporate satellite data into a statistical model to offer a timely estimate of malaria transmission for different ecosystems at state and district levels of major malaria risk areas in India, Bangladesh, Brazil, and Colombia. The study utilizes the AVHRR-based Vegetation Health (VH) indices (VCI, TCI) possible early warning indicators of epidemics of

malaria due to specific weather patterns. Furthermore, we investigate the past 15-year association between the monthly sea surface temperature (SST) anomalies in the tropical pacific and southern oscillation index (SOI) with the vegetation health (condition) index for those study areas in the world. We additionally incorporate SST, SOI and vegetation health data into a statistical model for long-term malaria transmission forecasting.

1.6 Thesis organization

Chapter two will provide an overview for study areas and data sets that have been used in this study. In this chapter a brief description of environmental pattern, landscape, ecology, topology, climate conditions, weather conditions, parasite and mosquito characteristics of different study areas has been presented. Moreover, how selection of latitude and longitude for satellite data has been made is also discussed.

Chapter three explains the estimation methodology by using two types of data (satellite, and malaria statistics) for correlation dynamics, multiple regressions and principal component regression. It also explains detection of multicolinearity, validation and testing of the model.

Chapter four presents results, analysis, and a summery of models for eight regions of South Asia and South America. In this chapter detail explanations of dynamic features of the relationship between satellite data and malaria condition in particular ecosystems have been given. Chapter five presents ENSO signals, VCI, TCI and correlation dynamics of different regions in South Asia and South America. Chapter six discussed the contribution and conclusion.

CHAPTER TWO

STUDY AREA AND DATA SET

The study areas consist of eight different sites in South Asia and South America. These eight sites have eight different land surface/ecosystem types (Table 2.1). Two types of dataset were used in this research: satellite data collected from NOAA and malaria statistics collected from Ministry of Health of Bangladesh, India, Brazil and Columbia. The malaria data were collected from local administrative health centers and aggregated to district and later state level. This malaria data included the number of people tested while they visited health center with fever and number of malaria positive cases. Slide positive rate (% of malaria) is expressed as malaria positive rate in percent of number of people tested. Malaria data of India were collected from statistical information based website (http://www.indiastat.com). Malaria statistics of Bangladesh are published in the year book of Bureau of Statistics Bangladesh. Health Ministry of Brazil posts malaria data in their website at (http://tabnet.datasus.gov.br/cgi/tabcgi.exe?idb2007/d04.def). I have obtained malaria data of Colombia from Dr. Felix Kogan, a NOAA scientist and also a member of my doctoral committee. Dr. Atiqur Rahman, a NOAA-CREST research assistant professor at City College helped me find Bandarban, Bangladesh malaria data set. Malaria epidemiology varies across states and regions. Therefore, the data of a specific country is not necessarily representative of the situation in other regions. My goal is to describe the dynamics of malaria transmission and develop a model for its seasonal variability, which utilizes satellite based vegetation health (VH) data. Such knowledge could be instrumental to control strategies. Satellite data have been extracted for the regions with the most populated among the malaria affected areas in that ecosystem. Moreover, proper analysis of the malaria affected area shall allow us to select

satellite data collection area efficiently. Since malaria transmission depends on topography, climate, rainfall, temperature, humidity, socioeconomic background, parasite, and mosquitoes, we shall analyze vectors, parasite, ecosystem, environmental condition and malaria epidemic briefly for the regions (study areas) and their peak malaria transmission seasons following the discussion of satellite data collection. Environmental conditions are keys to understanding the malaria transmission because the parasite, the mosquito, has to develop and progress through a cycle to transmit malaria. We shall compare our correlation of malaria incidence with VHI from satellite data with the major malaria transmission window and see if we can get early warning of high seasonal malaria risk.

Table 2.1: Land surface types and area (Latitude and Longitude) from where satellite data has been selected

Site #	Latitude	Longitude	Climate/Ecosystem	Regions
1	22.5^0N- 22^0 N	91.5^0E - 92.5^0 E	Sub-tropical, Forest fringe, highland	Bandarban, Bangladesh
2	27°11'-29°3'N	71°54'- 74°12'E	Arid or Desert	Bikaner, Rajsthan, India
3	23.00-24.00 N	91.0-91.80E	Sub tropical, Evergreen forest, Cropland, Humid	Tripura, India
4	21.3 - 22.05N	70.70-71.1E	Semi Arid, Monsoon	Gujarat, India
5	22.0 - 22.9 N	84.32-85.10E	Tropical wet and dry	Orissa, India
6	10– 12 S	62 – 66W	Tropical rain forest (always wet), Humid	Rondonia, Brazil
7	7.5 - 10S	69 – 72W	Hot and wet with short dry period	Ampa, Brazil
8	0.5 – 2.5N	75– 77W	High temperature and heavy rainfall, highland	Narino, Columbia

ltln(0.5,2.5,75,77.0);

2.1 Satellite Data

Radiances of the earth surface received from Satellite and measured by the AVHRR on board NOAA-9, NOAA-11, and NOAA-14 polar orbiting spacecrafts have been taken for this study; these data were collected from the NOAA/NESDIS Global Vegetation Index (GVI) product developed from April 1985 (Kidwell 1997). The GVI is produced by the AVHRR–based 4-km global area coverage format (GAC) daily radiances in the visible (VIS, 0.58 μm to 0.68 μm), near infrared (NIR, 0.72 μm to 1.1 μm), and infrared (IR, 10.3 μm to 11.3 and 11.5 and to 12.5 μm), which have been quantized to 8-bit precision and mapped to a latitude / longitude grid (16 sq km). To get rid of cloud effects, these maps are composited over a 7-day period by storing radiances of the day having the largest difference between NIR and VIS (Kidwell 1997).

2.2 Vegetation Health Indices

The ecology and the weather are two components of NDVI. The first component of NDVI, the ecosystem, is primarily influenced by some slowly changing factors relating to environment including climate, soil, topography and vegetation type. These factors affect the amount and distribution of vegetation on Earth. On the other hand, the second component, namely, the weather component, is governed by a set of weather parameters such as rain temperature, wind etc. These parameters reflect the vegetation state and greenness in the annual cycle. Optimal weather condition is necessary for maximum vegetation to occur by allowing efficient use of ecosystem resources. On the other hand, unfavorable weather conditions can result in minimum vegetation characterized by mostly dryness because of fall in ecosystem resources. The dearth or

efficiency in ecosystem resources can also be characterized as extreme conditions which can be measured using the absolute maximum and minimum NDVI derived from data containing extreme weather events over a number of years. It is important to note that every geographic region on the earth contains specific level of ecosystem resources that indicate the ecosystem potential, which is also known as the carrying capacity of an area. The NDVI value can also measure the amount of vegetation signifying the carrying capacity of particular geographic region.

Temperature condition index (TCI) and vegetation condition index (VCI) are the vegetation health indices that describes the thermal and moisture conditions of a region. These indices are the components of BT and NDVI, respectively referred as random components (short time variation for weather fluctuation). For extracting the weather component from NDVI and BT three–channel algorithm has been developed by Felix Kogan, a NOAA scientist. The three-channel (VIS, NIR, IR) algorithm consists of processing of NDVI and BT, which includes complete removal of temporal high-frequency noise and detection of medium to low-frequency fluctuations in vegetation condition associated with weather variations (Kogan 1995, Kogan 1997, and Kogan 2000).

A statistical method was used to remove effects of varying transparency of the atmosphere (clouds, aerosols, etc.), viewing geometry (satellite and sun angles), bidirectional reflectance by smoothing NDVI and BT time series with a combination of compound medium filter and least square techniques (Kogan and Sullivan 1993). After smoothing, weather-dependent inter annual differences in NDVI and BT become apparent; NDVI is lower (reduced vegetation greenness) and BT is higher (thermal stress) in drought vis-à-vis wet and normal years. The assumption was that the maximum and minimum NDVI and BT observed for each pixel at a given time of a year

characterizes the extreme NDVI and BT based vegetation state associated with extreme weather impacts. These criteria were used to describe and classify weather-related ecosystems (Kogan 1997, Kogan 1995). Therefore, an important step in algorithm development was to single out weather components based on MAX-MIN criteria.

After the thresholds were set up, the NDVI and BT based vegetation condition and health were estimated relative to the MAX-MIN interval for each pixel and week. If NDVI is close to the MIN and BT to the MAX, conditions are said to be stressed; and for the opposite combination they are favorable. This was formalized by the following three indices: vegetation condition index (VCI). Temperature condition index (TCI), and vegetation and temperature index (VH), an average of VCI and TCI. These indices describe moisture and thermal and vegetation health conditions, respectively. NDVI and BT interpret oppositely extreme weather events (for example, in case of drought, the NDVI is low and BT is high due to both vegetation deterioration and higher contribution of a soil signal; conversely, in a no drought year, the NDVI is high while BT is low).

$$VCI = \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \times 100 \quad (1)$$

$$TCI = \frac{BT_{\max} - BT}{BT_{\max} - BT_{\min}} \times 100 \quad (2)$$

$$VT = a.VCI + b.TCI \quad (3)$$

In order to facilitate early detection of drought and examine its impacts on crop and pasture production around the world, reflectance measurements were found to be very effective. (Liu and Kogan 2002, Domenikiotis et al.2004). The VHI can be used as proxy data for monitoring vegetation health, drought, moisture and thermal conditions, fire risk, greenness of vegetation cover, vegetation fraction, leaf area index, start and end of the growing season, crop and pasture productivity, teleconnection with ENSO, desertification, mosquito-borne diseases, invasive species, ecological resources, land degradation, etc.

2.3 Mosquito, Malaria and Climate in Bandarban, Bangladesh

The physical and cultural environments in Bangladesh permit the survival of at least two malarial parasites, *Plasmodiam malariea* and *Plasmodiam falciparum*, and three vectors *A. philippinesis*, *A. sundacious*, and *A. minimus* during most of the year (Dutt et al. 1980).

Malaria is one of the biggest health concerns in Bangladesh. The country has 4180-km border with India on the west, north and eastern territory and a 190-km border with Myanmar in the south-east. Twelve million (8% of the total population of Bangladesh) people are at high risk of malaria. Out of a total 64 administrative districts, 13 (Figure 2.1) (http://www.bangladeshgov.org/bdmaps) are in the malaria endemic areas and about 98% of the total malaria morbidity and mortality reported from Bangladesh each year originate from these districts (Rahman et al. 2006). These districts are extremely populated (27 million) (Rosenberg and Maheswary 1982) which is one fifth of total population of Bangladesh. Bangladesh has sub-tropical warm, wet and humid climate with 140 million population (Pampana 1969, Rosenberg and Maheswary, 1982). In the endemic malaria areas, annual rainfall is 3000 mm, temperature

25^0 C and humidity 85%. There are two malaria seasons in Bangladesh: April-October with wet and warm conditions and December-February with cool and dry conditions. Chittagong Hill tracts (CHT) are the districts (three districts) with the highest level of incidence of Malaria. In 2001, 67% (37180) of the countrywide malaria cases occurred in this area, which has only 1% of country population (Ingrid et al. 2004). I examine the regional pattern of the incidence of malaria at Bandarban hill tracts (CHT), a district in Bangladesh for the period from 1992-2005 and describe the changing characteristics.

These districts receive more than 2000 mm of Rainfall annually which contributes to a high relative humidity throughout the year particularly in the high sun season when the relative humidity stays above 70%. The hilly areas are mostly covered with tropical evergreen forests that provide the considerable shade needed for breeding grounds of *Anophylies*.

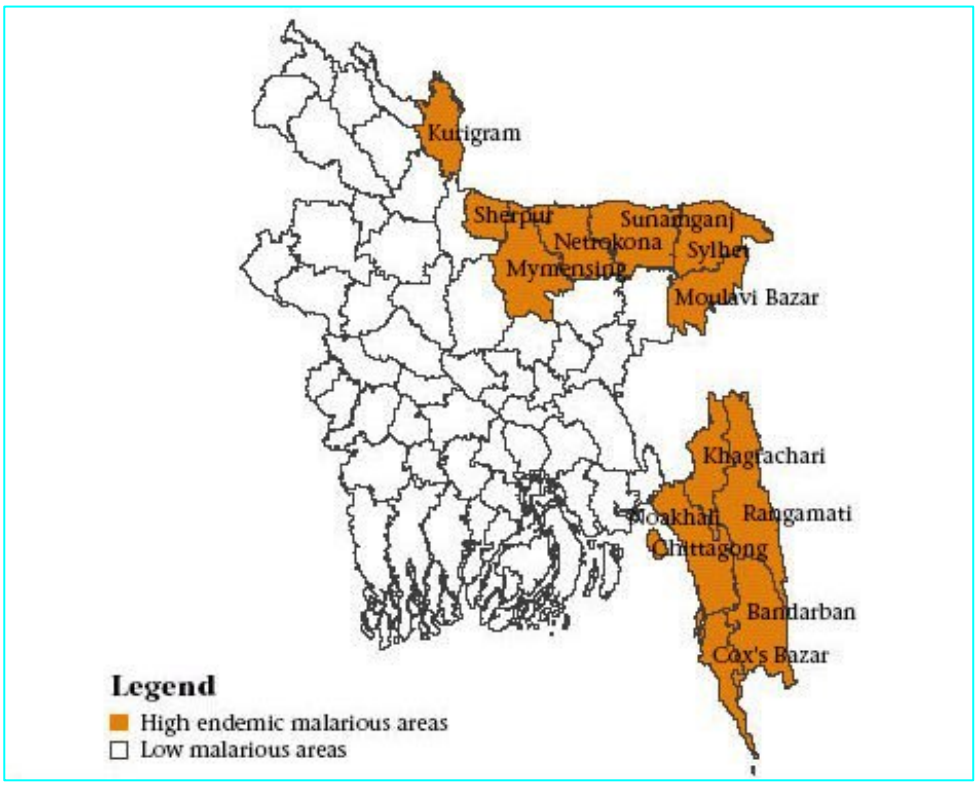


Figure 2.1: Distributions of Malaria in Bangladesh

Most precipitation falls during April-October, when average daily temperature is around 25º C with maximum temperature reaching 30º C during daytime (Faiz et al. 2001). In April-October wet, warm and humid climate is quite favorable for mosquito activity that would increase in the number of malaria cases.

Malaria parasite (plasmodium) in Bandarban is transmitted by female *Anopheles* mosquitoes (Elias and Rahman 1987, Rosenberg and Maheswary 1982). Mosquitoes in Bandarban transmit malaria throughout the year. However during the cooler season (November-March) mosquitoes are less active and the number of malaria cases is relatively small (Rosenberg and Maheswary 1982). This number increases considerably during warm and wet season (Faiz et al. 2001). The study area is Bandarban district, south eastern border of Bangladesh, bounded between 22^0N-22.5^0N latitude and 91.5^0E - 92.5^0 E longitudes, a hilly area mostly covered with tropical evergreen forest.

2.4 Mosquito, Malaria and Climate in India

Of 1.4 billion people in Southeast Asia, living in 11 countries, 1.2 billion are at risk of Malaria (Ashwani et al. 2007). According to an estimate (Kondrachine 1992) as high as 85.7% of these people (around 1.2 billion) are exposed to the risk of malaria, most of whom living in India only. Malaria was the prime cause for an estimated loss of 1.85 million life years in India alone out of a total 4.2 million disability–adjusted life years lost attributable to all vector borne diseases (Kumar et al. 2007). The Malaria epidemiology in India is complex because of geo-ecological

diversity, multi-ethnicity, and wide distribution of nine *Anopheline* vectors transmitting three *Plasmodium* species (http://www.malariasite.com). Sharma (1996) has pointed out that the entire geographical area of India is vulnerable to the spread of malaria with only exceptions of those parts of the nation that are near the coastal area and have an elevation of 1800 meters or more.

Rao (1984) identifies six main vectors of malaria in India. These are *Anopheles culicifacies, An.stephensi, An.dirus, An.uviatilis, An.minimus* and *An.Sundaicus* . Four other vectors are identified to have secondary importance for malaria in India, namely, *An.annularis, An.varuna, An.jeyporiensis* and *An.philippinensis*. These Malaria vectors in India are classified as: *Anopheles culicifacies*, a rural vector; *Anopheles stephensi*, an urban vector; *Anopheles fluviatilis,* prevalent in the foothills; *Anopheles sundaicus* in the Andaman and Nicobar Islands; and *Anopheles minimus and Anopheles dirus* in forest hill (Operational Manual for Implementation of Malaria Program 2009). The distribution of various Malaria vectors (Figure 2.2), diversity and environmental conditions result in variable endemicity of Malaria.

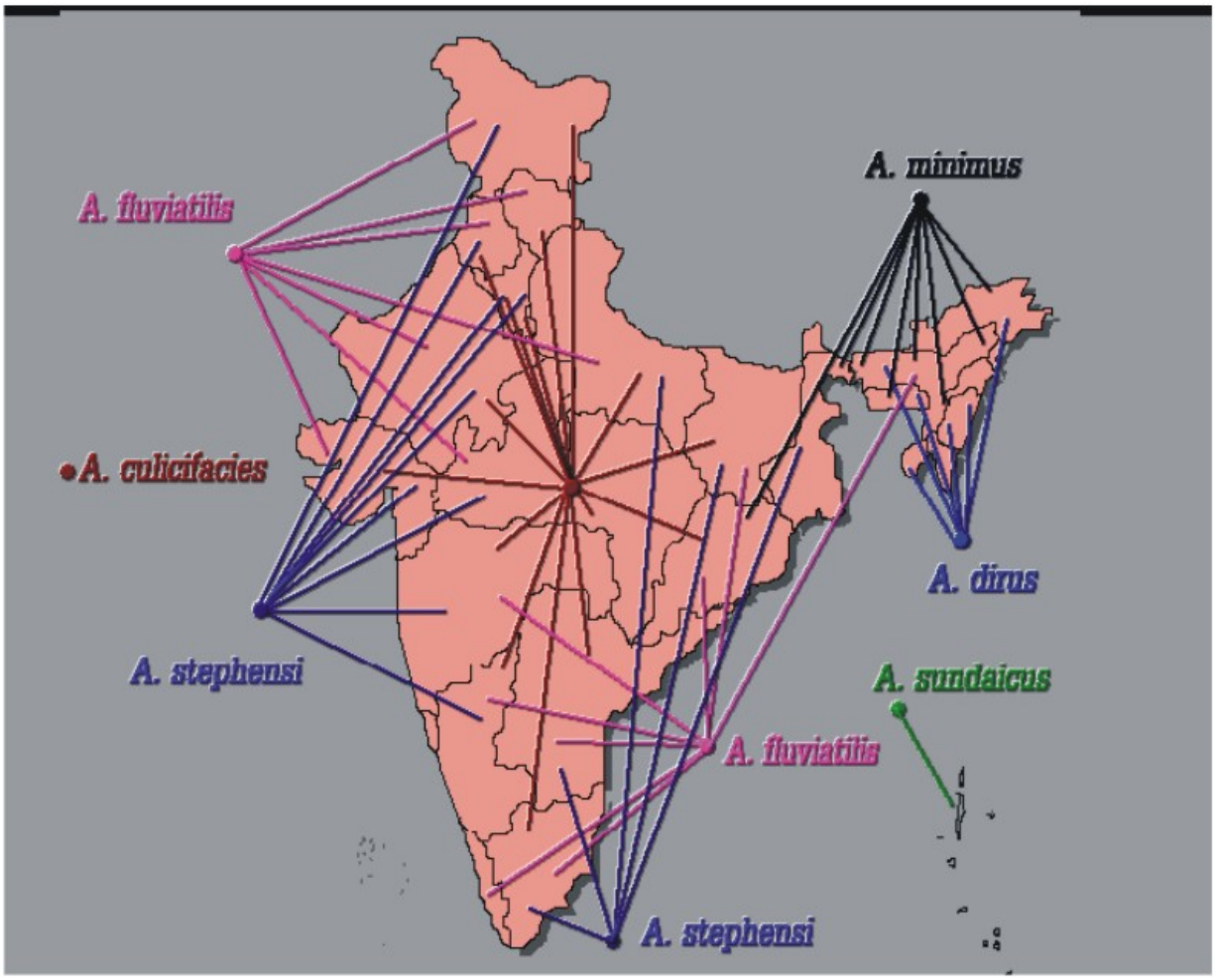


Figure2.2: Principal vectors of malaria in India (Operational Manual for Implementation of Malaria program 2009)

Apart from elevations above two kilometer and in some coastal places, Malaria is endemic in all regions of India (Sharma 1996). In India, Orissa has contributed the most to Malaria. (Source National Vector Borne Disease Control Program (NVBDCP), India). Figure 2.3 shows District wise distribution of malaria during 1997.

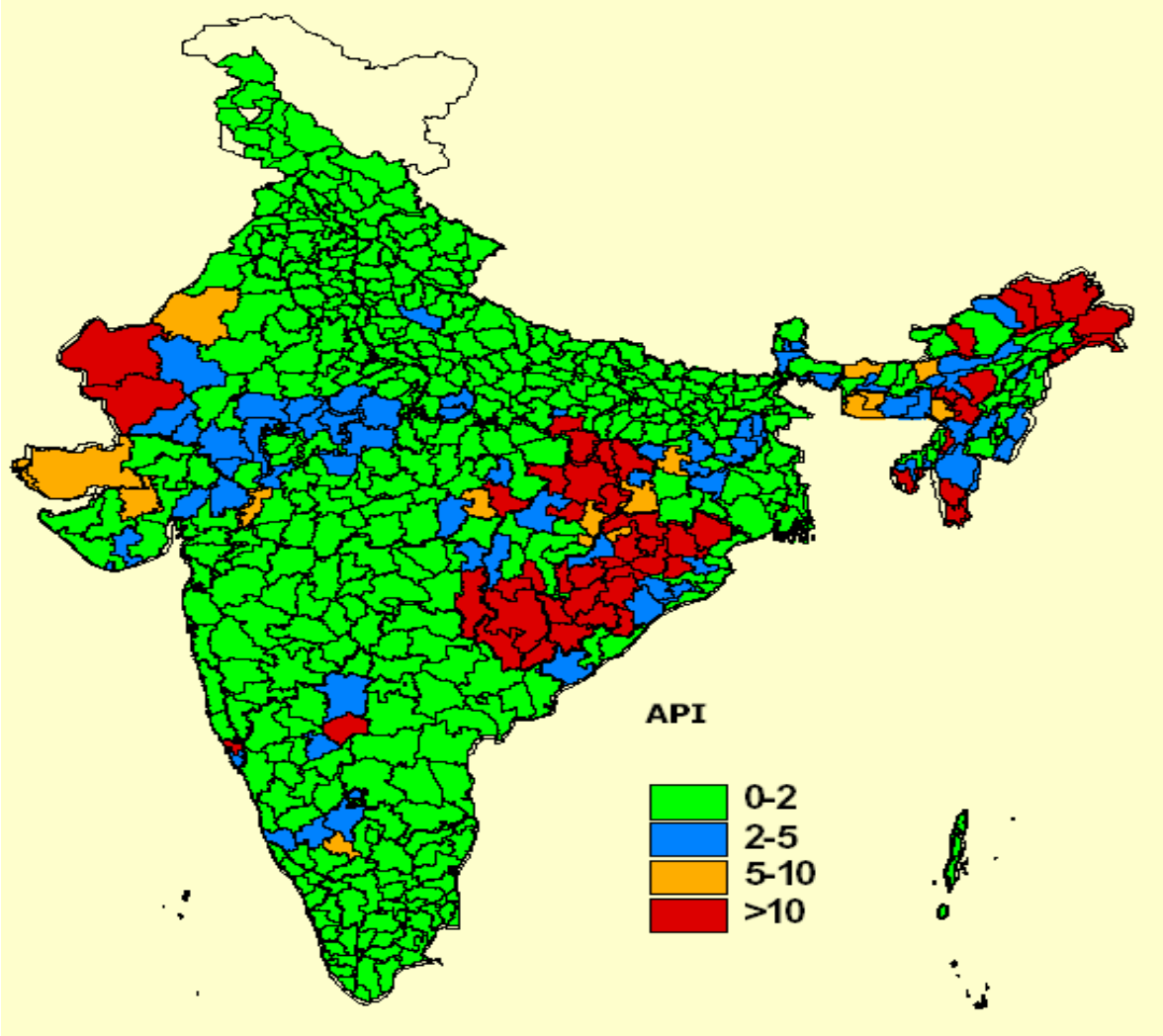


Figure 2.3: District wise distribution of malaria in India 1997(Malaria country profile, India 1997- 2007)

Environmental conditions are keys to understanding the malaria transmission because the parasite, the mosquito, has to develop and progress through a cycle to transmit malaria. The window of opportunity for the mosquito to spread malaria will likely extend over one - three months. This is largely because low temperatures (below 18 °C) are not suitable for transmission. During winter months November to February the density of malaria

vectors in rural places (*Anopheles culicifacies)* and urban places (*Anopheles stephensi)* is also relatively low.

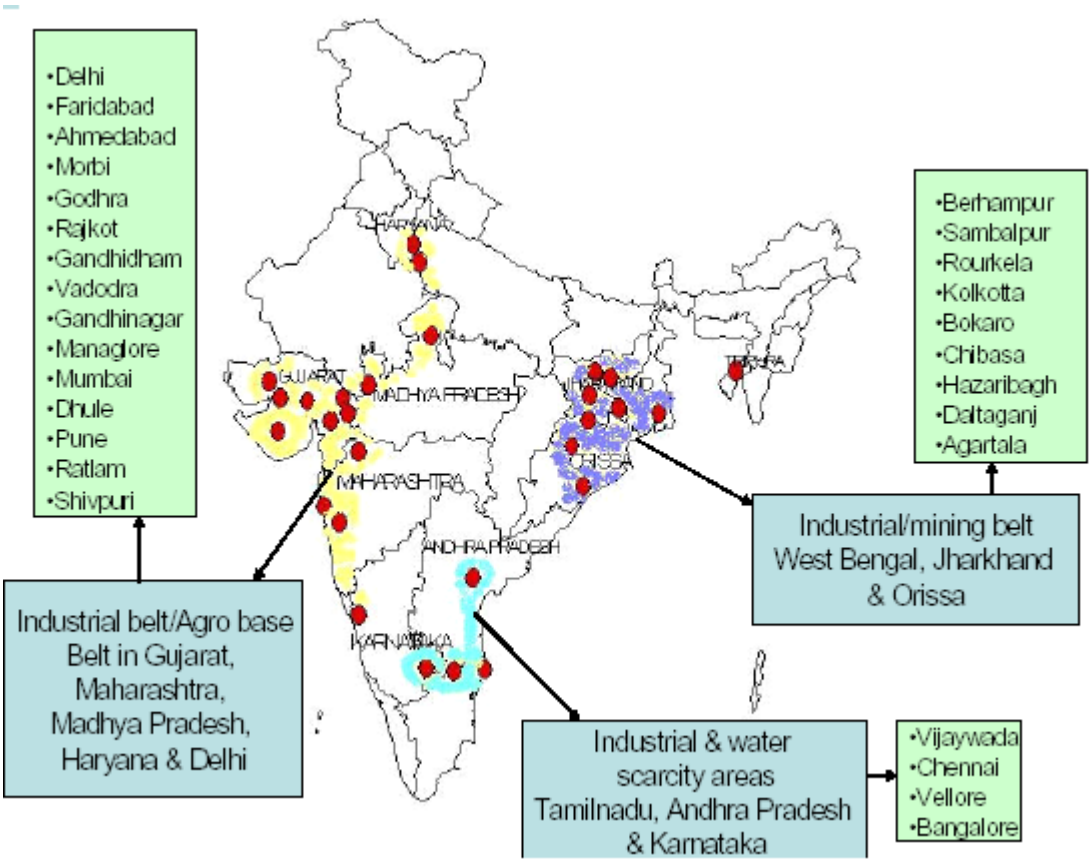
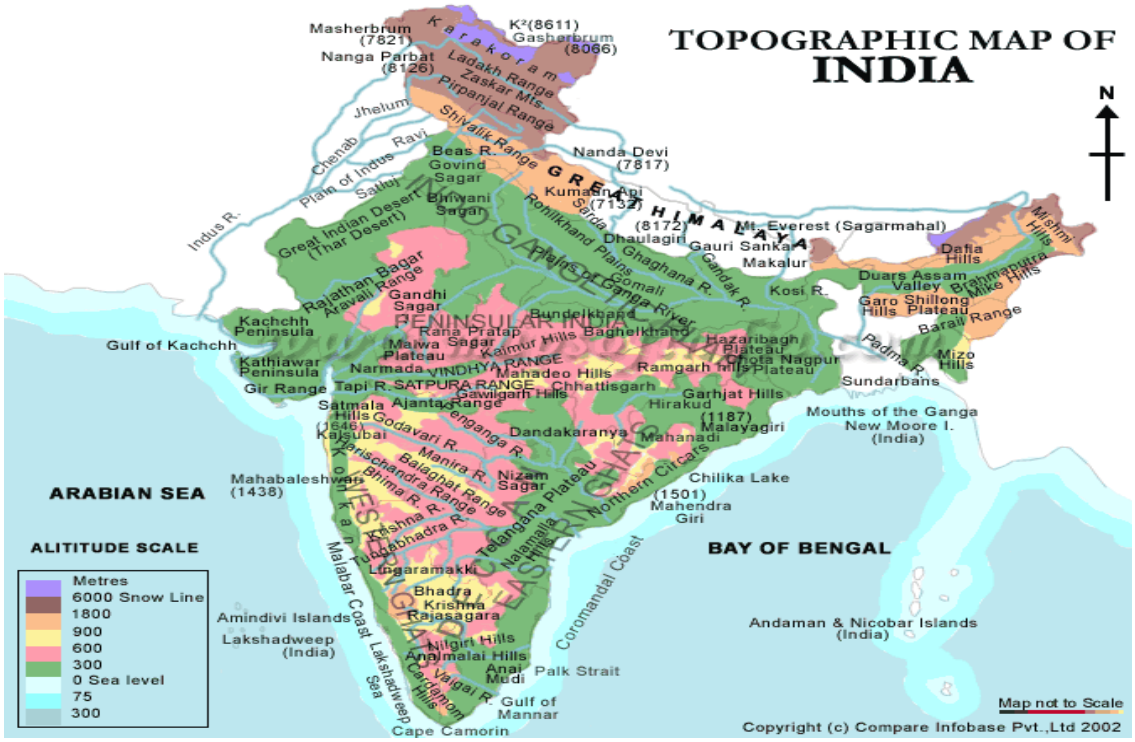


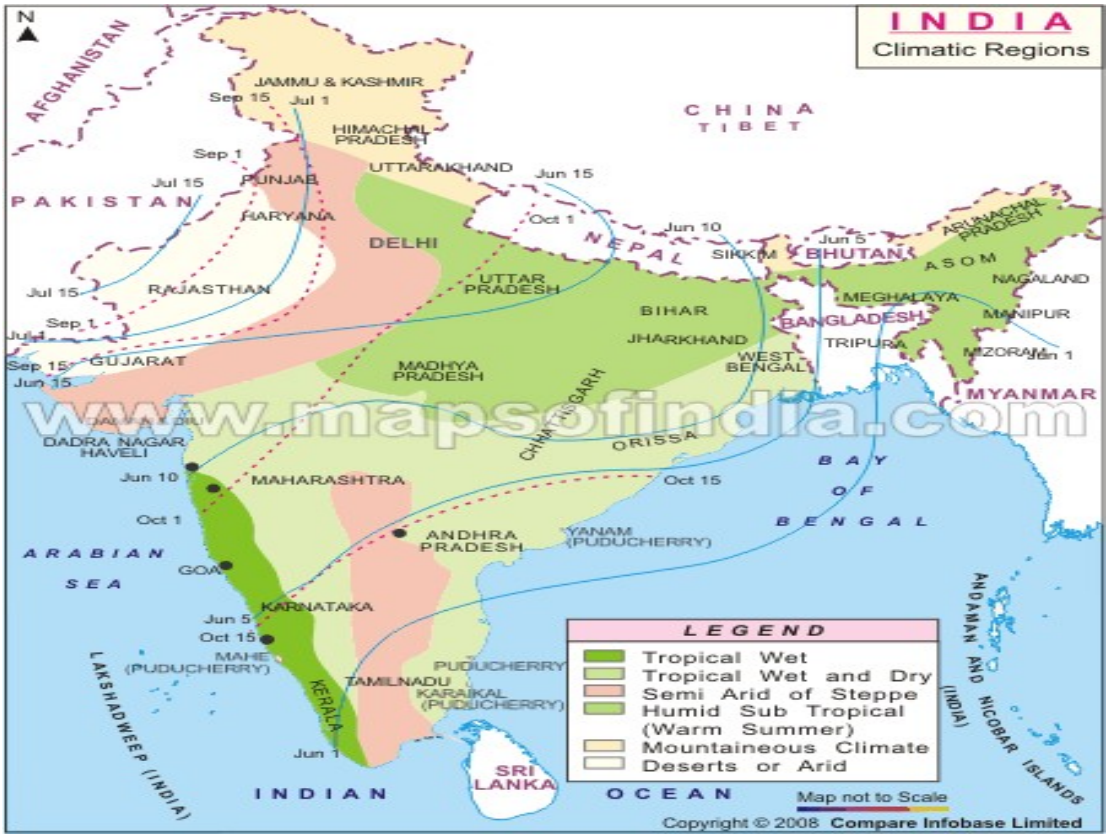
Fig 2.4: Malaria high risk towns in India (Malaria country profile 2005-2007)

India is a tropical country and it has three primary climatic classifications. The humid and semi-humid places are the west coast, the Himalayas, and the north-eastern region around Bangladesh. These places receive a lot of rain from the monsoon from the southwest. The sub-humid places include the states of West Bengal, Orissa, Madhya Pradesh, and the northern parts of Andhra Pradesh, Punjab, and Uttar Pradesh, and a narrow coastal strip of Tamil Nadu. The arid and semi-arid regions include Gujarat, Rajasthan, parts of the Himalayas, and the center of the country (Takahashi and Arakawa 1981). Figure 2.5

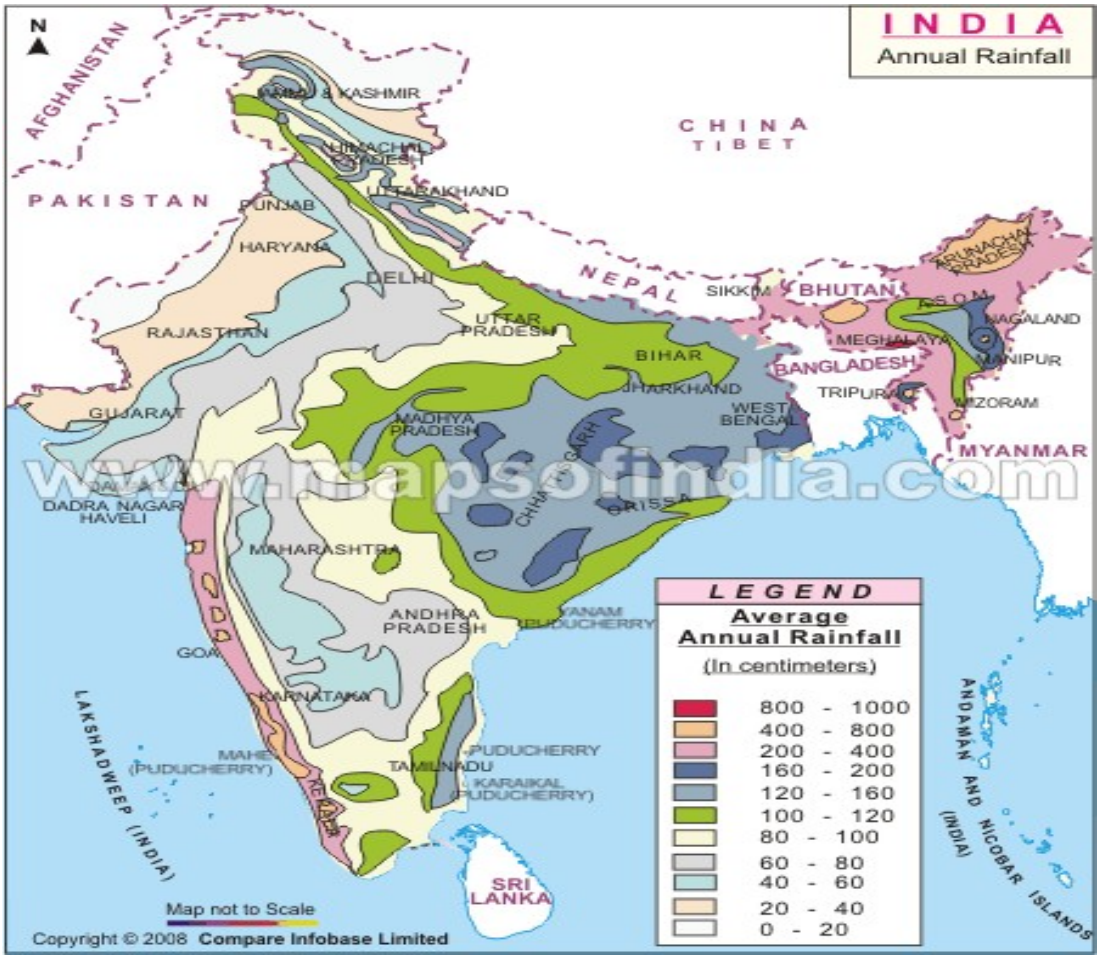
shows topographic map, climatic regions, yearly rainfall and yearly temperature in various regions in India. The four regions (Rajsthan,Gujarat, Orissa and Tripura) I chose in India can be differentiated in their ecosystem, rainfall, temperature, and other weather conditions. These maps (Figure 2.5) allow us to analyze and justify seasonal behavior of malaria transmission with its regional moisture and thermal condition. Furthermore, behavior of different types of vectors and parasites can also be detected through thermal and moisture condition. The analysis has been presented in the result and discussion (chapter four) section for each particular region.



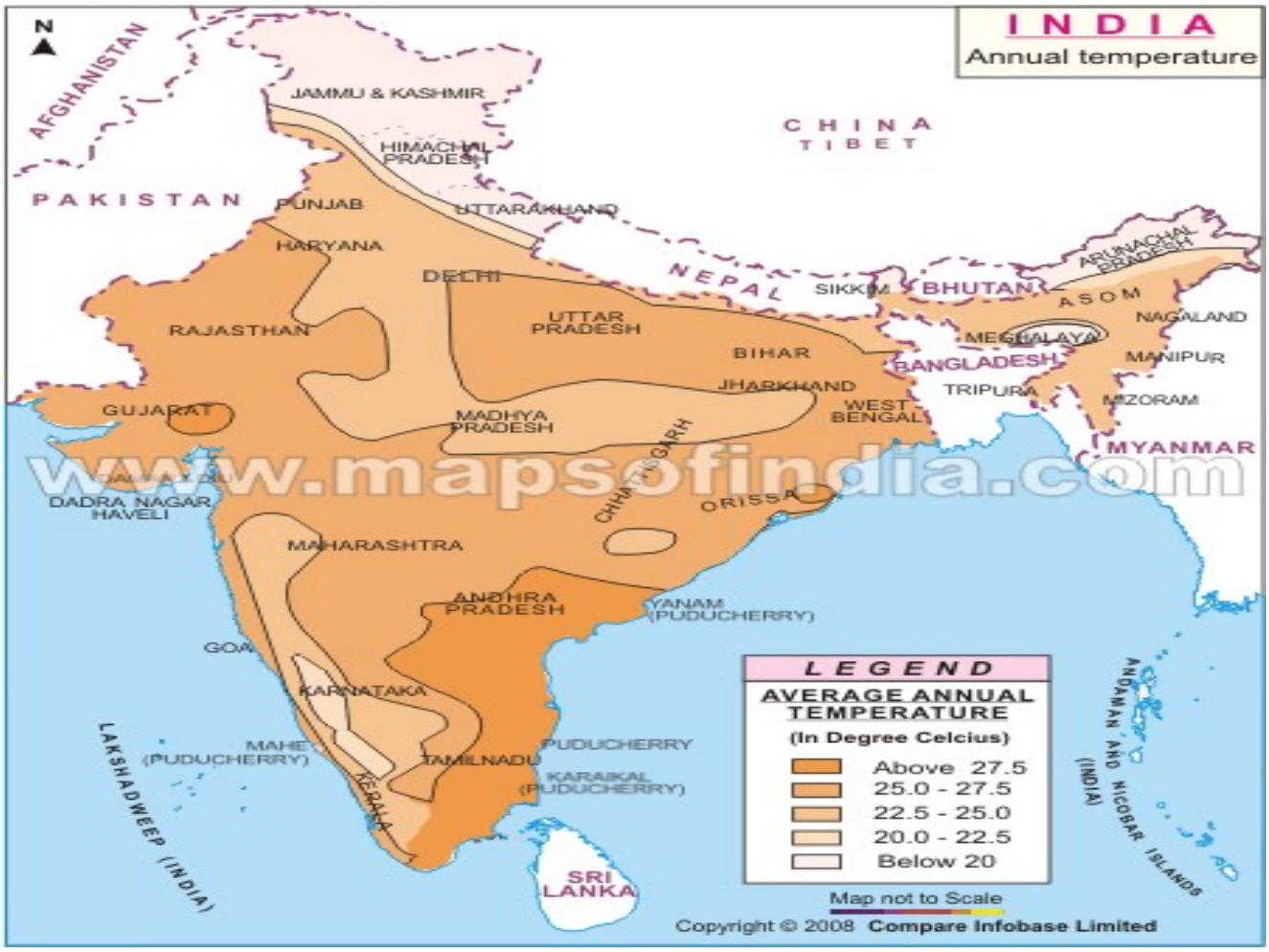
(a)


INDIA
Climatic Regions
N
AFGHANISTAN
PAKISTAN
CHINA
TIBET
NEPAL
BHUTAN
SIKKIM
BANGLADESH
MYANMAR
SRI LANKA
JAMMU & KASHMIR
HIMACHAL PRADESH
UTTARAKHAND
PUNJAB
HARYANA
DELHI
RAJASTHAN
UTTAR PRADESH
BIHAR
JHARKHAND
WEST BENGAL
ARUNACHAL PRADESH
ASOM
NAGALAND
MEGHALAYA
MANIPUR
TRIPURA
MIZORAM
GUJARAT
MADHYA PRADESH
CHHATTISGARH
ORISSA
DADRA NAGAR HAVELI
MAHARASHTRA
ANDHRA PRADESH
YANAM (PUDUCHERRY)
GOA
KARNATAKA
MAHE (PUDUCHERRY)
TAMILNADU
PUDUCHERRY
KARAIKAL (PUDUCHERRY)
KERALA
LAKSHADWEEP (INDIA)
ANDAMAN AND NICOBAR ISLANDS (INDIA)
ARABIAN SEA
BAY OF BENGAL
INDIAN OCEAN
Jul 1
Sep 15
Jul 15
Sep 1
Jun 15
Oct 1
Jun 10
Jun 5
Jun 1
Oct 15
LEGEND
Tropical Wet
Tropical Wet and Dry
Semi Arid of Steppe
Humid Sub Tropical (Warm Summer)
Mountaineous Climate
Deserts or Arid
Map not to Scale
Copyright © 2008 Compare Infobase Limited
www.mapsofindia.com

(b)


INDIA
Annual Rainfall
N
AFGHANISTAN
PAKISTAN
CHINA
TIBET
NEPAL
BHUTAN
SIKKIM
PUNJAB
HARYANA
DELHI
RAJASTHAN
UTTAR PRADESH
BIHAR
JHARKHAND
WEST BENGAL
GUJARAT
MADHYA PRADESH
BANGLADESH
MEGHALAYA
TRIPURA
MIZORAM
MYANMAR
DADRA NAGAR HAVELI
MAHARASHTRA
ANDHRA PRADESH
GOA
KARNATAKA
KERALA
TAMILNADU
PUDUCHERRY
LAKSHADWEEP (INDIA)
SRI LANKA
ANDAMAN AND NICOBAR ISLANDS (INDIA)
LEGEND
Average Annual Rainfall
(In centimeters)
800 - 1000
400 - 800
200 - 400
160 - 200
120 - 160
100 - 120
80 - 100
60 - 80
40 - 60
20 - 40
0 - 20
Map not to Scale
Copyright © 2008 Compare Infobase Limited

(c)



(d)

Figure 2.5: a) Topography map of India, b) Climatic region of India c) Average annual rainfall in India d) Average annual temperature India

2.4.1 Malaria Dynamics in Bikaner, Rajsthan, India

Bikaner, a part of the Thar desert area, has an area of an area of 27,000 square km and is

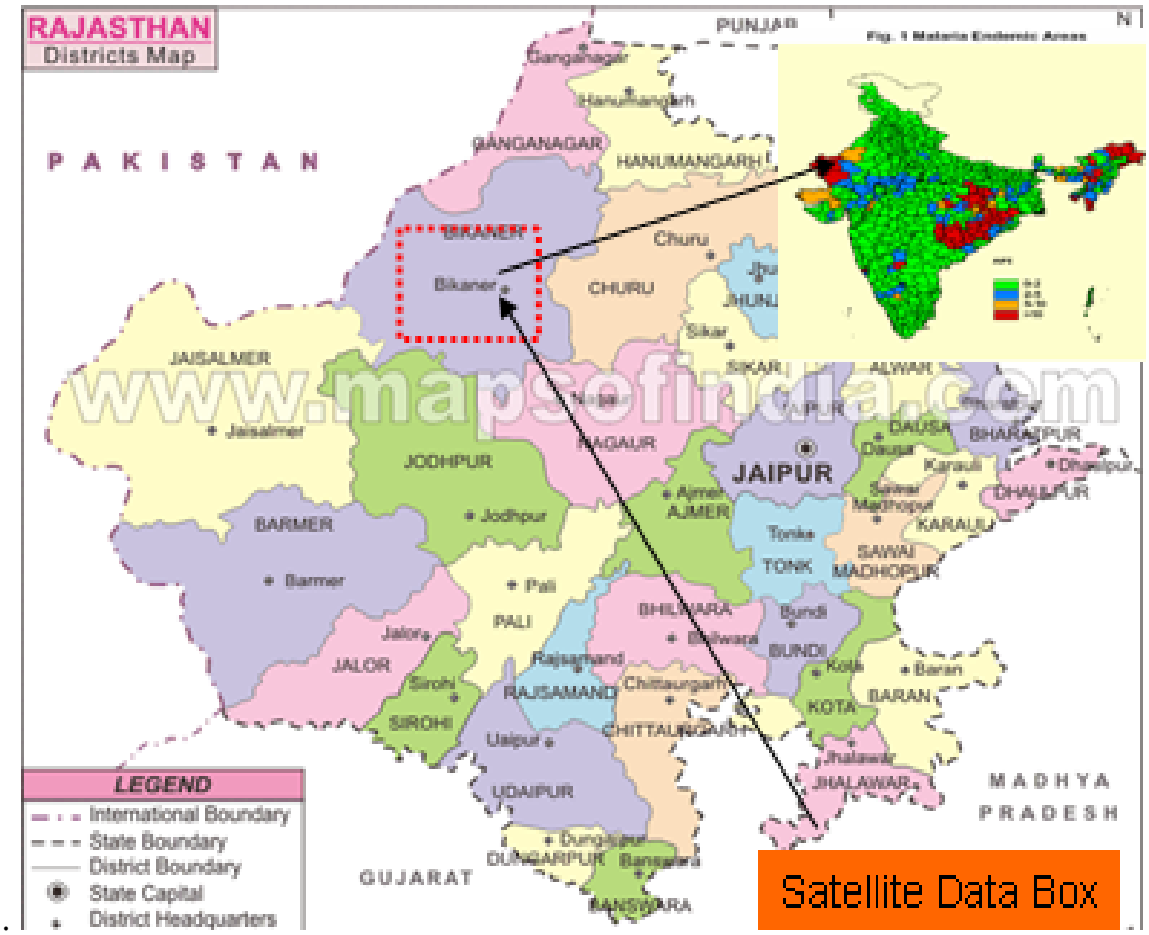


Figure 2.6: District Map of Rajsthan, District Bikaner is pointed with the rectangular box where Satellite data collection has been collected.

located between 71°54' and 74°12' east and latitude 27°11' and 29°3' north. It experiences great variation in terms of temperature ranging from 2° Celsius during the winter to as high as of 48° Celsius during the summer. The variation in temperature in the region results in very hot summer and extremely cold winter with yearly rainfall as low as 250 mm. Bikaner receives most of the rainfall between the months of July and September, commonly known as the rainy season. As documented in Kochar et al. (2007), two

species of Anopheles mosquitoes, namely, *Anopheles culicifacies* and *An.stephensi*, are very common in Bikaner and account for the most of the malaria transmission in this region. Kochar et al. (2007) use 17 years of malaria epidemiological data spanning between1990 and 2006 and offer a comprehensive trend analysis. Incidence of malaria transmission is the highest immediately after the rainy season and the lowest during January and February.

2.4.2 Malaria Dynamics in Tripura, India

Tripura generally experiences sub tropical weather, typified by warmth and high humidity. The state's three seasons are summers, winters, and monsoon. Summers extend from the end of March to May, when the temperature soars to the maximum of 35°C. In late-May, pre-monsoon showers soak the region (www.iloveindia.com/states/tripura/weather). Monsoon season formally arrives in June and continues until September. Heavy downpours are common and their average rainfall is 2100 mm.

The winter season lasts from November to February and the average minimum temperature is 10.5°C. Compared to the plains or urban areas, Malaria incidence is quite high in and around the forest. Tripura is hyperendemic for Malaria with perennial transmission and has more than one operating vectors. Both *P falciparum and P vivax* are present, but *P vivax* predominates (Ghose et al. 2006**)**. Figure 2.7 shows satellite data collection area since most of the malaria occurs in this area as seen from Figure 2.3 and Figure 2.4.

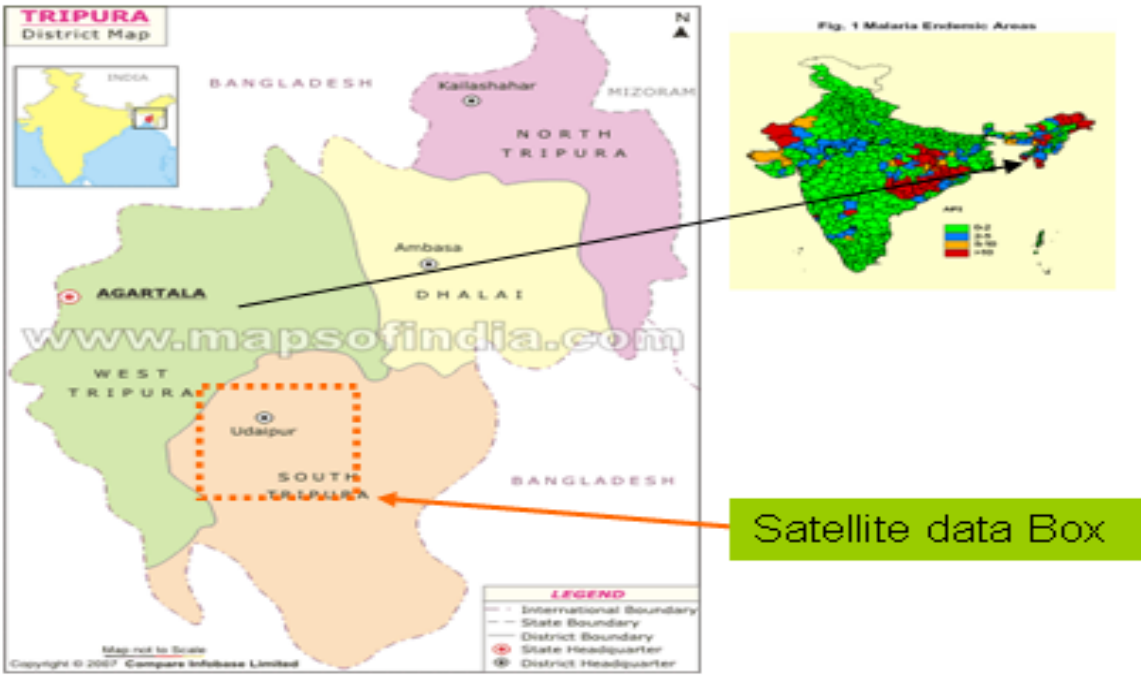


Figure 2.7: District Map of Tripura, rectangular box is pointed where Satellite data has been collected

2.4.3 Malaria Dynamics in Gujarat, India

Malaria incidence in Gujarat state has been on a general decline since 1989 (Srivastava and Yadav 2000). Nevertheless, in some tribal villages in forested places of Valsad district, southern Gujarat, there was a malaria outbreak in September 1995. In the absence of health agencies in or near affected villages, the malaria parasite load continued to build up leading to an outbreak towards the end of monsoon season (Srivastava 2000). Most arid and semi-arid places in western India fall in an unstable Malaria zone in the nation. Apart from monsoon, in other months the area remains generally dry. The Malaria affected villages were either on the bank of hill-streams or rivers, which stabilized post-monsoon and provided high vector breeding ground (Srivastava 2000).

Southern Gujarat is the most malarious region, comprising the districts of Valsad, Dangs, Surat, Bharuch, Vadodara, Kheda and Panchmahal. Gujarat has a tropical climate - hot summers and cold winters. Summer is from April to June and temperatures range from 27^0 to 42^0 C. Winters are more or less conducive to Malaria, with a temperatures variation of 14^0 to 29^0 C. Monsoons usually begin in June and last till September. Figure 2.8 shows satellite data collection area where most of the malaria occurs in this area as seen from Figure (2.3) and Figure (2.4).

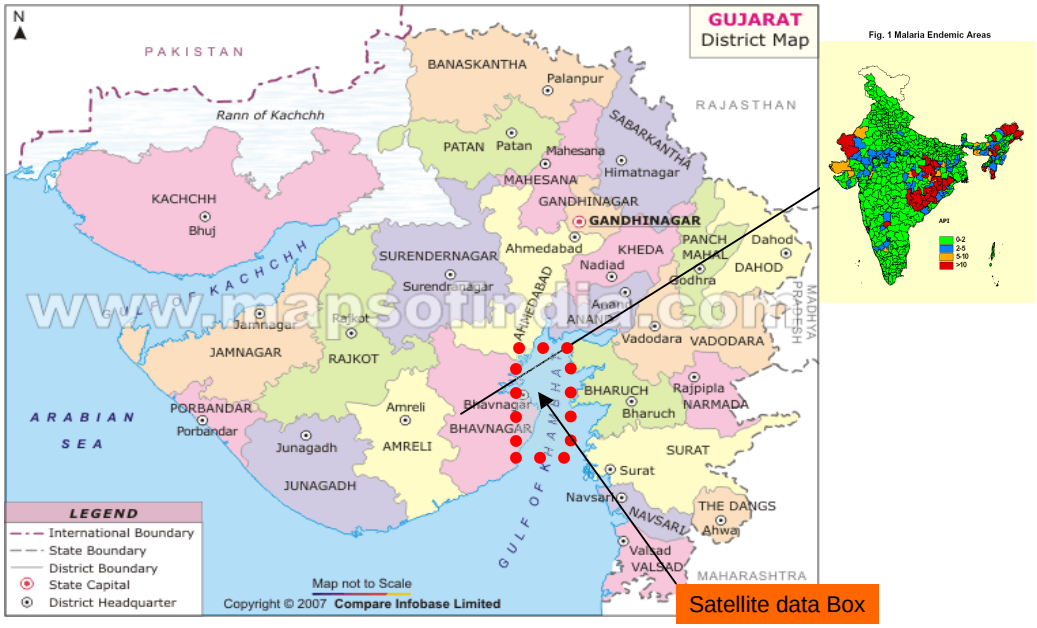


Figure 2.8: District Map of Gujarat, the rectangular box where Satellite data collection has been performed.

2.4.4 Malaria Dynamics in Orissa, India

Malaria has become increasingly serious during the past 10 –15 years, and it is currently the number one public health problem in Orissa. The magnitude of the problem can be

viewed in terms of 1998, when Orissa (with a share of less than 4 % of all-India population), accounted for 28.6 per cent of the detected cases of Malaria in all of India (two million), and 62.8 per cent of all Malarial deaths in the country, according to the National Malaria Eradication Programme Report for 1998 (Government of India 1999). Kandhamal, Keonjhar, Sundargarh, and Mayurbhanj districts of Orissa account for 50% of all Malarial instances and 40% of the Malarial deaths in the state (Orissa Voluntary Health Association 1995, p.25). Orissa and other forested areas in India primarily inhabited by ethnic tribes have experienced the highest incidence of mortality attributable to malaria. (Kumar et al. 2007).

There are three major seasons in Orissa- very hot summer (March-June), rainy season (July-September) and the winter (October-February). It is warm almost throughout the year in the Western districts with maximum temperature hovering between 40-46°C and in winter temperature goes down to 8°C . The climate is equable but highly humid and sticky. The average rainfall is 1500 mm, experienced as the result of south west monsoon during July-September. The month of July is the wettest and the major rivers often get flooded. The state also experiences small rainfall from the retreating monsoon in the months of October-November. January and February are dry.

Nanda et al. (2000) reports that the empirical studies have demonstrated existence of unpredictability in malaria endemicity in forested and broken forested villages. In the beginning of August, *An.fluviatilis* becomes thickly dense and reaches to the peak in December. The intensity of malaria declines in January and reaches to its lowest level between May and July. In contrast to the behavior of *An.fluviatilis*, the concentration of *An.culicifacies* is high between March and September with its peak falling between

March and April. The incidence of *An.culicifacies* is generally low during November to February. Various relevant studies have shown that these vectors of malaria pulsate alternatively and maintain recurrent spread in forested villages.

2.5 Introduction to Malaria Transmission in South America

2.5.1 Malaria Dynamics in Brazil

In the Americas and in the Caribbean, 38% of the population (308 million people), in 21 countries, live in areas with malaria transmission, with a mean of 1.3 million cases per year. Thirty-six percent of these are in Brazil (WHO 2002). Since 1970, when just over 52,000 cases were recorded, malaria has gradually been increasing in this country. In the 1990s, the number of cases surpassed 500,000. In 1999, there were 610,000 notified cases of malaria in Brazil, and 99% of them were in the Amazon region (Ministério da Saúde Brazil 2003). Administrative, topographic, and vegetation map are shown in Figure 2.9 A. Our study areas in Brazil are department of Rondonia, and Ampa. The general topography of Rondônia state, situated in the southwestern Amazon region, is flat, with an average elevation of 300 m above sea level. The climate is characterized as tropical, with a long rainy season from January till May. There is also some rainfall during other months of the year. Rondônia is mostly covered with near-complete lowland rainforest and sparsely populated (Willem et al. 2005). The climate of Ampa is characterized as tropical rainforest with a rainy season, in which 85% of the precipitation generally occurs, and a dry season. From December to March it rains several times every day with

high intensity but short duration; from April to June the rain is continuous and heavy; and from July to November it is dry season (Póvoa et al. 2001).

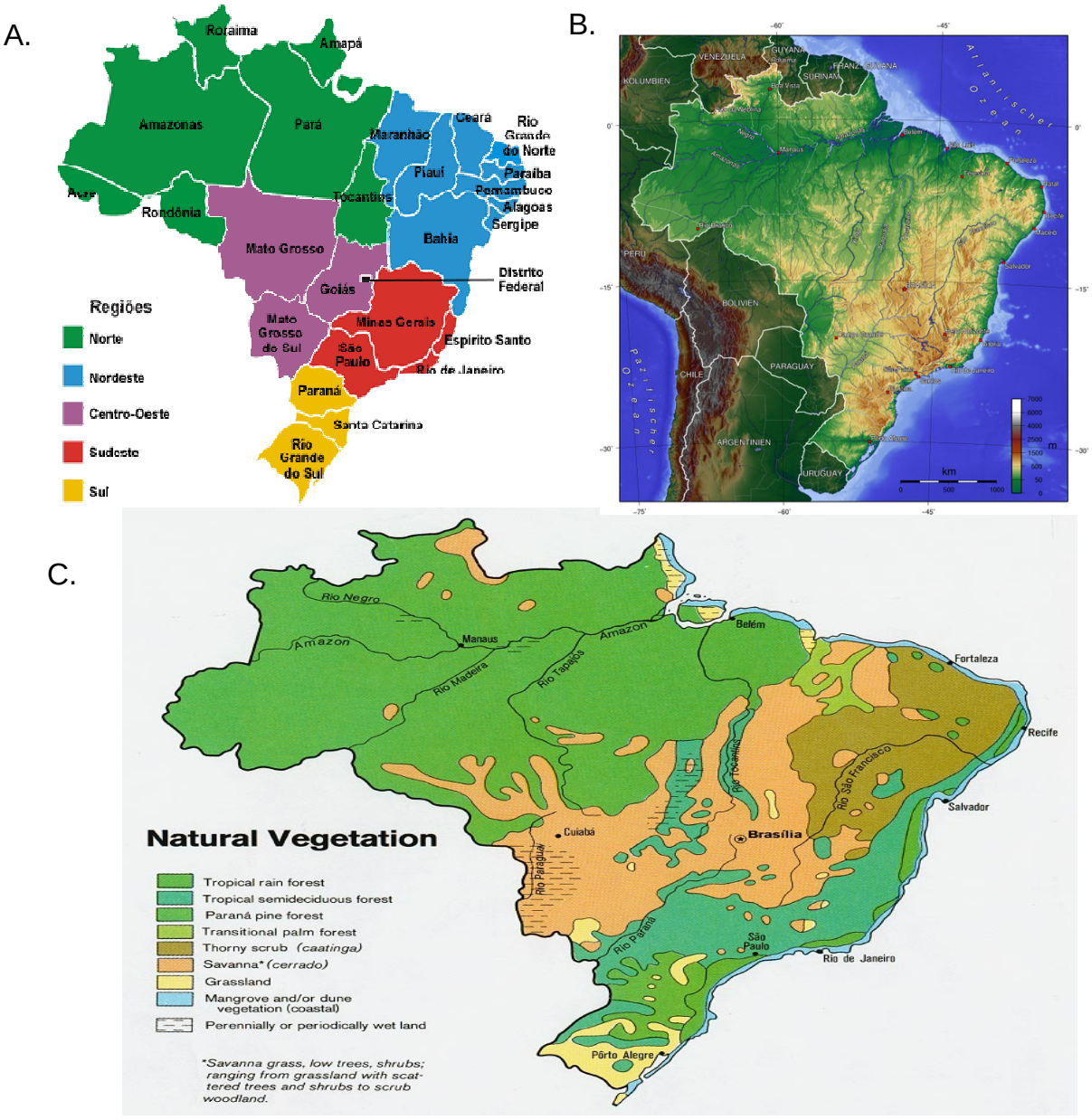


Figure 2.9 A. Departmental map of Brazil B. Topography map of Brazil C. Vegetation map of Brazil

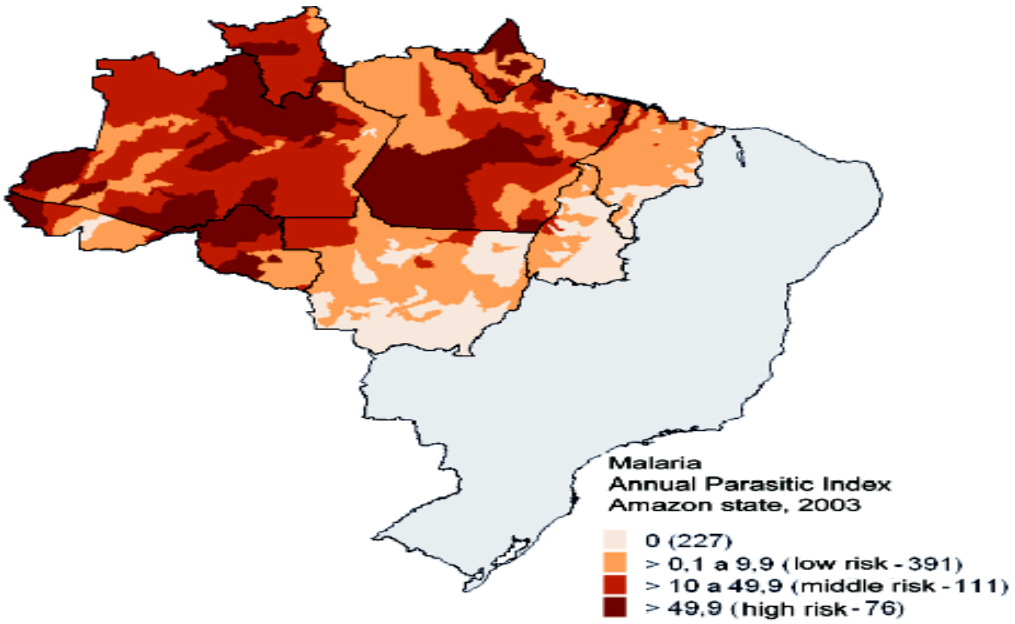


Fig. 1: malaria risk areas in Amazon states, Brazil, 2003. Source: Sivep-Malária. Ministry of Health.

Figure 2.9 B. Malaria risk area in Brazil (Ministry of Health Brazil)

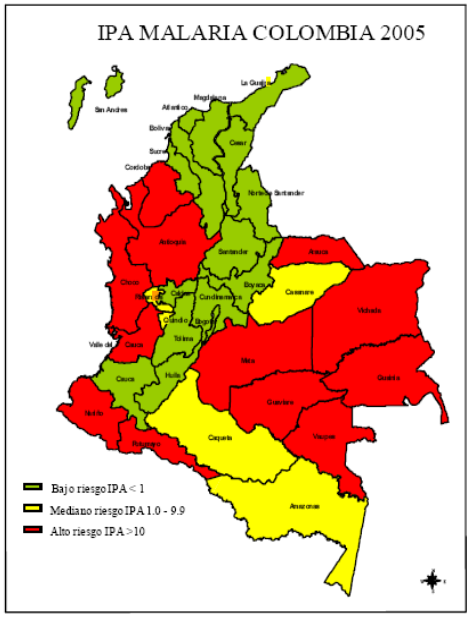
2.5.2 Malaria Dynamics in Narino, Colombia

More than five million people in Colombia live in endemic malaria regions. The most important malaria vectors in the country are *Anopheles albimanus*, *Anopheles darling*i, and *Anopheles nuñeztovari* that transmit *Plasmodium falciparum* (46.5%) and *Plasmodium vivax* (53.5%), and rare cases (8–10 per year) of *Plasmodium malariae*. The geographical distribution of malaria in Colombia is associated with prevalent climatic conditions. The study is conducted on the Pacific coast (Narino Department) of Colombia where both *P. vivax* and *P. falciparum* are transmitted in approximately similar proportions throughout the year. In this region, there are peaks in malaria transmission between April and May and between September and October. Figure 2.10 shows, A.

Administrative map of Colombia, B. Malaria risk area in Colombia, C. Climatic region of Colombia

A.

B.



C.

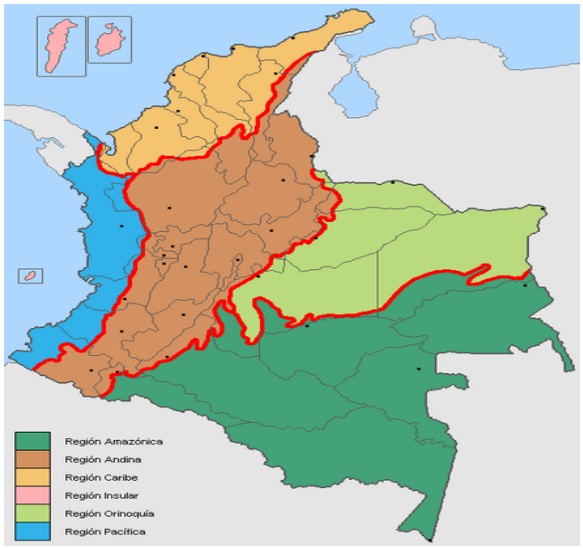


Figure 2.10 A. Administrative map of Colombia B. Malaria risk area in Colombia C. Climatic region of Colombia

2.6 Performance Indicator for malaria

Malaria control program have been traditionally monitoring fever surveillance, malaria case load and other disease burden and impact indicators. For this research I used slide positive rate (SPR) which is repeatedly mentioned as percent of malaria in this thesis. I use Annual Parasite Incidence (API) only for Rondonia, and Ampa in Brazil since the Ministry of Health of Brazil does not publish number of blood smear examined.

Annual Parasite Incidence (API) indicates disease burden & impact population used for areas to be brought under spray

$$\text{Annual parasite Incidence (API)} = \frac{\text{Number of blood smears positive for malaria Parasite in a year}}{\text{Total population}} \times 1000$$

Slide Positive rate (SPR) indicates Disease burden and impact and it is independent of surveillance activity, therefore a better indicator for impact assessment

$$\text{Slide Positive rate (SPR)} = \frac{\text{Number of blood smears positive for malaria Parasite in a year}}{\text{Number of Blood smears examined}} \times 100$$

The transmission potential, endemicity levels, vectors of malaria and other factors of malaria differ from area to area. The multiple vectors require different intervention strategy for malaria control.

CHAPTER THREE

Methodology

Malaria (slide positive rate) time series data have been separated in two components; short term (random function) variation which is associated with weather related fluctuation (temperature, precipitation, and relative humidity) and long term (slowly changing function) variations, for example, environmental factors (ecosystem, desertification, sea level rise, changing vegetation, and agricultural practices), and the human population (migration, spread of drug resistance, change in immune status, and spread of pathogens into new areas). In my research I use time series of malaria statistics to extract weather related variations and compare with the weather components of satellite data that have been extracted from NDVI and BT time series discussed in chapter two. The weather components of NDVI and BT are expressed as vegetation health (VH) indices (Kogan 1997). I attempt to investigate the intensity of the relationship and find if there are significant correlations between weather components of malaria data and satellite data before the major transmission seasons of the malaria for the particular ecosystem of that area. In order to explain the methodology, results for malaria in Bandarban district has been presented. Two data types were used in this study: malaria statistics, and satellite data. The methods of data collection and data process have been discussed in chapter two.

3.1 Malaria Time Series

Following Brockwell and Davis (2000) malaria time series (Figure 3.1 A) for Bandarban district, Bangladesh was approximated by equation 3.1. The weather-related variations around the trend were expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 3.3).

$$Y_t = T_t \bullet DY_t \tag{3.1}$$

Where T_t *is a* slowly changing function representing the deterministic component of trend that is regulated by ecosystem, malaria eradication program, and socio economic background, and the DY_t is a random component regulated by weather fluctuations.

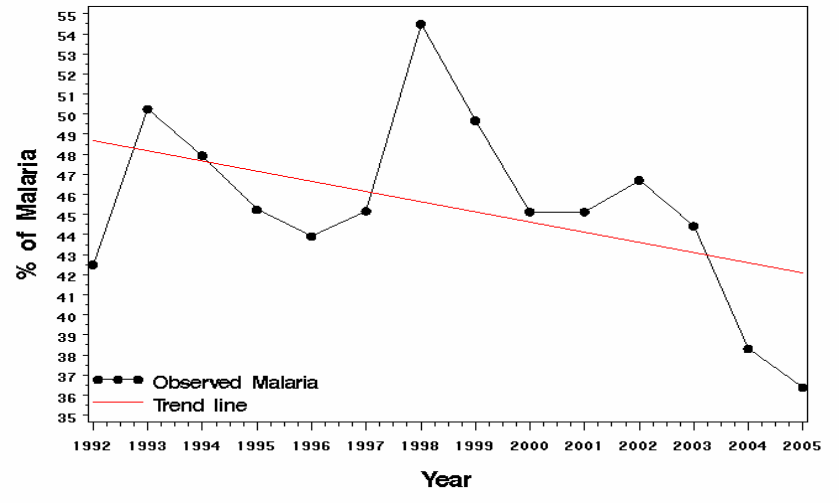
We use a linear approximation for T_t . The parameters in linear equation 3.2 are shown for the malaria time series in Bandarban district, Bangladesh

$$Y_{trend} = 1059.95 - 0.51* \text{Year} \tag{3.2}$$

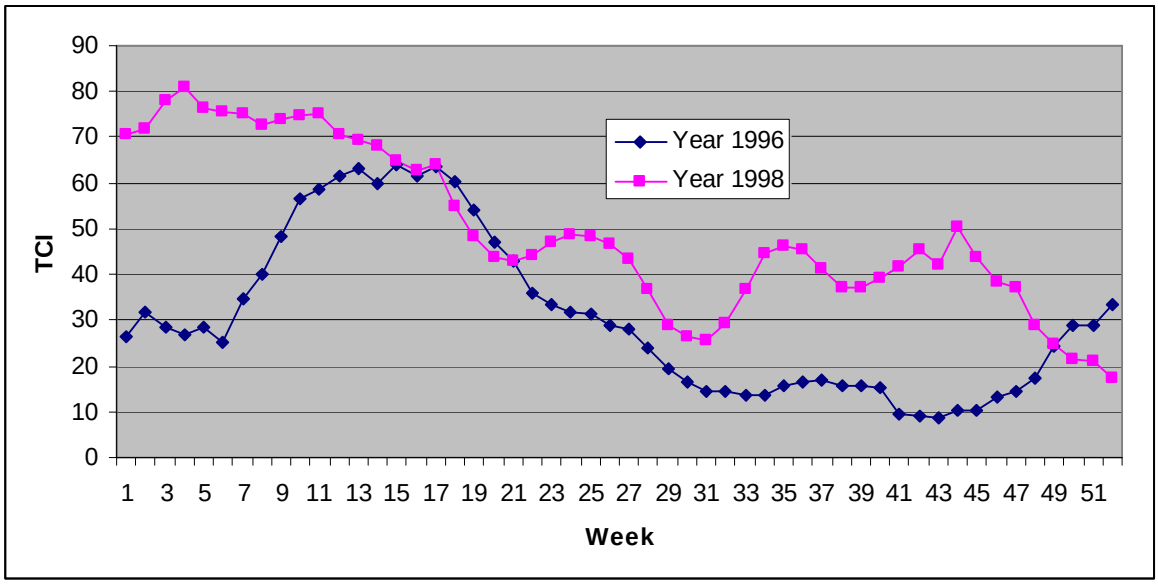
The random component (DY_t) was expressed as

$$DY_t = (Y / Y_{trend})*100 \tag{3.3}$$

Where Y is actual percent of malaria cases, Y_{trend} is the malaria estimated from trend; Year is in year number; *DY* is deviation (%) from the trend. DY is a good relative metric of malaria cases.



(A)



(B)

Figure 3.1: (A)Percent of malaria in Bandarban district, Bangladesh and trend line (B) Temperature condition index (TCI) for 52 weeks during 1996(unfavorable year for malaria) and 1998 (favorable)

From Figure (3.1 A) it can be demonstrated that in 1998, *DY* was only 119% or 19% above the trend, whereas in 1996, DY was 87% or 13% below the trend. These estimations indicate that the 1996 (less % of malaria) was an unfavorable year for malaria transmission whereas 1998 (higher % of malaria) was favorable malaria transmission year. We can also see from Figure (3.1 B) that the values of temperature condition index (TCI) week 32- 36 during 1996 which are smaller (DY is also smaller in these weeks) where as during 1998 TCI for the same weeks are larger (DY is also larger in these weeks). These phenomena explain that during week 32-36 there are consistently positive relations among TCI and malaria incidence. Furthermore it is important to emphasize that more malaria cases are associated with cooler conditions (TCI>60) when vegetation is under favorable condition, than during drought years when vegetation is under stress (TCI<40). However, for better understanding correlation dynamics between DY and TCI, VCI for all 52 weeks have been calculated.

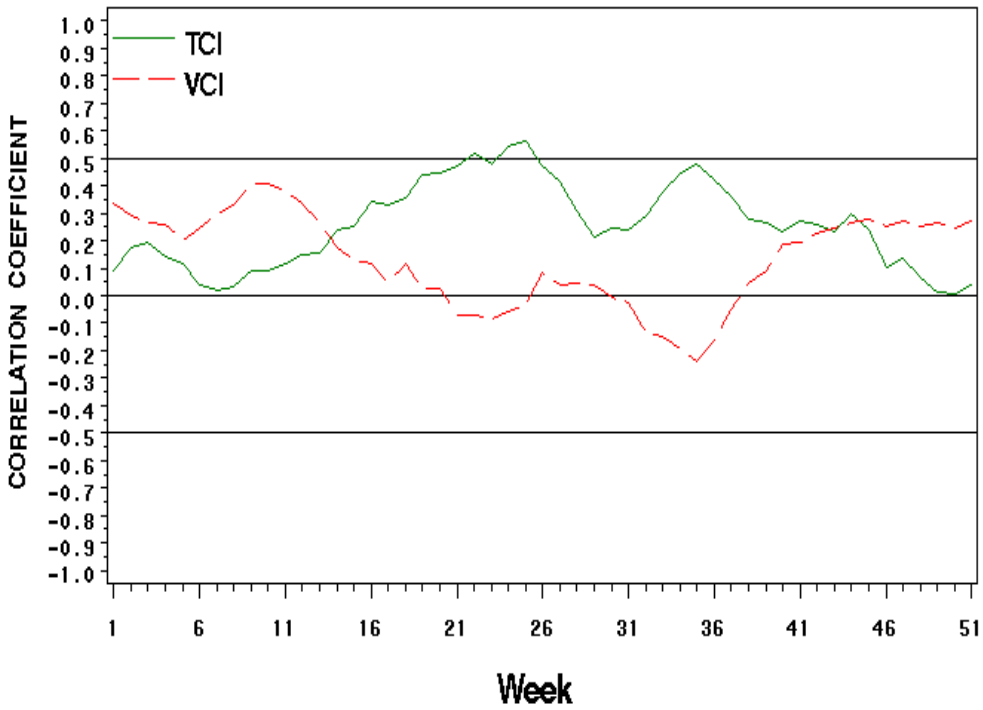


Figure 3.2: Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI

3.2 Correlation Dynamics

Once I was able to extract the detrended value DY (deviation from climatology) I compute the correlation dynamics between VH indices (which is also expressed as deviation from climatology) and DY (detrended value from malaria time series).

The weekly TCI that had the largest number of statistically significant correlations with high R values (0.55) during summer are not random but can be explained in terms of the

seasonal climate of Bandarban district. The summer (May and June) is typically the two hottest months of the year. July and August are two months of rainy season and December and January are the two coldest months. One would expect that variation in these months strongly affects malaria activities. To illustrate, if the cold months are too cold or the hot months too hot, vectors and parasites cannot survive. The correlation analysis of trend malaria cases (DY) versus VCI and TCI, are shown in Figure 3.2. It should be emphasized that during cooler month (spring and fall) when mosquitoes are less active; correlation is low for both indices. In spring (week 16) when mosquito activity season starts, correlation for Bandarban district starts to increase reaching maximum (0.5 for TCI and 0.4 for VCI) at the end of June (week 25-26). In week 31 correlation for Bandarban district starts to increase again reaching maximum (0.55 for TCI) at the middle of September (week 35) TCI. These results are compatible with mosquito's activity in Bandarban district. After laying eggs in April-May, new female mosquitoes reach adult stage in 7-15 days. After that their biting activity increases, leading to enhanced malaria transmission. Due to larva being flushed away by too much rainfall during July and August malaria transmission decreases, and increases again during September and October because of stagnant water creates more breeding places. The association of TCI with malaria incidence can be explained as follows. Because of more rain (higher precipitation causes more breeding places) more cloud covers the area resulting in the reduction of the temperature, which ultimately increases TCI. Thus increasing TCI is associated with increases malaria transmission in this region. Although the correlation of DY with VCI during the same wet period is weaker than with TCI, the dynamics of correlation coefficient (CC) in Figure 3.2 rightly emphasizes seasonal cycle

(increase of CC at the beginning of the warm season, maximum CC, during the critical period and decrease of CC during the end of the warm season). Since correlation dynamics for VCI is not significant enough (Figure 3.2) we shall not use VCI for further analysis for this area.

Further examination includes regression analysis of these deviations (DY) to develop a model of the association between malaria cases and thermal and moisture conditions at Bandarban district in Bangladesh. I shall consider the week 32 through week 36 as potential predictor variables for further analysis for forecasting malaria transmission.

3.3 Multiple Linear Regression Analysis

In order to see if multiple linear regression analysis would be applicable I estimated correlation dynamics which provides a preliminary feel for the simple relationship between malaria incidence and satellite derived thermal condition index (TCI) of certain weeks of the year. Following the correlation analysis, the result of fitting the ordinary least squares (OLS) model is approximated by equation 3.4, for the second breeding periods during July and August (week 32-36).

$$DY=b_0 + b_1TCI_{32}+b_2TCI_{33}+b_3TCI_{34}+b_4TCI_{35}+ b_5TCI_{36} \quad (3.4)$$

Table 3.1. Results of multiple linear regressions (OLS) of DY on the equation (3.4), Bandarban district Bangladesh R^2=0.88

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	86.226	3.11859	27.65	<.0001	.	0
TCI_{32}	-1.1168	0.59254	-1.88	0.0962	0.01203	83.12697
TCI_{33}	-0.00702	1.19173	-0.01	0.9954	0.00307	326.03281
TCI_{34}	0.22426	1.27245	0.18	0.8645	0.00254	393.64865
TCI_{35}	2.33509	0.81987	2.85	0.0215	0.00617	162.00663
TCI_{36}	-1.26114	0.61886	-2.04	0.0759	0.01051	95.18932

Table 3.1 shows coefficients of equation 3.4 that the value of R^2 is large 0.88 but p-values for the regression coefficients is high (in this case 0.99, 0.86 for coefficient of TCI week 33 and 34), which is not significant at $p<0.05$ level. Very small tolerance with very high variance inflation for the regression coefficient indicates lack of stability for the model. These kinds of problems are the natural consequence of multicolinearity which will produce misleading results (Ryan 1997).

3.3 Multicollinearity

The high degree of correlation (linear dependency) among predictor variables is known as multicollinearity. Multicollinearity occurs when a large number of independent variables are used in a regression model where some of them may measure the same concepts or phenomena. According to OLS assumptions predictor variables must be orthogonal. So existence of multicollinearity is a violation of OLS assumptions. It is also implied that existence of multicollinearity can produce wrong signs of the estimated coefficients of predictor variables (Ryan 1997).

It is important to note that independent variables with high correlation coefficients do not necessarily imply multicollinearity. We need to check related statistics, such as tolerance value or variance inflation factor (VIF), eigenvalues, and condition number. Low tolerances correspond to high VIF since variance inflation factor (VIF) is just the reciprocal of a tolerance value. Thus, VIF shows how multicollinearity has increased the instability of the coefficient estimates (Freund and Littell 2000: 98). Another way to diagnose multicollinearity is to investigate correlation matrix among predictor variables (Ryan 1997).

3.4 The Correlation Matrix

Here I try to detect multicollinearity looking at the table of correlation matrix. Using this method I can easily Figure out weather predictor variables (VH indices) are completely independent or the predictor variables of the neighboring weeks are highly correlated as in the Table 3.2

Table 3.2. Correlation Matrix among Weekly TCIs, Bandarban district Bangladesh

	TCI32	*TCI33*	*TCI34*	*TCI35*	*TCI36*
TCI32	1	0.9901	0.9759	0.9688	0.9704
TCI33	0.9901	1	0.9935	0.9847	0.9825
TCI34	0.9759	0.9935	1	0.995	0.9915
TCI35	0.9688	0.9847	0.995	1	0.9938
TCI36	0.9704	0.9825	0.9915	0.9938	1

For example, the correlation coefficient between TCI32 with TCI33 (Week 32 and 33) is 0.99. This high correlation among the independent variables is an indication of

collinearity. Because of existence of severe multicollinearity, ordinary least square (OLS) can produce unbiased estimators only at a potentially high cost of estimator variability (Gunst and Mason 1980, Chatterjee et al. 2000). Therefore, I choose an alternative forecasting model that will considerably reduce multicollinearity.

3.6 Principal Component Analysis

The necessity for transforming the data arises as the original variables or the model in terms of the original variables violates one or more of the standard regression assumptions. In this case existence of multicolinearity among explanatory variables has caused the violation of one of the OLS assumptions.

In order to remove collinearity the variables in OLS model equation 3.4 are decomposed and transformed into a new set of orthogonal or uncorrelated variables called principal component of the correlation matrix. This transformation sets the new orthogonal variables in order of their importance. The matrix of the original predictor variables in OLS model is decomposed into eigenvalues, eigenvectors (loadings), principal components PC and residuals.

3.6.1 Eigenvectors of the correlation Matrix

Table 3.3 shows the eigenvectors for each of PCs. These coefficients, which relate the components to the original variables in the first column, are scaled so their sum of squares is unity. This allows for finding which of the original variables dominates a

component. The coefficients of the first PC show positive relationships with all variables with somewhat larger contributions from TCI34 (0.4487), TCI33 (0.4483) TCI35 (0.4475) and TCI32 (0.444) as seen from Table 3.3. The second component is dominated by TCI32 (0.73). Since eigenvectors are components of original predictor variable, this Table 3.3 is also necessary for recalculating the coefficient of original variable after the completion of principal component regression. This transformation is mathematically explained in the following section.

Table 3.3: Eigenvector of correlation matrix Bandarban district Bangladesh

	Prin1	*Prin2*	*Prin3*	*Prin4*	*Prin5*
TCI32	0.444186	0.734379	0.377374	0.242585	0.249254
TCI33	0.448321	0.291132	-0.484485	-0.314843	-0.616764
TCI34	0.448793	-0.202352	-0.485925	-0.173533	0.701
TCI35	0.447566	-0.425024	-0.022899	0.749047	-0.239674
TCI36	0.447186	-0.392858	0.621463	-0.500841	-0.092894

3.6.2 Principal Component Regression

Principal component regression (PCR) has been used for forecasting even in the presence of multicollinearity with smaller prediction error. PCR method has the ability to estimate regression coefficient more precisely and efficiently than the predictions that are generated using data other than those used for estimations (Draper and Smith 1981, Raymond 1986).

Initially the predictor variables in model equation 3.4 are transformed into a new set of orthogonal uncorrelated variables called principal component of the correlation matrix. This transformation scaled the new orthogonal variables in order from high to low

variance which involves eliminating some of the PCs to get reductions in variance. In other words, we don't need to use all the PCs for further regression analysis. We can also use another method (discussed in the next section) to eliminate some of PCs other than just choosing first couple of PCs which have higher variances. Now, regression analysis of the response variable DY against the reduced set of predictors (PCs) was performed using OLS method. These new regression coefficients are principal component estimators (Gunst and Mason 1980). Finally these coefficients of newly performed regression analysis are mathematically transformed into a new set of coefficients (using eigenvector Table 3.3) correspond to the original correlated set of variables in model equation. The following mathematical expressions explain how the total transformation occurs (generally) in principal component regression.

This model equation (3.4) can be expressed as

$$Y = \beta_o + \sum_{i=1}^{n} \beta_i X_i + \varepsilon \tag{3.5}$$

Where DY= Y, TCI_j = X_j n= number of weeks (independent variables). Let $\overline{y}$ and $\overline{x}_j$ be the means of Y and X_j and let

$$S_j = \left(\sum_{i=1}^{n} \left(x_{ij} - \overline{x}_j \right)^2 \Big/ (n-1) \right)^{1/2} \tag{3.6}$$

and

$$S_y = \left(\sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2 \Big/ (n-1) \right)^{1/2} \tag{3.7}$$

be the standard deviations of the response and Jth predictor variable, respectively. Equation (3.6) can be written in terms of standardized variables as

$$\tilde{Y} = \theta_1 \times \tilde{X}_1 + \theta_2 \times \tilde{X}_2 + \theta_3 \times \tilde{X}_3 + ... + \theta_m \times \tilde{X}_m + \varepsilon' \qquad (3.8)$$

where

$$\tilde{Y} = \frac{\left(y_i - \bar{y}\right)}{s_y} \qquad (3.9)$$

is the standardized version of response variable (dY) and

$$\tilde{X} = \frac{\left(x_{ij} - \bar{x}_j\right)}{s_j} \qquad (3.10)$$

is the standardized version of Jth predictor variable.

The estimated coefficients satisfy

$$\beta_j = (s_y / s_j)\theta_j \qquad (3.11)$$

j= 1,2,……..n

$$\beta_o = \bar{y} - \beta_1 x_1 - \beta_2 x_2 - ..... - \beta_n x_n \qquad (3.12)$$

The principal components of the standardized predictor variables are given by

$$Z_j = \sum_{i=1}^{n} c_{ij}\theta_i \qquad (3.13)$$

j= 1,2,……..n

where c_{ij} are elements of the eigenvectors of the matrix of bivariate correlation between pairs of explanatory variables. The model in equation (3.8) may be written in terms of principal components as

$$\tilde{Y} = \alpha_1 \times Z_1 + \alpha_2 \times Z_2 + \alpha_3 \times Z_3 + ... + \alpha_m \times Z_m + \varepsilon'$$

(3.14)

where α_j's and θ_j's are related as

$$\alpha_j = \sum_{i-1}^{m} c_{ij}.\theta_i, \tag{3.15}$$

$$\theta_j = \sum_{i-1}^{m} c_{ij}.\alpha_i, \tag{3.16}$$

3.6.3 Percent Variation Accounted for By Principal Component

Instead of looking at the variance to reduce Principal Component (PCs) we could also choose the number of components in the model by specifying the percent of variance that we want to be explained, either in the predictors or in the model effects.

Table 3.4: Percent Variation accounted for By Principal Components

Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	98.7710	98.7710	16.3950	16.3950
2	0.8759	99.6469	42.7172	59.1122
3	0.2287	99.8756	11.5443	70.6564
4	0.0976	99.9731	11.3525	82.0090
5	0.0269	100.0000	0.2211	82.2301

For example in our case from Table (3.4) one can see that first four components are better to use since these four components can extract 82% of the information from predictor variables. In case of only two components a total of 59% information can be extracted from predictor variables.

3.7 Cross Validation and Choosing the Number of Factors

Validation of a fitted regression equation is used to see the capability of the model to be used for future predictions. Validation of the model is to assess the effectiveness of the fitted equation with an independent set of data. R square, mean square error values obtained from regression analysis do not necessarily reflect the degree of agreement one might obtain from future applications of the equation. If the number of parameters in a model increases, the model can bend in more complicated ways.

None of the regression methods fits the observed data any better than ordinary least squares (OLS) regression; in fact, all of the methods approach OLS as more factors are extracted. However, when there are a large number of predictors, OLS can over-fit the observed data leading to large prediction errors. Regression methods with fewer extracted factors can provide better predictability of future observations. In order to simplify the final model and to make it more reliable and stable, non–significant predictors were eliminated using leave-one-out cross validation method. This method is generally preferred because it uses as much of the data as possible during model fitting and still allows for cross validation on the left-out points. In each cross validation iteration n=1,2,….14 a set of the structure models parameters estimates were obtained. Each set of parameter estimates was based on all the samples except one which was kept out for each cross validation iteration. Using leave-one-out validation method I choose the number of factors that minimizes the predicted residual sum of squares (PRESS) (Raymond 1986). However, often models with fewer factors have PRESS statistics that are only marginally larger than the absolute minimum can be chosen.

Table 3.5. Cross Validation for the number of extracted factors

Number of Extracted Factors	Root Mean PRESS	Prob > PRESS
0	1.076923	0.062
1	1.076793	0.024
2	0.866256	0.185
3	0.827262	0.217
4	0.758663	1
Minimum root mean PRESS		***0.7587***
Minimizing number of factors		4
Smallest number of factors with p > 0.1		2

Table 3.5 shows that minimizing number of factors to four, which also justifies the results found from the Table (3.4) that explains percent variation accounted for by principal components. Finally, model equation has been developed using four principal components. Table 3.6 shows the results of PCR using predictor variables TCI_{32} through TCI_{36} for first four principal components. The R^2 (0.86) is high and p values of the regression coefficients are low ($p<.05$) which are significant. The final set of coefficients for variables in model equation (3.4) are calculated and used to generate estimation model equation and (3.17).

The final malaria case is determined during activities weeks which occurs in week 32 to week 36.

$$DY= 86.48- 0.98\ TCI_{32}-0.36\ TCI_{33}+0.61\ TCI_{34}+ 2.20TCI_{35}-1.31TCI_{36} \quad (3.17)$$

Table 3.6. Results of principal component regression for equation (3.17) Bandarban district Bangladesh (R^2=0.86)

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	99.9894	1.23302	81.09	<.0001
Prin1	1.64897	0.57579	2.86	0.0187
Prin2	-28.26484	6.11437	-4.62	0.0012
Prin3	-28.75508	11.9657	-2.4	0.0397
Prin4	43.6613	18.32133	2.38	0.041

3.8 Simulation Vs Observed

Figure 3.3 displays observed versus simulated percent of malaria time series using predictor variables TCI_{32} through TCI_{36} (equation 3.17). Correlation analysis indicates malaria cases are more sensitive to thermal condition (TCI). Principal component regression could successfully eliminate the multicollinearity of the predictor variables in equation 3.17 for second peak at the end of September, the second cycle of malaria season, with high R^2 (0.86) value, and small error of estimates (1.78) of malaria cases. Since major malaria transmission season in Bandarban district, Bangladesh is September and October, VH indices from NOAA polar orbiting satellites can provide valuable information one month prior to malaria spreading largely in this area.

Table 3.7 Results of independently simulated malaria cases

year	observed	simulated
1992	42.5	44.4526
1993	50.25	51.1086
1994	47.94	46.7846
1995	45.24	45.1159
1996	43.91	41.3541
1997	45.17	45.8887
1998	54.49	52.0769
1999	49.68	50.5226
2000	45.13	45.8599
2001	45.13	43.8535
2002	46.71	48.3926
2003	44.42	41.8836
2004	38.31	38.8024
2005	36.39	39.0683

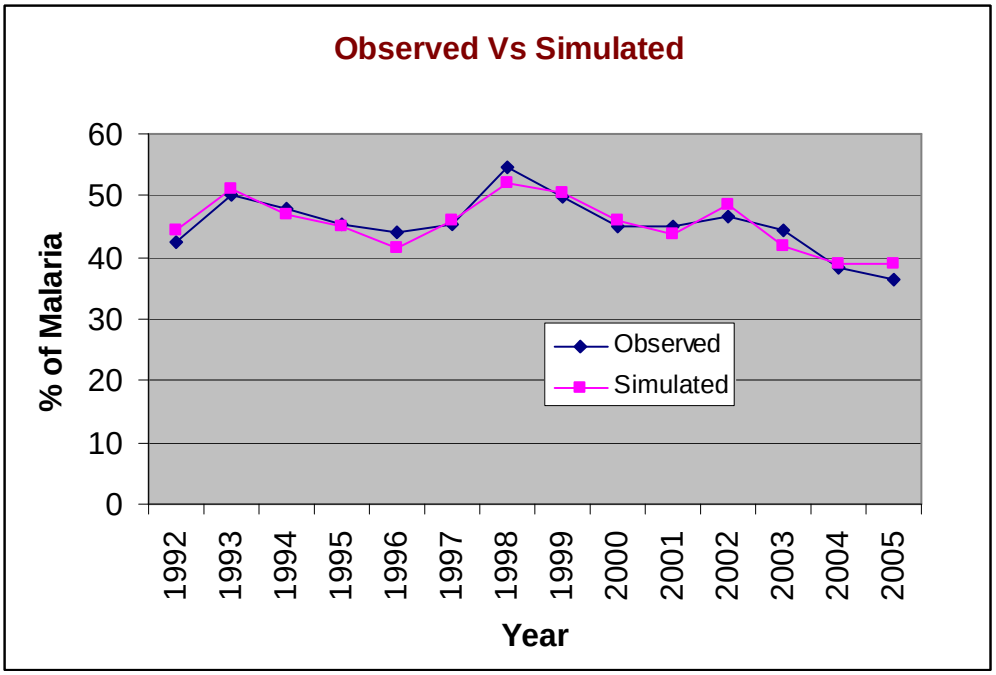


Figure 3.3: Independently simulated and observed percent of malaria Bandarban district, Bangladesh (R^2 =.86)

3.9 Model Test for 2005

In order to see if the correlation dynamics of TCI of week 32 through 36 is properly applicable for developing model for malaria forecasting I develop a model using 13 years (1992-2004) of malaria statistics instead of using 14 years (1992-2005) and use malaria data for 2005 for validation . The TCI of 2005 for week 32 through week 36 were used as a predictor in the model and has been found simulated % of malaria with error of less than 10% while comparing with collected malaria data for 2005. Figure 3.4 shows results found from the PCR model developed using 1992-2004 dataset.

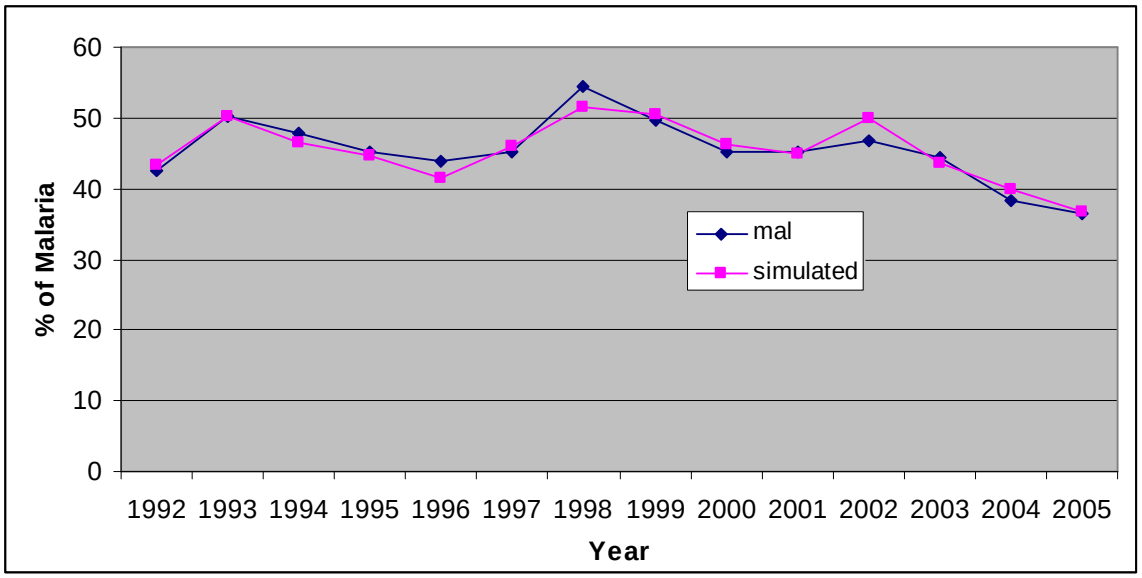


Figure 3.4: simulated and observed percent of malaria Bandarban district, Bangladesh (1992-2004)

Table 3.8 Coefficient of predictor variables for 13 years model to tests for year 2005

Intercept	88.198
TCI32	-0.788
TCI33	-0.414
TCI34	0.513
TCI35	2.31
TCI36	-1.466

$$DY = 88.19 - 0.79\,TCI_{32} - 0.41\,TCI_{33} + 0.51\,TCI_{34} + 2.31\,TCI_{35} - 1.46\,TCI_{36} \quad (3.18)$$

Table 3.9 Observed Numerical values TCI weeks 32-36, and observed % of malaria incidence for the year 2005

Year	% mal(Observed)	DY	TCI32	TCI33	TCI34	TCI35	TCI36
2005	36.39	86.485	25.6923	25.6296	25.3846	25.9643	24.44

Table 3.8 shows the coefficient for the model (equation 3.18) developed using dataset 1992-2004 and Table 3.9 shows observed values of TCI and malaria for 2005.The predictor value (TCI) of 2005 for equation 3.18 can be found from Table 3.9. Finally the simulated malaria was found using equation 3.3 is 36.76 where as expected malaria was 36.39.

3.9 Summary

For early warning of malaria using satellite derived vegetation health indices (thermal and moisture condition indices) we analyze malaria time series to extract weather related variation which has been used to compute correlation dynamics with TCI and VCI. For

all 52 weeks, correlation coefficients demonstrate a clear information about the critical period of major malaria transmission which could be explained through physical and environmental feature of the region (Bandarban). Following the correlation analysis multiple linear regressions has been performed for the model development. The diagnosis test for the reliability of the model was developed using p-value, variance inflation and tolerance statistics. Finally, correlation matrix was used to test for multicollianearity in predictor variables. From the results of all these tests it was clear that using OLS method for regression model would not be reliable and stable. Thus, predictor variable datasets should be transformed to achieve a new set of orthogonal variables for the elimination of collinearity. We use principal component regression which ultimately reduce collinearity and get a statistically reliable and stable model. For ensuring that our forecast model is robust and stable I use leave one out cross validation and calculate PRESS statistics for model development. The root mean square error for this model is 1.74. Furthermore, to test the stability and reliability of our model leaving 2005 dataset aside another model using the same predictor variables (TCI week 32 through 36) has been developed in which predictor variable (TCI week 32 through 36) of 2005 were used to find simulated percent of malaria and compared with observed malaria of 2005. The simulated malaria cases was 36.76 where as the expected malaria cases was 36.39. Finally, we can conclude that the results of this study, AVHRR-based vegetation health indices (VCI and TCI), can be used as a proxy for numerical estimation of number of malaria cases in coastal area of Bangladesh (Bandarban district) at least one month before the major transmission season (post monsoon October-November) starts.

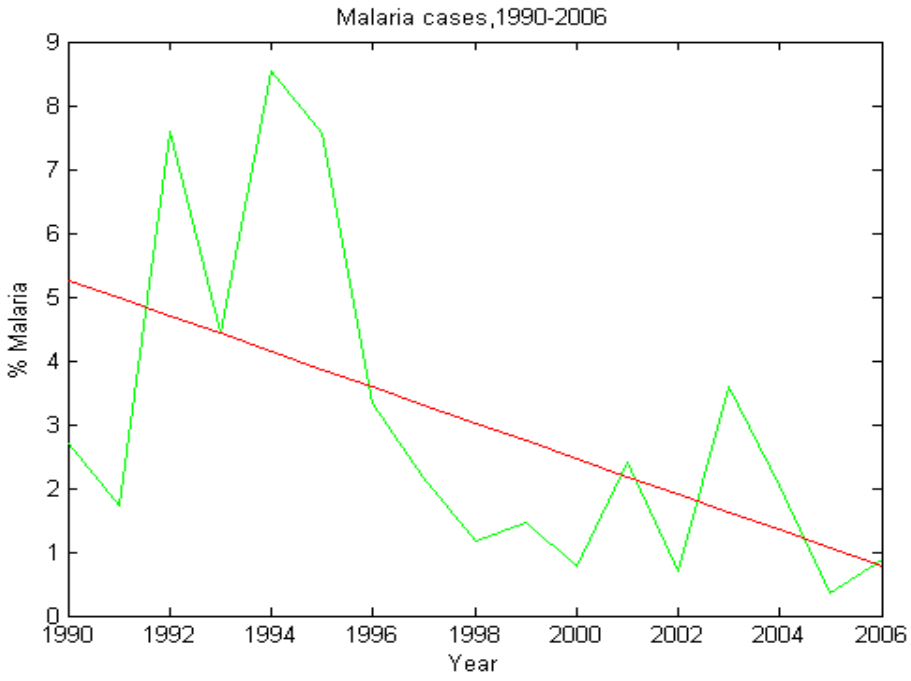
CHAPTER FOUR

Results and Discussion

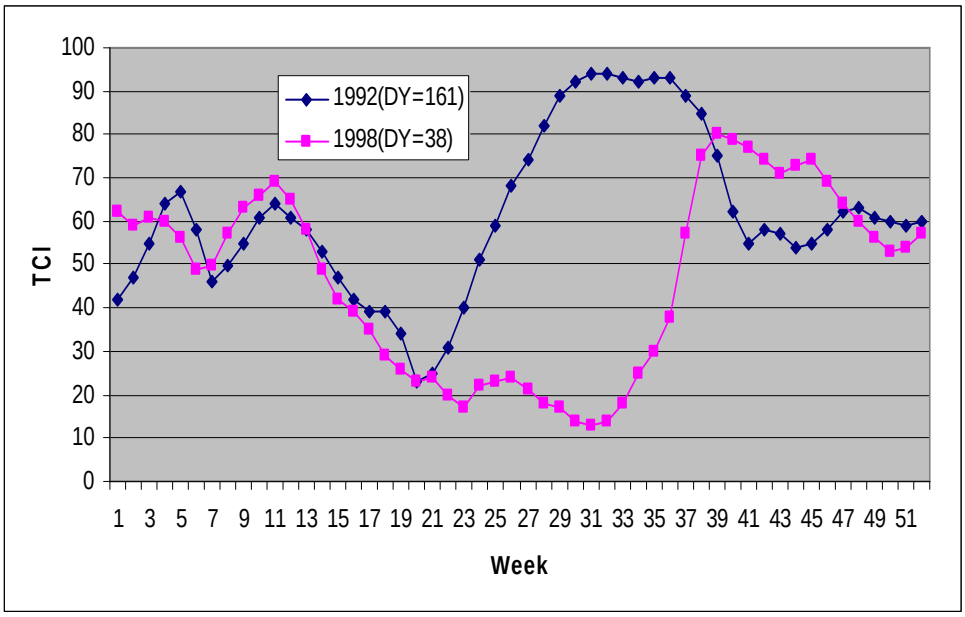
4.1 Modeling for malaria transmission in Bikkaner, Rajsthan, India

4.1.1 Trend analysis

The Malaria time series (Figure 4.1 A) for Bikkaner, Rajsthan was approximated by the linear equation (4.1). The weather-related variations around the trend were expressed as a ration of actual percent of Malaria instances to the estimated from the trend (Equation 4.2)



(A)



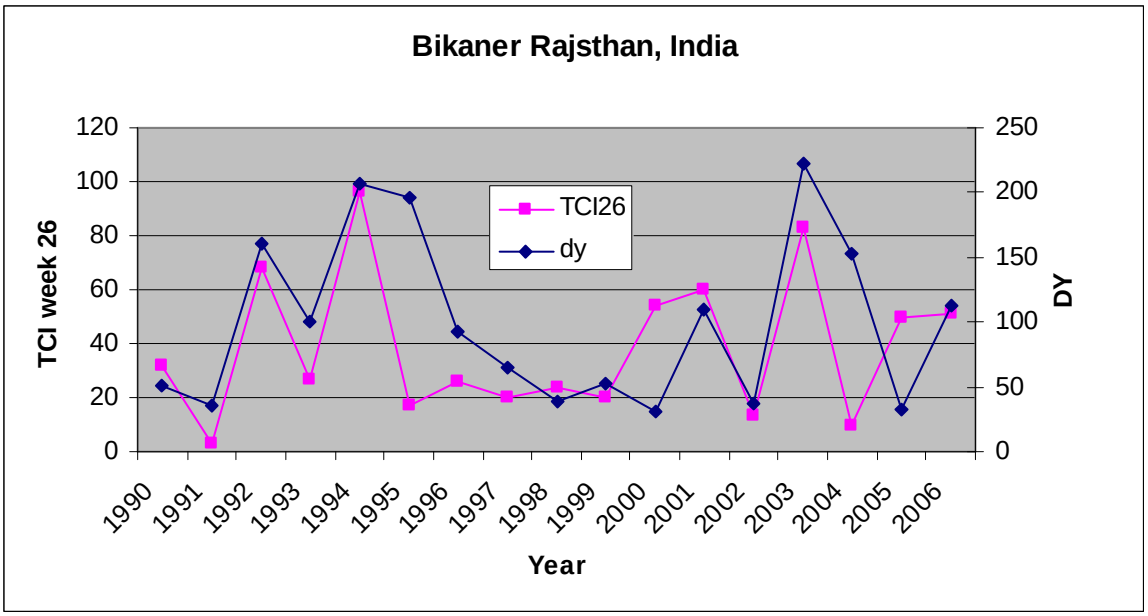
(B)

Figure 4.1. A. Percent of malaria in Bikaner district, Rajsthan, India and trend line, B. Temperature condition index (TCI) for 52 weeks during 1998(unfavorable year for malaria transmission) and 1992 (favorable)

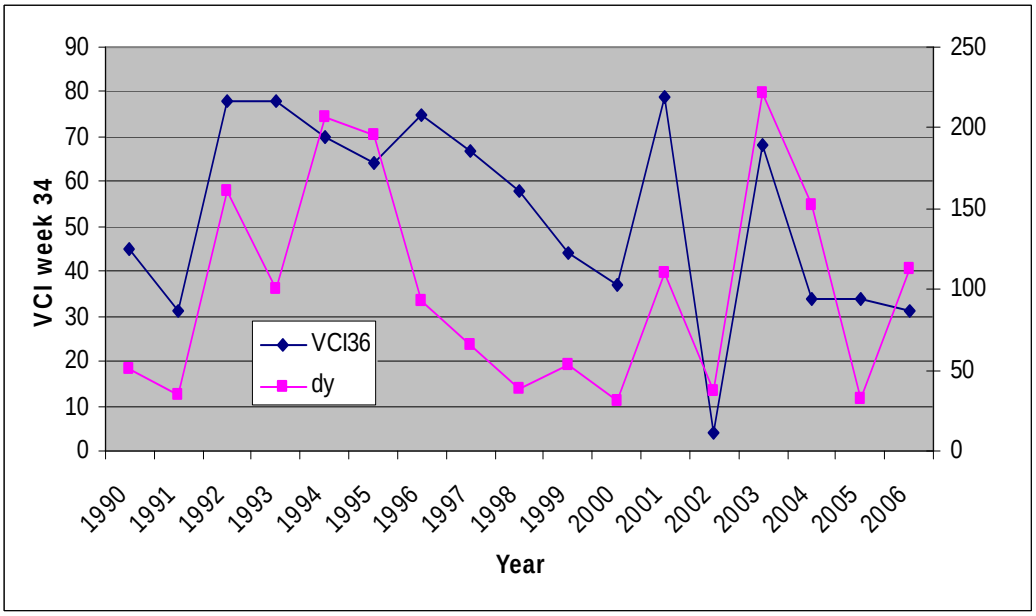
$$Y_{trend} = 562.26 - 0.279* \text{ Year} \quad (4.1)$$

$$DY = (Y / Y_{trend})*100 \quad (4.2)$$

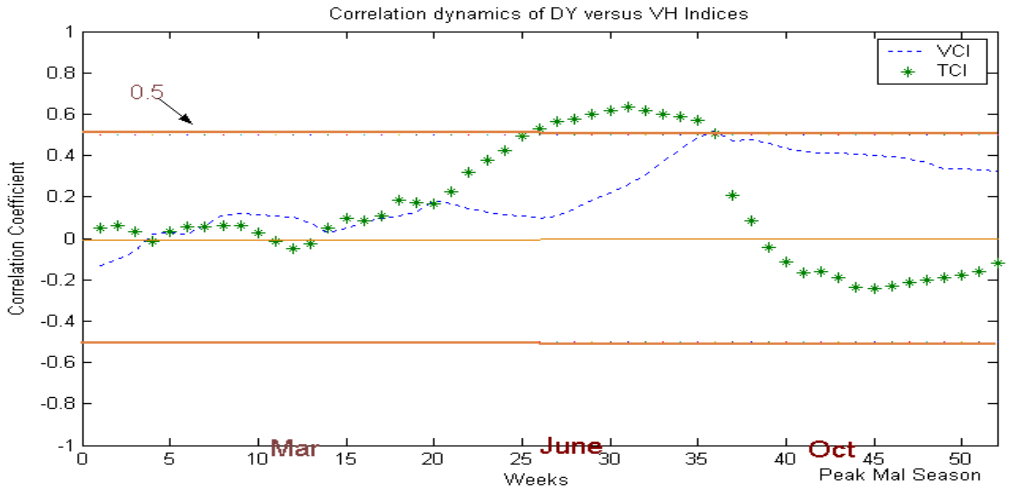
Figure 4.1 (B) shows temperature condition index (TCI) for 52 weeks during 1998 (unfavorable year for malaria transmission) and 1992 (favorable) which clearly indicates there are some positive correlation among TCI (week 26 through 32) and malaria detrended value (DY). While detrended value (DY) is 161%, TCI is high whereas detrended value (DY) is 38% TCI is very low. In 1994, DY was only 206 % or 106% above the trend, whereas in 1991, DY was 35% or 65% below the trend. These estimations indicate that the 1991(less % of malaria) was an unfavorable year for malaria whereas 1994 (higher % of malaria) was favorable.


Bikaner Rajsthan, India
TCI week 26
DY
Year
TCI26
dy
120
100
80
60
40
20
0
250
200
150
100
50
0
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

(A)


VCI week 34
Year
VCI36
dy
90
80
70
60
50
40
30
20
10
0
250
200
150
100
50
0
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

(B)



(C)

Figure 4.2: (A) Detrended value (DY) and Temperature condition Index (TCI) for weeks 26 from year 1990-2006, (B) Detrended value (DY) and Vegetation condition Index (VCI) for weeks 34-36 from year 1990-2006, (C) Correlation coefficient dynamics of the percent deviation of malaria from the trend versus TCI and VCI

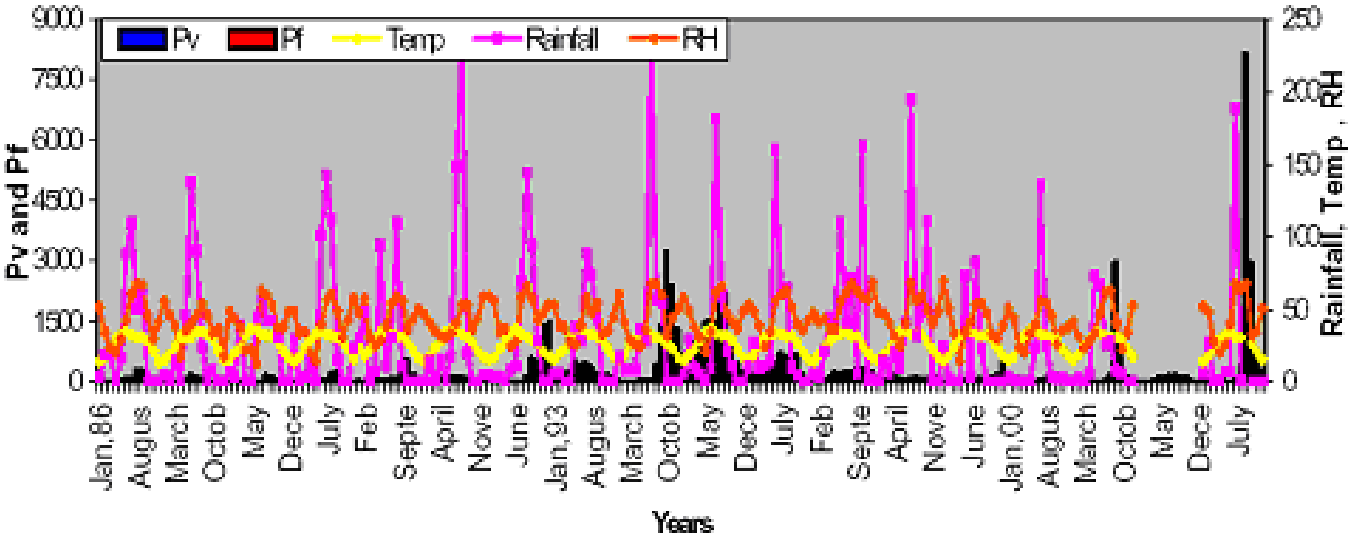


Figure 4.3: Relationship between Malaria and Met. Parameters: Bikaner district (Rajsthan) 1988-2002; Rainfall two months prior was found as an important indicator for early warning. Source: WHO-NEERI National Workshop on Climate change and its impacts on health. India, 2007

4.1.2 Correlation Dynamics

Figure 4.2 (A) shows detrended value (DY) and Temperature condition Index (TCI) for weeks 26 from year 1990-2006 which has clear indication that TCI(week 26) is positively correlated to variation in malaria(DY) involved for weather condition. On the other hand, similar phenomena are found for VCI (week 34) Figure 4.2 (B). Further investigation included correlation analysis of detrended malaria cases (DY) versus VCI and TCI, shown in Figure 4.2 (C). During cooler months (winter season December - February) when mosquitoes are less active; correlation is low for both indices. In spring (week 25)

when mosquito activity season starts, correlation for Bikaner district increases reaching maximum (0.6 for TCI and 0.3 for VCI) at the end of July (week 30-32). In week 31 correlation for Bekaner district starts decreasing for TCI whereas for VCI kept increasing reaching (0.5 for VCI) at the middle of September (week 36) TCI. These results are compatible with mosquitoes' activity in this region. This is also compatible with some of the results already found by other scientists as seen from Figure 4.3 which shows the relationship between malaria and meteorological parameters. Most cases of malaria occur in the post-rainy season and least number of cases occurs in January and February. Explanation of these relationships of VCI and TCI with malaria incidence is discussed in the summary and contribution section.

4.1.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 26 through 32 and to VCI for week 34 to 36. Therefore DY was regressed on the linear combination of TCI (Weeks 26 to 32), VCI (week 34 to 36) values. Table 4.1 presents the results of fitting the ordinary least squares (OLS) regressions model approximated by equations (4.3) for Bikaner district

$$DY=b_0 + b_1TCI_{26}+b_2TCI_{27}+b_3TCI_{28}+b_4TCI_{29}+ b_5TCI_{30} + b_6TCI_{31}+b_7TCI_{32}+b_8VCI_{34}+b_9VCI_{35}+ b_{10}VCI_{36} \quad (4.3)$$

Table 4.1. Results of multiple linear regressions (OLS) of DY on the equation (4.3), R^2=0.97

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	7.45761	25.9365	0.29	0.7834	.	0
TCI26	-14.71468	4.15969	-3.54	0.0123	0.00288	347.0034
TCI27	43.23332	12.6474	3.42	0.0142	0.000286	3500.0452
TCI28	-55.92391	16.9055	-3.31	0.0162	0.000155	6449.2997
TCI29	43.51318	8.99163	4.84	0.0029	0.000571	1750.2831
TCI30	-22.04618	16.3074	-1.35	0.2251	0.000196	5089.4082
TCI31	4.6315	15.847	0.29	0.7799	0.000235	4256.9735
TCI32	3.06023	6.70976	0.46	0.6644	0.00145	690.36098
VCI34	-27.44166	6.58006	-4.17	0.0059	0.00151	662.07195
VCI35	55.01534	12.4958	4.4	0.0046	0.000449	2229.0872
VCI36	-27.89067	6.35426	-4.39	0.0046	0.00184	543.43425

Table 4.1 shows coefficients of equation 4.3 where the value of R^2 is large (0.97) and p values for the regression coefficients are high, which are not significant at $p<0.05$ level with very small tolerance and high variance inflation. The high p-values (e.g. 0.78 and 0.77) are clear indication of collinearity among the predictor variables. In addition VH indices of neighboring weeks are highly correlated as seen in Table 4.2.

Table 4.2. Correlation Matrix among weekly TCI and VCI, Bikener district Rajsthan

	TCI26	TCI27	TCI28	TCI29	TCI30	TCI31	TCI32	VCI34	VCI35	VCI36
TCI26	1	0.9895	0.9578	0.9144	0.8637	0.8012	0.7411	0.3682	0.391	0.4076
TCI27	0.9895	1	0.9875	0.9576	0.9158	0.8581	0.8015	0.3459	0.3805	0.4102
TCI28	0.9578	0.9875	1	0.99	0.9631	0.9168	0.8698	0.2914	0.3367	0.3795
TCI29	0.9144	0.9576	0.99	1	0.9901	0.9595	0.9234	0.2704	0.3243	0.3767
TCI30	0.8637	0.9158	0.9631	0.9901	1	0.9886	0.965	0.2588	0.3196	0.3791
TCI31	0.8012	0.8581	0.9168	0.9595	0.9886	1	0.9925	0.2342	0.3006	0.3662
TCI32	0.7411	0.8015	0.8698	0.9234	0.965	0.9925	1	0.191	0.2594	0.3285
VCI34	0.3682	0.3459	0.2914	0.2704	0.2588	0.2342	0.191	1	0.9925	0.9686
VCI35	0.391	0.3805	0.3367	0.3243	0.3196	0.3006	0.2594	0.9925	1	0.9912
VCI36	0.4076	0.4102	0.3795	0.3767	0.3791	0.3662	0.3285	0.9686	0.9912	1

To avoid this multicollinearity problem, we used an alternative method of estimation, principal component regression (PCR), which results in better estimation and prediction

than OLS. The Eigenvectors of correlation matrix (Table 4.3) are calculated for recalculation of the coefficients of PCR model.

Table 4.3. Eigenvector of correlation matrix Bikener district Rajsthan

	Prin1	*Prin2*	*Prin3*	*Prin4*	*Prin5*
TCI26	0.3467	-0.06667	-0.54734	0.49787	-0.24373
TCI27	0.35802	-0.0889	-0.40794	-0.02181	-0.14242
TCI28	0.36315	-0.13042	-0.21311	-0.33841	0.10738
TCI29	0.3646	-0.14452	-0.00754	-0.40938	0.26206
TCI30	0.36169	-0.14817	0.19454	-0.22893	0.22504
TCI31	0.35151	-0.15464	0.38823	0.15801	-0.03576
TCI32	0.33682	-0.17203	0.52203	0.44948	-0.13971
VCI34	0.18107	0.55833	-0.04828	0.29016	0.61461
VCI35	0.19937	0.5428	0.04729	0.00414	0.01318
VCI36	0.21555	0.51762	0.15128	-0.32544	-0.6251

	Prin6	*Prin7*	*Prin8*	*Prin9*	*Prin10*
TCI26	-0.346276	0.355981	-0.065214	0.107841	0.07732
TCI27	0.311978	-0.553266	0.142198	-0.287558	-0.412803
TCI28	0.361603	-0.079189	-0.056813	0.443924	0.585082
TCI29	0.011739	0.535762	-0.093517	-0.556161	-0.067983
TCI30	-0.444151	-0.089414	-0.007325	0.512872	-0.487824
TCI31	-0.406999	-0.375188	0.219428	-0.330933	0.459872
TCI32	0.517796	0.199035	-0.136115	0.106715	-0.158913
VCI34	0.116087	0.048954	0.417698	0.029341	0.002689
VCI35	-0.085316	-0.219747	-0.772822	-0.095637	0.03496
VCI36	-0.009027	0.186552	0.353692	0.073129	-0.027121

4.1.4 Principal Component Regression

Using the PCR methodology, the coefficients of principal component estimators are shown in Table 4.4. We found simulated percent of malaria time series using (equation 4.4) predictor variables TCI_{26} through TCI_{32} and VCI_{34} through VCI_{34} for the three principal components (Prin1, Prin8, and prin9) only .

$$DY=18.15 -5.91\ TCI_{26} + 17.26\ TCI_{27}-32.88TCI_{28}+49.93\ TCI_{29}-44.07\ TCI_{30}+19.69TCI_{31} -2.87TCI_{32} - 25.07\ VCI_{34}+52.56VCI_{35}-27.59VCI_{36} \quad (4.4)$$

Table 4.4: Results of principal component regression for equation (4.4) Bikener district Rajsthan R^2=0.74, RMSE = 1.36

Variable	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*
Intercept	99.77645	7.73079	12.91	<.0001
Prin1	16.26859	2.9926	5.44	0.0001
Prin8	-1255.47081	474.89617	-2.64	0.0203
Prin9	-2285.4783	530.70424	-4.31	0.0009

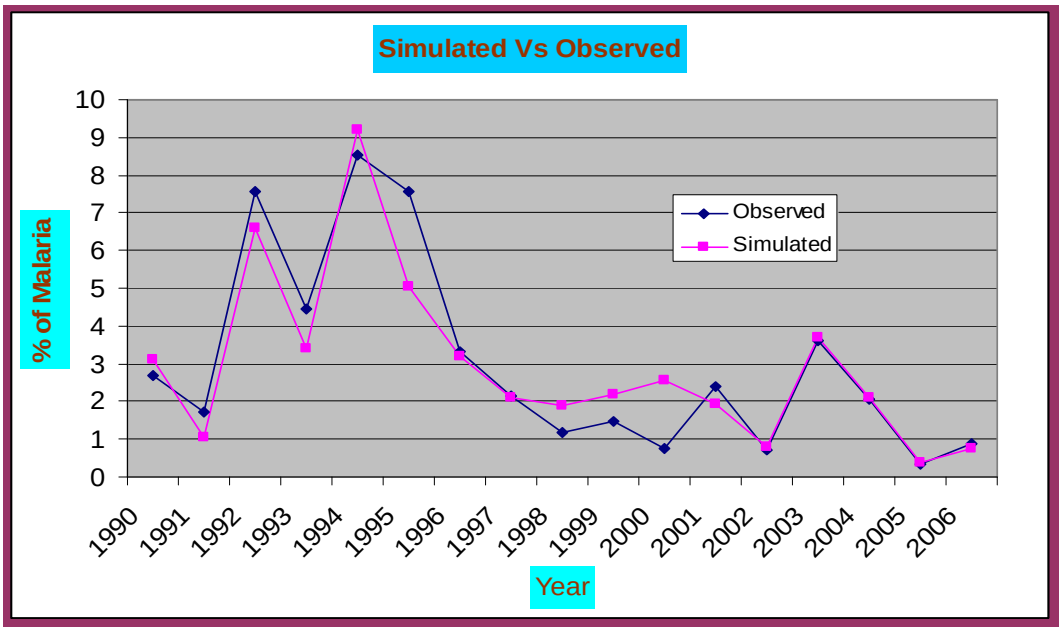


Figure 4.4: Independently simulated and observed percent of malaria Bikener Rajsthan (R^2 =0.74)

Figure 4.4 displays observed versus simulated percent of malaria time series using predictor variables TCI_{26} through TCI_{32} and VCI_{34} through VCI_{34} (equation 4.4). Correlation analysis indicates malaria cases are more sensitive to thermal condition (TCI). The earliest assessment can be done using model equation 4.4 at least two months prior to the malaria peak season. Principal component regression could successfully eliminate the multicollinearity of the predictor variables in equation 4.4 with high R^2 (0.86) value, and small error of estimates (.96) of malaria cases.

4.1.5 Contribution and Summary

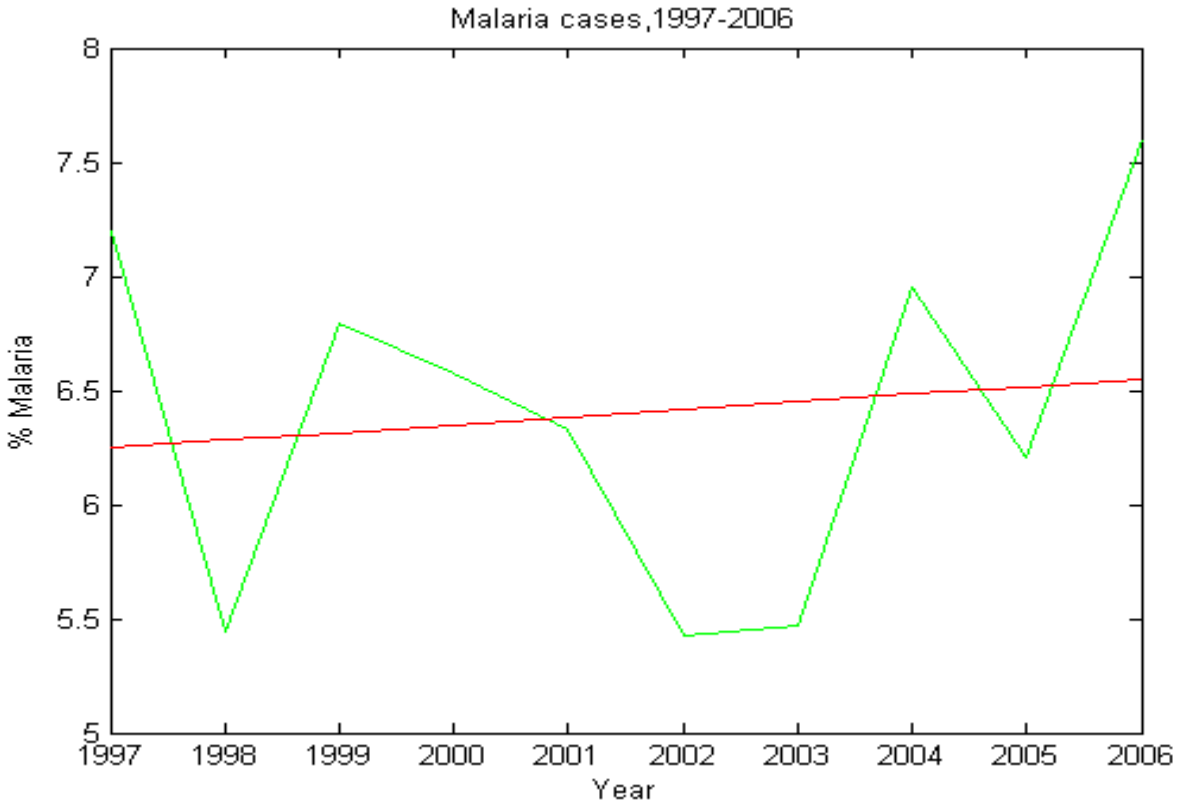
The correlation dynamics of malaria and TCI and VCI shows that malaria transmission in Bikaner, Rajsthan state is sensitive to thermal and moisture condition of that area which

can be measured using AVHRR based satellite data. The results of this study also indicate that AVHRR-based vegetation health indices (VCI and TCI) can be used as a proxy for numerical estimation of a number of malaria cases in Bikaner, Rajsthan. There are two species of *Anopheles* mosquito, namely, *Anopheles culicifacies* and *An. stephensi* in this region responsible for transmission of malaria and the main transmission season is post monsoon (October-November). The results can be explained using the climatic features of the area. The correlation between malaria and TCI is positive and high during week 30-36 (June-July) which signifies that lower temperature will increase malaria proliferation. The possible explanation is that the maximum temperature in this area is around 45 degree celsius (too high for parasite and mosquito survival) during summer which needs to be reduced for malaria transmission to occur. In other words, as the temperature decreases, the malaria transmission gets more favorable condition. Thus positive correlation for detrended value DY with TCI signifies the climatic features of this area. Since ecosystem in this area is dry, increasing moisture conditions obviously will provide better condition for malaria transmission which can be seen from positive correlation of VCI and malaria during malaria transmission season. In order to get a robust model, I have performed principal component regression which reduces multicolinearity in the model and have done cross validation using the leave one out method. Finally, my model can estimate malaria incidence at least two months earlier than the main transmission season October post monsoon.

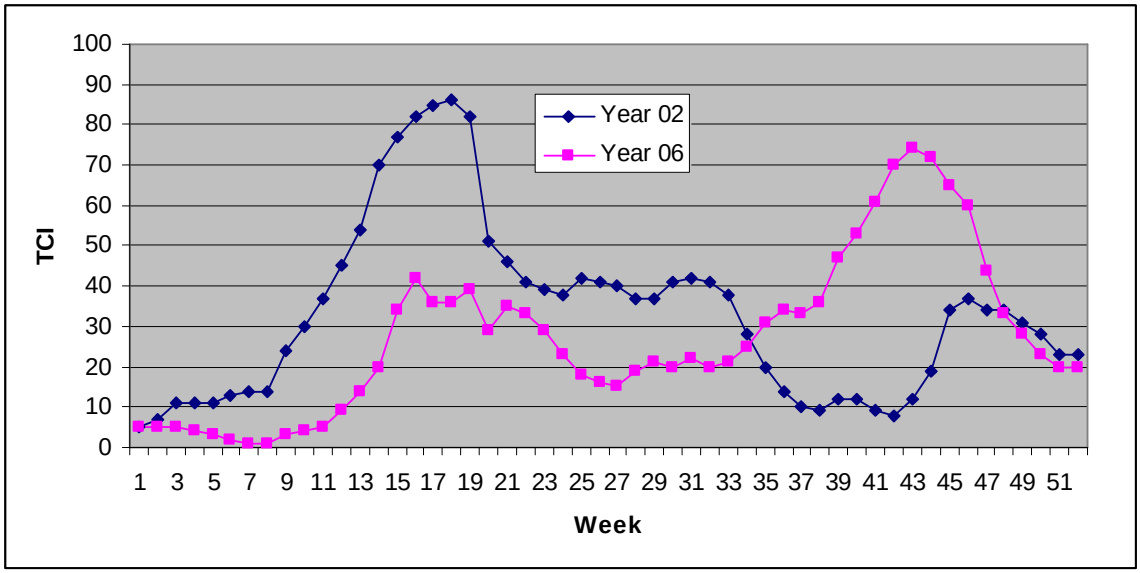
4.2 Modeling for Malaria Transmission in Tripura, India

4.2.1 Trend Analysis

The time series (Figure 4.5 A) for percent of malaria in Tripura state was approximated by the linear equation 4.5. The weather-related variations around the trend were expressed as a ration of actual percent of malaria cases to the estimated from the trend (Equation 4.6).



(A)



(B)

Figure 4.5 (A) Percent of malaria in Tripura state, India and the trend line, (B) Temperature condition index (TCI) for 52 weeks during 2002 (unfavorable year for malaria) and 2006 (favorable condition)

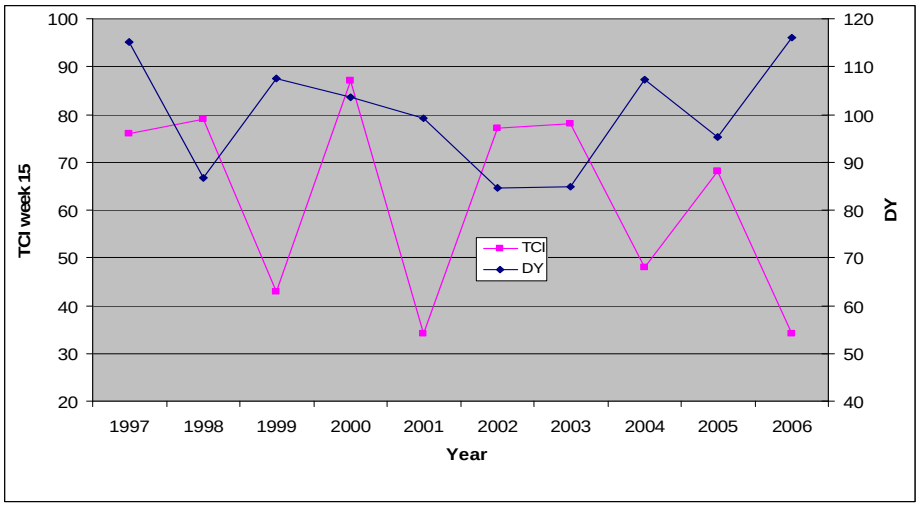
$$Y_{trend} = -61 + 0.034 * \text{Year} \quad 5.5$$

$$DY = (Y / Y_{trend}) * 100 \quad 5.6$$

Figure 4.5 (B) shows temperature condition index (TCI) for 52 weeks during 2002 (unfavorable year for malaria transmission) and 2006 (favorable) which clearly indicates that there is some negative correlation between TCI (week 15 through 20) and malaria detrended value (DY). While detrended value of 2002 is lower, TCI is higher whereas detrended value (DY) of 2006 is higher, TCI is very lower.

4.2.2 Correlation Dynamics

Figure 4.5 (B) shows the values of TCI of week 15- 20 during 2002 are larger (favorable condition and smaller DY) whereas during 2006 TCI for the same weeks are smaller (unfavorable condition and larger DY) which reflects a negative correlation whereas positive correlation exists during the end of the year.



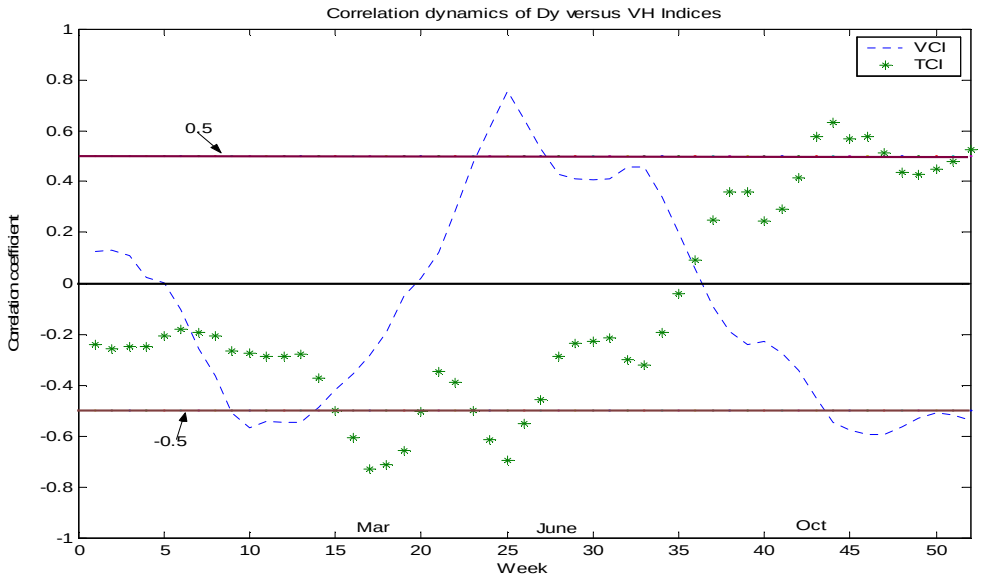


Figure 4.6 (A) Detrended value (DY) and Temperature condition Index (TCI) from year 1997-2006, (B) Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI

Figure 4.6 (A) shows detrended value (DY) and TCI for week 15 for year 1997-2006. There is a clear indication that TCI (week 15) is negatively related to DY. Investigation included correlation analysis of trend malaria cases (DY) versus VCI and TCI, shown in Figure 4.6 (B). Tripura is humid, subtropical, and has warm summer (chapter two). During cooler month (January – February) when mosquitoes are less active; correlation is low for both indices. In spring (pre monsoon) when mosquito activity season begins, correlation (negative) starts increasing to reach the maximum (-0.6 for VCI and -0.3 for TCI) at the beginning of March (week 10-15). In week 16 correlation for Tripura for TCI with DY is -0.6 whereas for VCI goes down reaching (-0.2 for VCI). Correlation coefficient for TCI gets lower and two weeks later at the beginning of June (week 25) and it again reaches maximum (-0.6). High positive Correlation (0.6) has seen with VCI at the same time week 25. During the rainy season (July-September) the correlation starts decreasing and during post rainy season (October) correlation increases (0.6) for TCI.

4.2.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 15 to 20 and to VCI for weeks 8 to 13. Therefore in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 15 to 20). I Initially attempted to

develop OLS model using VCI (week 8 to 15) but it didn't give a unique solution. Table 4.5 presents the results of fitting the ordinary least squares (OLS) regressions model approximated by equations (4.7) for Tripura

$$DY=b_0 + b_1TCI_{15}+b_2TCI_{16}+b_3TCI_{17}+b_4TCI_{18}+ b_5TCI_{19} + b_6TCI_{20} \quad (4.7)$$

Table 4.5: Results of multiple linear regressions (OLS) of DY on the equation (4.7) for Tripura, R^2=0.87

	Parameter				Variance
Variable	Estimate	t Value	Pr > |t|	Tolerance	Inflation
Intercept	119.1941	10.02	0.0021	.	0
TCI15	0.8867	1.46	0.2407	0.0398	25.153
TCI16	-0.52946	-0.55	0.62	0.0179	56.031
TCI17	-0.68417	-0.77	0.4962	0.0207	48.346
TCI18	1.01124	0.56	0.6167	0.006	167.28
TCI19	-3.32523	-1.92	0.1501	0.0072	138.32
TCI20	3.68654	2.44	0.0925	0.0222	45.099

Table 4.5 shows coefficients of equation (4.7) and that the value of R^2 is large 0.87. However p-values for the regression coefficients are high, which are not significant at $p<0.05$ level with very small tolerance and high variance inflation. R^2 for the equation (4.7) is 0.87 very high with p-values (e.g. 0.62 and 0.49) which are the clear indications of collinearity among the predictor variables. In addition, VH indices of neighboring weeks are highly correlated as seen in Table (4.6).

Table 4.6: Correlation Matrix among weekly TCIs for Tripura, India

	TCI15	TCI17	TCI18	TCI19	TCI20
TCI15	1	0.8487	0.7537	0.5674	0.3756
TCI16	0.9501	0.9488	0.8697	0.7025	0.5477
TCI17	0.8487	1	0.9501	0.8238	0.6921
TCI18	0.7537	0.9501	1	0.9525	0.8598
TCI19	0.5674	0.8238	0.9525	1	0.9627
TCI20	0.3756	0.6921	0.8598	0.9627	1

The Eigenvectors of correlation matrix (Table 4.7) are calculated for recalculation of the coefficients of PCR model.

Table 4.7: Eigenvector of correlation matrix Tripura, India

	Prin1	Prin3	Prin4	Prin5	Prin6
TCI15	0.370958	0.618843	-0.197321	0.3502	-0.091
TCI16	0.414474	-0.085445	0.636202	-0.4863	0.1857
TCI17	0.435032	-0.699175	-0.137711	0.4995	0.1734
TCI18	0.44446	-0.144185	-0.359668	-0.4162	-0.683
TCI19	0.412763	0.224511	-0.382064	-0.2524	0.6424
TCI20	0.365282	0.222918	0.511872	0.393	-0.22

4.2.5 Principal Component Regression

Using PCR methodology, the coefficients of principal component estimators are shown in Table 4.8. We found simulated percent of malaria time series using (equation 4.8) predictor variables TCI_{15} through TCI_{20} for the only one principal component (Prin1). Figure 4.7 shows simulated versus observed malaria time series of Tripura.

$$DY=129.7 - 0.07\ TCI_{15}-0.08\ TCI_{16}-0.08\ TCI_{17}-0.09\ TCI_{18}-0.09\ TCI_{19}-0.12TCI_{20} \quad (4.8)$$

Table 4.8: Results of principal component regression for equation (4.8) Tripura R^2=0.48, RMSE = 0.58

	Parameter	Standard		
Variable	Estimate	Error	t Value	Pr > |t|
Intercept	100.00131	2.91215	34.34	<.0001
Prin1	-3.67089	1.37816	-2.66	0.0286

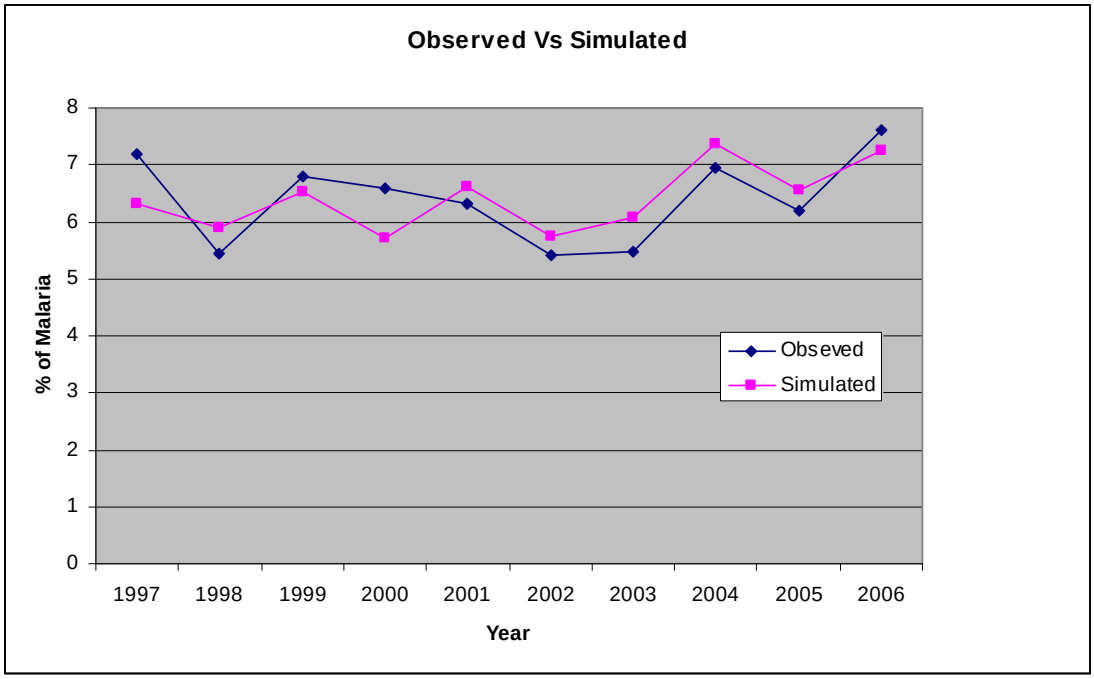


Figure 4.7: Independently simulated and observed percent of malaria Tripura (R^2 =.74)

4.2.6 Model Test for 2006

In order to see if the correlation dynamics of TCI of week 15 through 20 is properly applicable for developing a model, I use 9 years (1997-2005) of malaria data instead of using 10 years and 2006 malaria data has been left for validation (testing). Coefficients from Table 4.9 are used for this independent model test and it is found that

simulated malaria is 6.5 where as expected value was 7.6. Figure 4.8 (a) shows results found from the PCR model developed using 1997-2005 dataset.

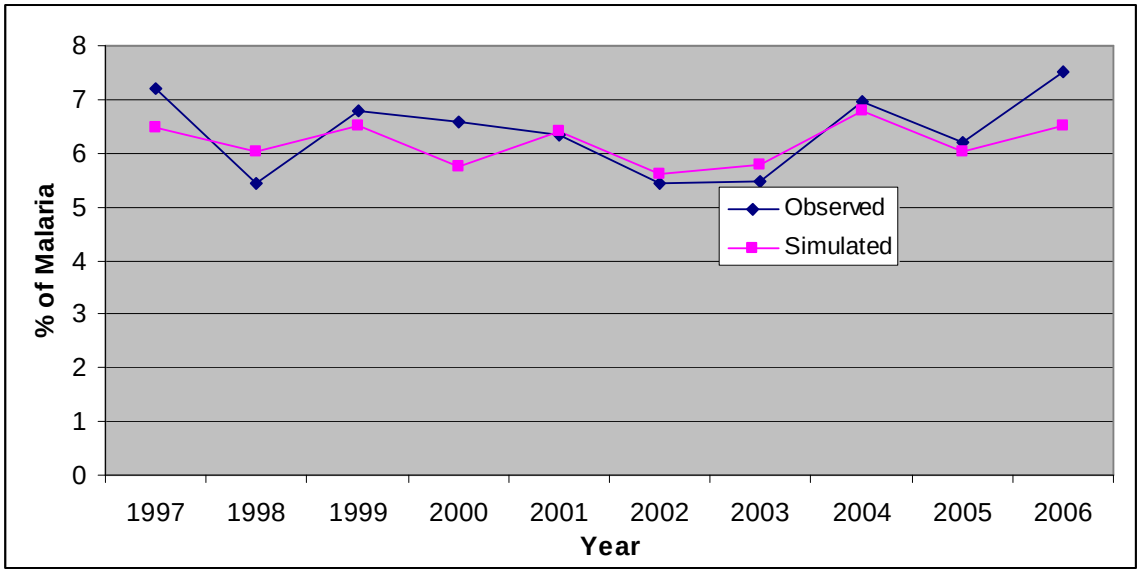


Figure 4.8(a): Simulated and observed percent of malaria using dataset from 1997-2006

Table 4.9. Coefficient of predictor variables from 1997-2005 model to tests for year 2006

Intercept	125.01
TCI15	0.057
TCI16	-0.066
TCI17	-0.072
TCI18	-0.084
TCI19	-0.078
TCI20	-0.099

Table 4.10: Numerical values TCI weeks 15-20, and % of malaria incidence for the year 2006

year	Mal(observed)	TCI15	TCI16	TCI17	TCI18	TCI19	TCI20
2006	7.6	4	42	36	36	39	29

4.2.7 Contribution and Summary

The correlation dynamics of malaria and TCI and VCI shows that malaria transmission in Tripura state is sensitive to thermal condition and also moisture condition of that area which can be measured using AVHRR based satellite data. *Anopheles Dirus* mosquito is responsible for malaria transmission in this area and the major transmission seasons for this vector are pre monsoon (April-May) period and post monsoon (October-November) period. The correlation between malaria and TCI is negative and high during week 15-20 which signifies temperature needs to increase for malaria proliferation. A possible explanation is that, maximum humidity during month of April becomes very high (near about 90%) which cause malaria transmission to terminate. So if the temperature increases humidity will decrease and will cause mosquito and parasite to get more active. Another possible explanation is that too much rainfall (3000 mm) in this area causes mosquito larva to get washed away quickly. So less rainfall in this area is necessary for larva to survive and can cause more malaria transmission. Thus, high temperature (less clouds and low rainfall) helps in this area to increase malaria. The correlation of DY with VCI during month of February also validates this explanation.

In order to get a robust model I have developed principal component regression (PCR) and have done cross validation using the leave one out method. I used the number of factors that has minimum PRESS (predicted square sum of residual). Later, I took 9 years (1997-2005) of data using the same variable to develop a PCR model using the same principal components in which I use TCI of 2005 for week 15 to week 20 to find percent of malaria. I found simulated malaria cases is 6.5 whereas expected malaria cases was

7.6. My model can estimate malaria incidence in this area than the main transmission season September and October (post monsoon). Finally, my model for malaria transmission can predict malaria incidence at least three months earlier than the main transmission season (post rainy season August and September).

4.3 Modeling for Malaria Transmission in Gujarat, India

4.3.1 Trend Analysis

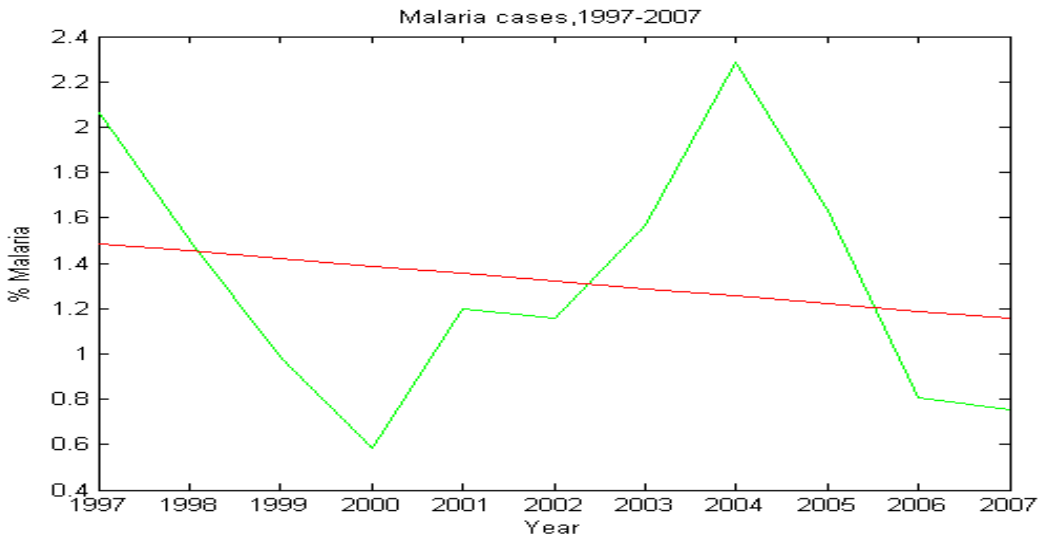
The malaria time series (Figure 4.9 A) of Gujarat was approximated by the linear equation 4.9. The weather-related variations around the trend were expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 4.10)

$$Y_{trend} = 67 - 0.033* \text{Year} \quad (4.9)$$

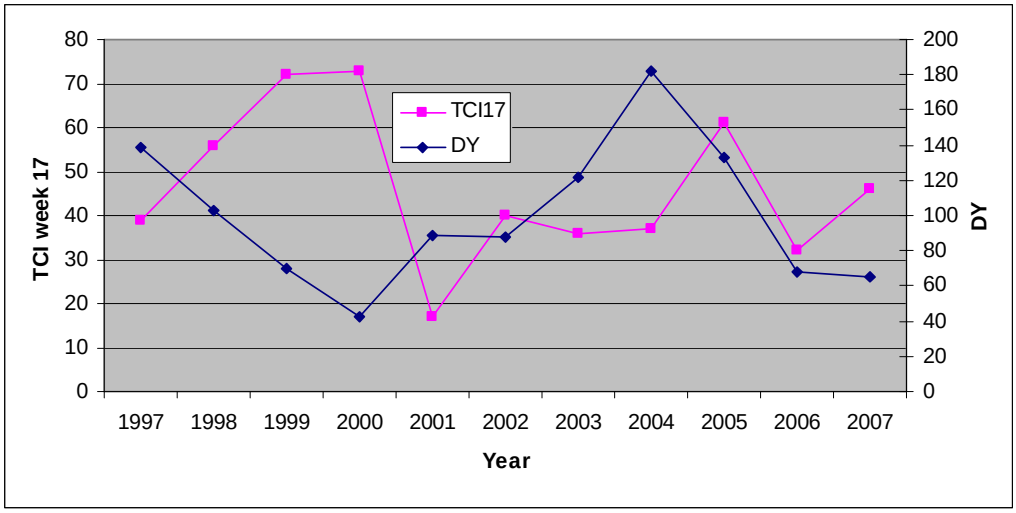
$$DY = (Y / Y_{trend})*100 \quad (4.10)$$

4.3.2 Correlation Dynamics

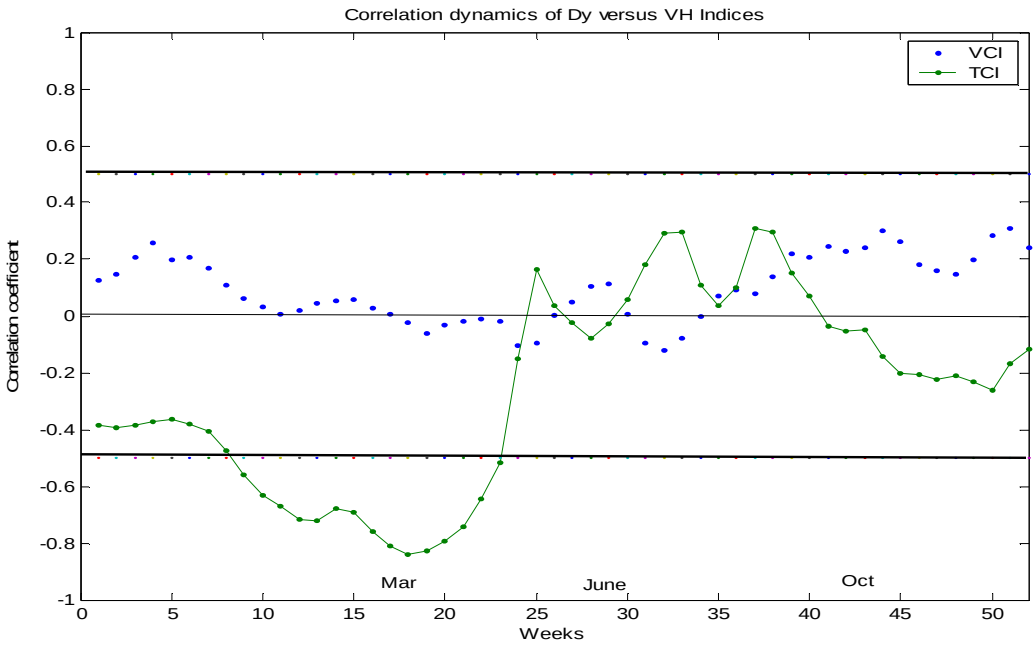
In 2004, DY was 181 % or 81% above the trend, whereas in 2000, DY was 42% or 58% below the trend. These estimations indicate that the 2000 (less % of malaria) was an unfavorable year for malaria whereas 2004 (higher % of malaria) was favorable.


Malaria cases,1997-2007
% Malaria
Year
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007

(A)


TCI week 17
DY
Year
TCI17
DY
80
70
60
50
40
30
20
10
0
200
180
160
140
120
100
80
60
40
20
0
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007

(B)



(B)

Figure 4.9: (A) Percent of malaria in Gujarat state, India and the trend line, (B) Detrended value (DY) and Temperature condition Index (TCI week 17) from year 1997-2006, (B) Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI for Gujarat state

Figure 4.9 (B) shows detrended value (DY) and Temperature condition Index (TCI) for weeks 17 for years 1997-2007 which clearly indicate that TCI (week 15) is negatively correlated to DY. Correlation analysis of detrended malaria cases (DY) versus VCI and TCI, shown in Figure 4.9(C). Since Gujarat is semi arid, subtropical, monsoon and has warm summer, during cooler month (November – January) when mosquitoes are less active; correlation is low for both indices. In spring (pre monsoon) when mosquito activity season starts, correlation (negative) starts increasing to reach maximum (-0.8 for

TCI and 0.1 for VCI) at the end of March (week 15-20). In week 16 correlation for Gujarat for TCI is -0.8. No significant relation has been found between moisture conditions (VCI) and malaria in this region.

4.3.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 10 to 17 therefore in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 10 to 18). Table 4.11 presents the results of fitting the ordinary least squares (OLS) regression model given by equations (4.11) for Gujarat

$$DY=b_0 + b_1TCI_{10}+b_2TCI_{11}+b_3TCI_{12}+b_4TCI_{13}+ b_5TCI_{14} + b_6TCI_{15}+b_7TCI_{16}+ b_8TCI_{17} + b_9TCI_{18} \quad (4.11)$$

Table 4.11. Results of multiple linear regressions (OLS) of DY on the equation (4.11) Gujarat state, R^2=0.95

	Parameter	Standard				Variance
Variable	Estimate	Error	t Value	Pr > |t|	Tolerance	Inflation
Intercept	117.16393	111.99943	1.05	0.4053	.	0
TCI10	-0.40561	10.23982	-0.04	0.972	0.0122	81.98942
TCI11	0.29234	22.87259	0.01	0.991	0.00251	398.3323
TCI12	2.49708	32.28462	0.08	0.9454	0.00138	726.00433
TCI13	0.08859	30.81281	0	0.998	0.0017	589.61722
TCI14	-15.16349	30.57239	-0.5	0.669	0.00189	528.9476
TCI15	21.72765	25.08345	0.87	0.4777	0.00274	365.12028
TCI16	-7.47178	11.21906	-0.67	0.574	0.01043	95.87372
TCI17	-2.30818	4.72368	-0.49	0.6734	0.05416	18.46345

Table 4.11 shows coefficients of equation (4.11) and that the value of R^2 is very large (0.95). However p-values for the regression coefficients is high, which are not significant

at $p<0.05$ level and very small tolerance and high variance inflation. R^2 for the equation (4.11) is 0.95 with largest p values of 0.99 which clearly indicates collinearity among the predictor variables. In addition VH indices of neighboring weeks are highly correlated as seen in Table 4.12

Table 4.12. Correlation Matrix among weekly TCI Gujarat, India

	TCI10	TCI11	TCI12	TCI13	TCI14	TCI15	TCI16	TCI17	TCI18
TCI10	1	0.9795	0.9185	0.8453	0.8117	0.7567	0.7365	0.6495	0.5377
TCI11	0.9795	1	0.9735	0.9178	0.8846	0.8366	0.8121	0.7192	0.6049
TCI12	0.9185	0.9735	1	0.9815	0.9556	0.9128	0.8791	0.7957	0.692
TCI13	0.8453	0.9178	0.9815	1	0.9863	0.9468	0.9079	0.8422	0.7451
TCI14	0.8117	0.8846	0.9556	0.9863	1	0.9814	0.9471	0.8902	0.7798
TCI15	0.7567	0.8366	0.9128	0.9468	0.9814	1	0.9849	0.9383	0.8147
TCI16	0.7365	0.8121	0.8791	0.9079	0.9471	0.9849	1	0.9619	0.8246
TCI17	0.6495	0.7192	0.7957	0.8422	0.8902	0.9383	0.9619	1	0.9383
TCI18	0.5377	0.6049	0.692	0.7451	0.7798	0.8147	0.8246	0.9383	1

The Eigenvectors of correlation matrix (Table 4.13) are calculated for recalculation of the coefficients of PCR model.

Table 4.13. Eigenvector of correlation matrix Gujarat, India

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
TCI10	0.307735	-0.49771	0.444218	0.310801	0.460446	-0.163972	0.35289	-0.010462	0.01094
TCI11	0.328979	-0.410049	0.215785	0.080861	-0.341954	0.269138	-0.646267	-0.063437	0.243881
TCI12	0.345423	-0.250438	-0.047756	-0.277356	-0.466391	0.128571	0.341504	0.153533	-0.603745
TCI13	0.348306	-0.103441	-0.239475	-0.492868	-0.056653	-0.51574	0.097127	-0.380785	0.379588
TCI14	0.351021	0.005827	-0.324303	-0.243029	0.533167	-0.0034	-0.352695	0.532917	-0.14086
TCI15	0.34818	0.140856	-0.355992	0.14563	0.235362	0.622501	0.181695	-0.483753	0.032539
TCI16	0.342961	0.199975	-0.267559	0.513597	-0.332277	-0.139381	0.237507	0.429595	0.369081
TCI17	0.328637	0.409126	0.125572	0.327926	-0.015203	-0.406527	-0.320147	-0.321069	-0.479828
TCI18	0.293883	0.532426	0.614182	-0.35416	0.005064	0.214358	0.124253	0.147111	0.210834

4.3.4 Principal Component Regression

Using PCR methodology, the coefficients of principal component estimators are shown in Table 4.14 . We found simulated percent of malaria time series using predictor variables TCI_{10} through TCI_{18} for the one of the principal components (Prin9, prin1 and Prin6).

$$DY=364.12 - 4.29\ TCI_{10}-43.77\ TCI_{11}+123.25\ TCI_{12}-88.52\ TCI_{13}+31.25\ TCI_{14}-1.68TCI_{15}-73.11\ TCI_{16}+83.63\ TCI_{17}-34.30TCI_{18} \quad (4.12)$$

Table 4.14. Results of principal component regression for equation (4.12) Gujarat R^2=0.82, RMSE = 0.82

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	99.98682	6.21275	16.09	<.0001
Prin9	-2923.77899	798.1203	-3.66	0.008
Prin1	-7.85186	2.32641	-3.38	0.0118
Prin6	192.96014	71.2186	2.71	0.0302

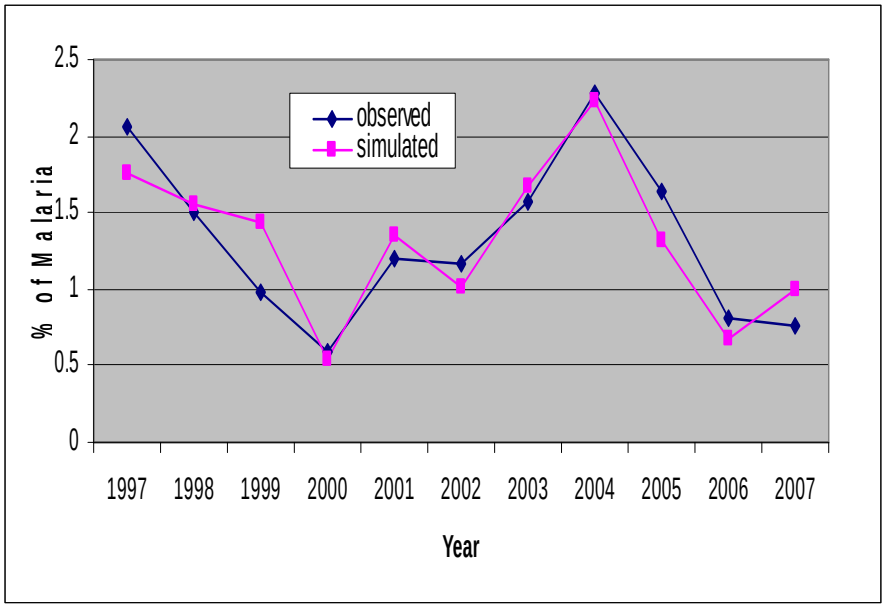


Figure 4.10: Independently simulated and observed percent of malaria Gujarat (R^2 =.82)

4.3.5 Contribution and Summary

The correlation dynamics of malaria and TCI and VCI demonstrate that malaria transmission in Gujarat is sensitive to thermal condition. However, since correlation of TCI and VCI in later months is poor and showing no continuities, moisture and thermal condition are a little bit unstable. Our model can predict malaria incidence much earlier (March- April) than the major transmission season August-September. The correlation between malaria and TCI is negative and high during week 10- 18 which signifies higher temperature will increase for malaria proliferation. A possible explanation is that most of the malaria affected villages were either on the bank of hill-streams or rivers, and the average temperature in these areas during week 10-20 is below 25 degree whereas the optimum temperature for malaria to transmit is 28 degrees. The PCR model I developed can estimate malaria incidence in this area much earlier than the main transmission season (post monsoon). Thus, AVHRR-based satellite data (TCI) can be used as a proxy for numerical estimation of number of malaria cases in Gujarat state.

4.4 Modeling for Malaria Transmission in Orissa, India

4.4.1 Trend Analysis

The malaria time series for Orissa (Figure 4.11 A) was approximated by the linear equation (4.13). The weather-related variations around the trend is expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 4.14)

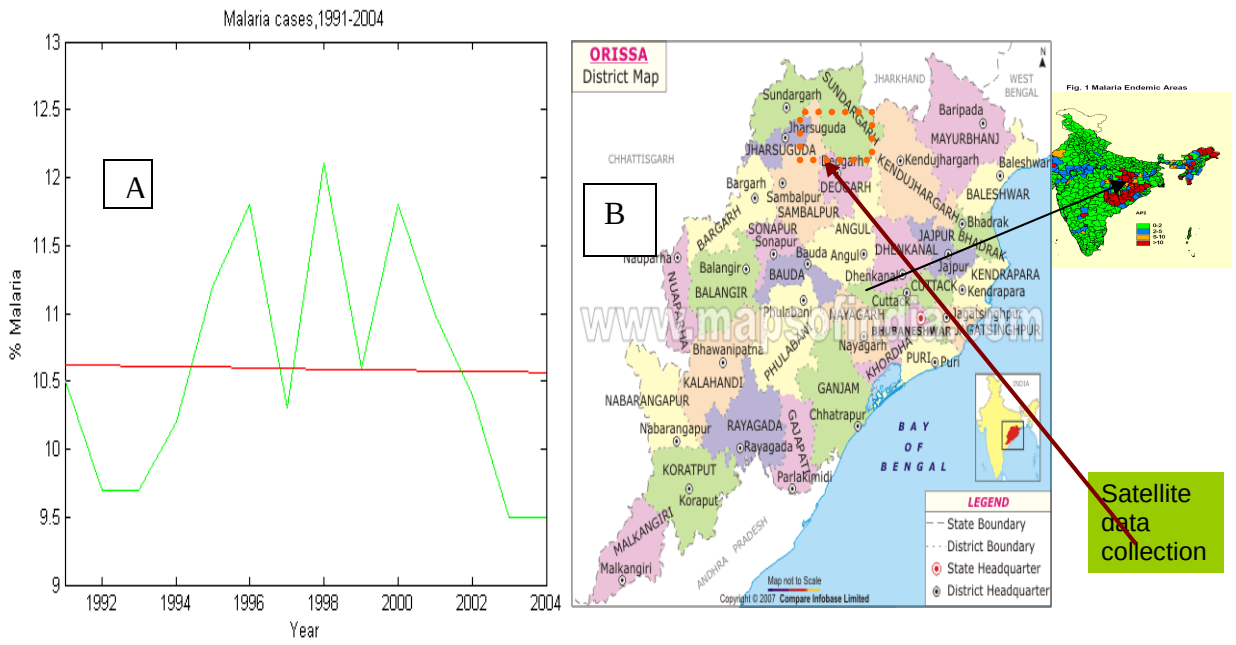


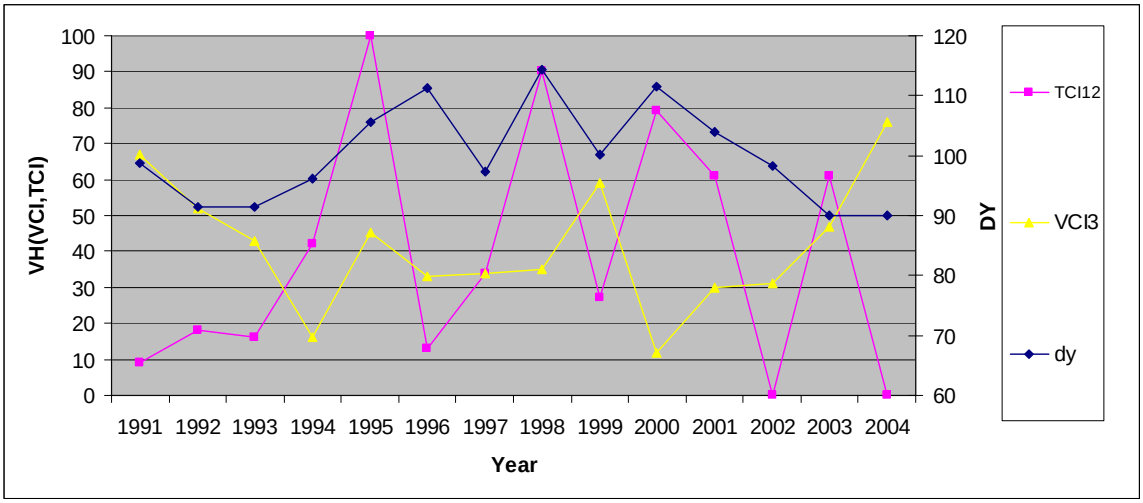
Figure 4.11. A. Percent of malaria in Orissa state, India and trend line (1991-2004) B. District Map of Orissa, the rectangular box where Satellite data collection has been performed.

$$Y_{trend} = 19 - 0.004 * Year \quad (4.13)$$

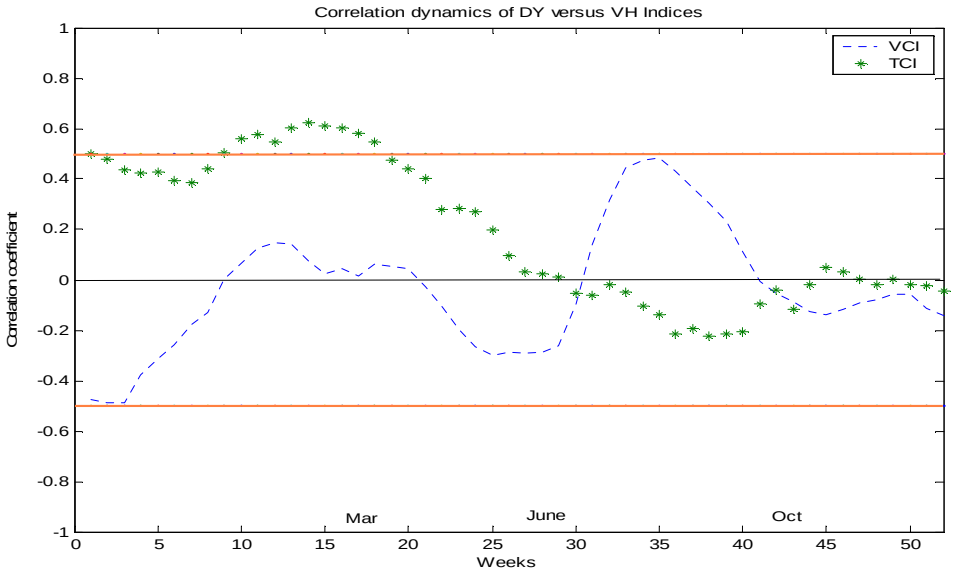
$$DY = (Y / Y_{trend}) * 100 \quad (4.14)$$

4.4.2 Correlation Dynamics

In 1998, DY was 114 % or 14% above the trend, whereas in 2004, DY was 90% or 10% below the trend. These estimations indicate that the 2004 (less % of malaria) was an unfavorable year for malaria whereas 1998(higher % of malaria) was favorable.



(A)



(B)

Figure 4.12: (A) Detrended value (DY) and Temperature condition Index (TCI week 13) , moisture condition index (VCI) from year 1991-2004, (B) Correlation coefficient

dynamics of the percent deviation of malaria from trend versus TCI and VCI for Orissa state

Figure 4.12 (A) shows detrended value (DY) and Temperature condition Index (TCI) for moisture condition index (VCI) from years 1991-2004, which has clear indication that TCI(week 12) is positively correlated to DY. Investigation included correlation analysis of trend malaria cases (DY) versus VCI and TCI, shown in Figure 4.12 (B). Since Orissa has monsoon, humid, subtropical warm summer, during cooler months (November – December) when mosquitoes are less active; correlation is low for both indices. In spring (pre monsoon, March and April) when mosquito activity season starts, correlation (positive) start increasing to reach maximum (0.6 for TCI and 0.2 for VCI) and high correlation persist until at the end of March (week 10-15). These phenomena can be explained as follows, if the temperature during this time reduces, malaria activities increase. In other words if more rainfall (more cloud cover and less temperature) occurs during this time malaria activities will increase. Higher amount of precipitation causes more breeding places for mosquito and malaria transmission increases.

4.4.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 15 to 20 and to VCI for week 3 to 6 therefore in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 15 to 20). Table 4.15 presents the results of fitting the ordinary least squares (OLS) regression model approximated by equation (4.15) for Orissa

$$DY=b_0 + b_1TCI_{15}+b_2TCI_{16}+b_3TCI_{17}+b_4TCI_{18}+ b_5TCI_{19} + b_6TCI_{20} \quad (4.15)$$

Table 4.15. Results of multiple linear regression (OLS) of DY on the equation (4.14) Orissa state, R^2=0.90

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	99.83444	16.83601	5.93	0.0096	.	0
VCI3	0.91571	0.53739	1.7	0.1869	0.02194	45.57015
VCI4	-3.16036	1.94513	-1.62	0.2027	0.00183	545.50029
VCI5	1.80515	3.36285	0.54	0.6287	0.00072873	1372.24929
VCI6	0.43761	1.74765	0.25	0.8184	0.00295	339.16084
TCI10	-0.09046	0.68777	-0.13	0.9037	0.00408	245.19655
TCI11	0.87113	1.25804	0.69	0.5384	0.00124	805.41076
TCI12	-1.91783	0.70888	-2.71	0.0734	0.00365	274.15286
TCI13	2.49771	1.10273	2.27	0.1084	0.0013	768.53066
TCI14	-2.13374	1.29062	-1.65	0.1968	0.0012	830.12221
TCI15	0.82447	0.55543	1.48	0.2344	0.00806	124.02533

Table 4.15 shows coefficients of equation (4.15) and that the value of R^2 is large 0.90. However p-values for the regression coefficients are high, which are not significant at $p<0.05$ level and very small tolerance and high variance inflation. R^2 for the equation (4.14) is 0.90 with largest p-value of 0.9 which clearly indicate collinearity among the predictor variables. In addition, VH indices of neighboring weeks are highly correlated as seen in Table 4.16 .

Table 4.16. Correlation Matrix among weekly TCI and VCI, Orissa, India

	TCI15	TCI17	TCI18	TCI19	TCI20
TCI15	1	0.8487	0.7537	0.5674	0.3756
TCI16	0.9501	0.9488	0.8697	0.7025	0.5477
TCI17	0.8487	1	0.9501	0.8238	0.6921
TCI18	0.7537	0.9501	1	0.9525	0.8598
TCI19	0.5674	0.8238	0.9525	1	0.9627
TCI20	0.3756	0.6921	0.8598	0.9627	1

In order to avoid this multicollinearity problem, we used an alternative method of estimation, principal component regression (PCR), which results in better estimation and prediction than OLS. The eigenvectors of correlation matrix (Table 4.13) are calculated for recalculation of the coefficients of PCR model.

Table 4.17. Eigenvector of correlation matrix Orissa, India

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
VCI3	-0.326349	0.293336	-0.100816	0.732765	0.113495	0.474784	0.114273	0.000133	-0.089708	-0.030569
VCI4	-0.277161	0.402217	0.130534	0.161867	0.163426	-0.712566	0.082799	0.169179	-0.087584	0.374909
VCI5	-0.265545	0.418343	0.21753	0.169746	0.032584	-0.083375	0.008537	-0.324736	0.503448	-0.55798
VCI6	-0.254499	0.425512	0.242732	-0.540976	-0.046823	0.444033	-0.116883	0.191767	-0.312316	0.235077
TCI10	-0.29165	0.335462	-0.579419	-0.212941	0.344727	-0.014875	0.380908	-0.338682	-0.018252	-0.205298
TCI11	-0.310711	0.334448	-0.305154	-0.015045	-0.09521	-0.030985	-0.281476	-0.687294	-0.017617	0.369216
TCI12	-0.324805	0.311793	-0.038905	-0.205864	-0.363462	0.02719	0.59628	-0.453726	-0.228719	-0.082587
TCI13	-0.345613	0.248904	-0.219434	-0.109677	-0.418716	0.140401	-0.296653	0.176063	0.573586	-0.339678
TCI14	-0.371585	0.111515	-0.376254	-0.064974	0.13628	-0.196202	-0.496576	-0.047878	-0.47993	-0.40917
TCI15	-0.369824	0.01053	-0.496469	-0.105756	-0.709405	-0.050837	0.227797	-0.053985	0.144885	-0.154324

4.4.4 Principal Component Regression

Using PCR methodology, principal component estimators are shown in Table (4.18). We found simulated percent of malaria time series using predictor variables TCI_{15} through TCI_{20} for four principal components (Prin1, Prin4, Prin6, Prin8).

$$DY=97.69 + 0.18\ VCI_3-1.56\ VCI_4+1.12\ VCI_5-0.005VCI_6-0.78\ TCI_{10} +1.85TCI_{11}+0.3TCI_{12}+0.03TCI_{13}+0.30TCI_{14}+0.03TCI_{15} \quad (4.16)$$

Table 4.18: Results of principal component regression for equation (4.16) Orissa R^2=0.70, RMSE = 0.49

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	99.99998	1.44207	69.34	<.0001
Prin1	1.98429	0.57972	3.42	0.0076
Prin4	-10.23794	6.1667	-1.66	0.1312
Prin6	24.96315	15.08825	1.65	0.1324
Prin8	-84.75777	40.53411	-2.09	0.0661

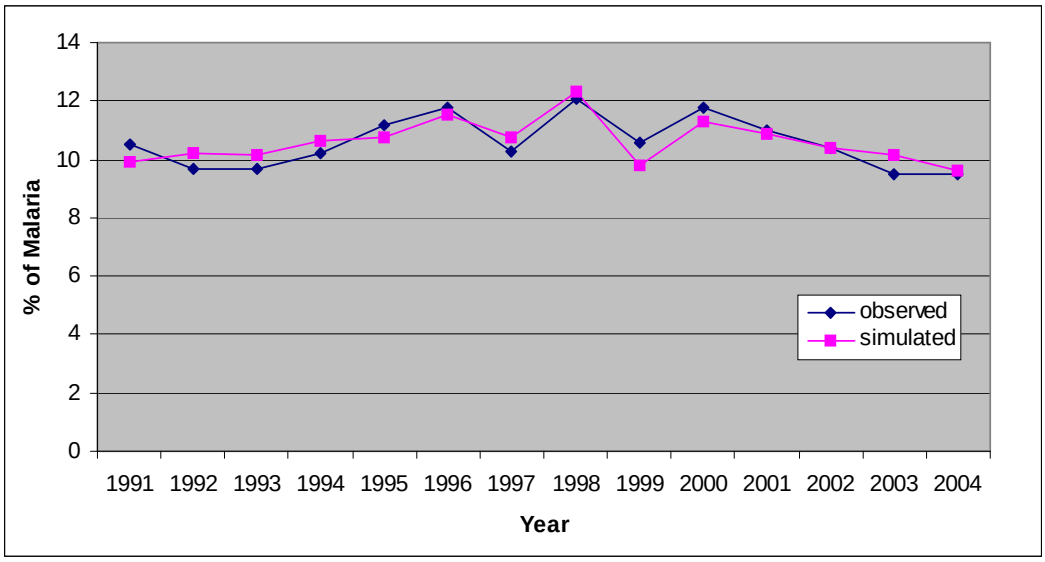


Figure 4.13: Independently simulated and observed percent of malaria Orissa (R^2 =.70)

4.4.5 Model Validation Test for Two Years

I have calculated the coefficients for the predictor variables for 12 years (1991-2002) and 13 years (1991-2003) of data sets to see if vegetation indices (VCI,TCI) could be used for future prediction. First I use 12 years (1991-2002) of data to develop a model comprising the same predictor variables as equation (4.5). TCI of 2003 dataset (Table 4.19) is used to find simulated malaria incidence for 2003 and then compared with observed cases. I have done the similar procedure using 13 years datasets (1991-2003) and find simulated malaria incidence for 2004 and compared with observed cases. In both cases I have found less than 10% of prediction error. Figure 4.14 (A) and 4.14 (B) below show both simulated and observed malaria cases for their respective years. Table 4.20 and 4.21

show the coefficients of the predictor variables for tests using coefficient of parameter estimation Tables 4.20 and 4.21.

Table 4.19. Numerical values of VCI from week 3 to 6 and TCI weeks 10-15, and % of malaria incidence for the year 2003 -2004

year	Mal(Obs)	VCI3	VCI4	VCI5	VCI6	TCI10	TCI11	TCI12	TCI13	TCI14	TCI15
2003	9.5	47	44	42	40	58	57	61	48	39	35
2004	9.5	76	75	72	69	3	0	0	0	0	0

Table 4.20. Coefficient of predictor variables for 13 years model to tests for year 2004

Intercept	94.48044895
VCI3	-0.0298383
VCI4	-0.00315351
VCI5	0.00513716
VCI6	0.00939507
TCI10	0.00794617
TCI11	0.01658807
TCI12	0.02259216
TCI13	0.02607084
TCI14	0.03342957
TCI15	0.04416186

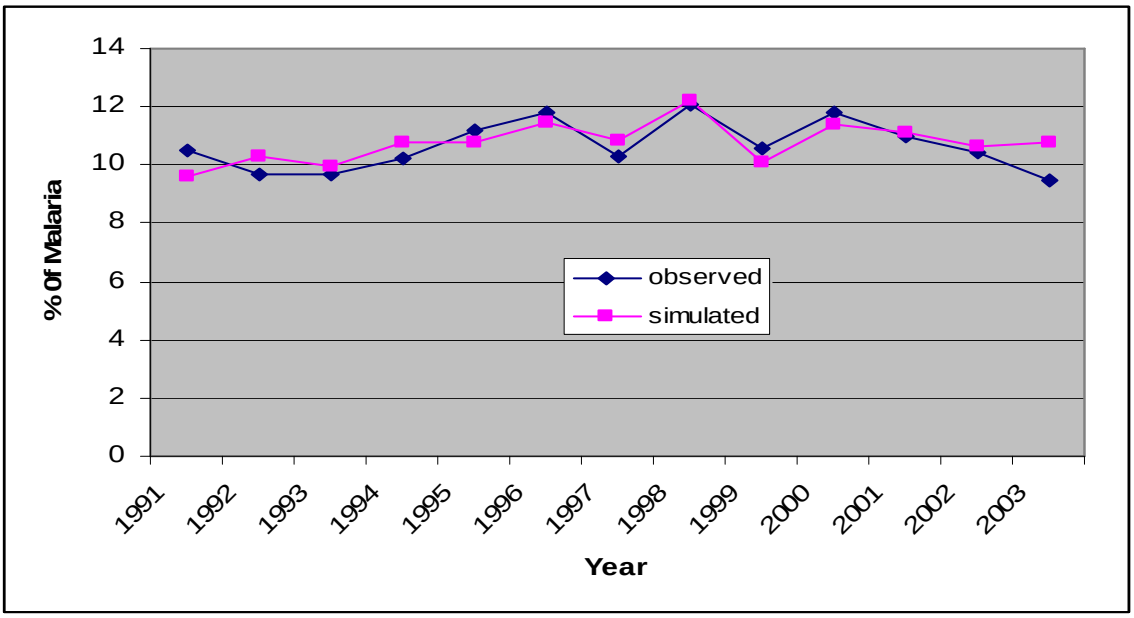


Figure 4.14 (A): Simulated and observed percent of malaria Orissa for 13 years model (1991-2003)

Table 4.21. Coefficient of predictor variables for 12years model to tests for year 2003

Intercept	99.6281103
VCI3	-0.00807946
VCI4	-0.02186629
VCI5	-0.03323182
VCI6	-0.03366654
TCI10	0.0683546
TCI11	0.04806437
TCI12	0.02462147
TCI13	0.01176858
TCI14	0.00301614
TCI15	-0.01352293

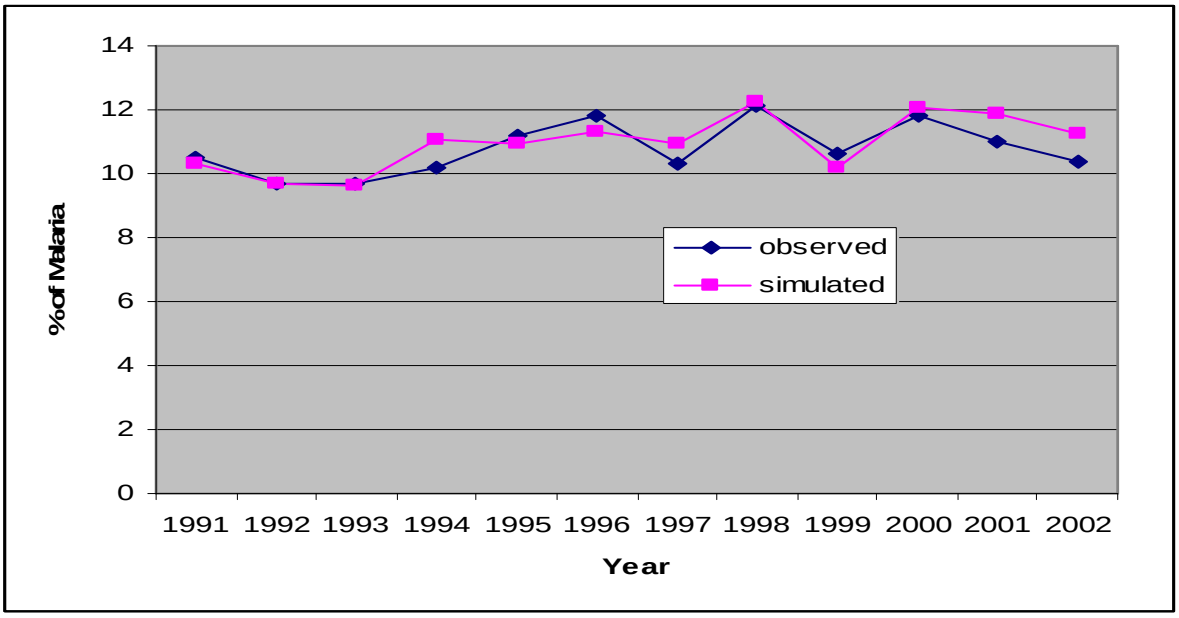


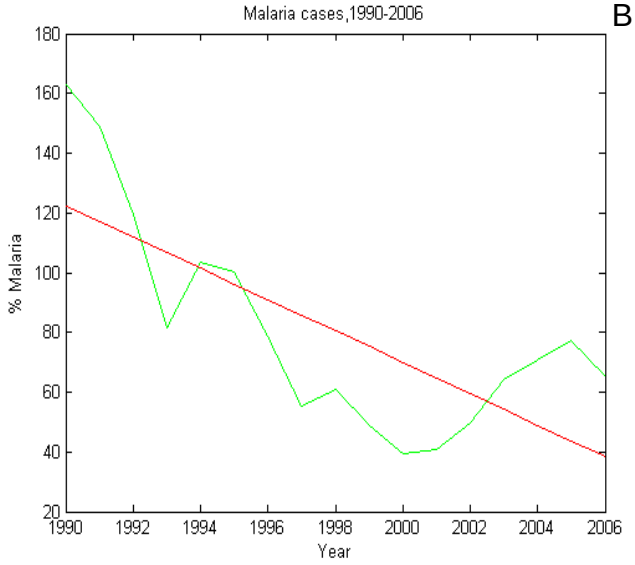
Figure 4.14(B): Simulated and observed percent of malaria Orissa for 12 years model (1991-2002)

4.4.6 Contribution and Summary

The correlation dynamics of malaria and TCI and VCI demonstrates seasonal malaria transmission in Orissa. Malaria is sensitive to thermal condition as well as moisture condition in this region which can be measured using AVHRR based satellite data. My model can also predict malaria much earlier than the major transmission season. Most importantly, using satellite data the peak transmission season for different vectors can also be detected. It is clear from correlation dynamics that two vectors of different seasonal behavior dominate malaria transmission in this area. The correlation between malaria and TCI is positive and high during week 10- 18 which signifies temperature needs to decrease for malaria proliferation. The possible explanation is that the average temperature is very high in Orissa during April and May which is not good for malaria transmission. So decrease of temperature in this area has a positive affect on malaria transmission. At the same time more rainfall during July, August shall prepare more breeding places which means that high VCI (better moisture condition, more rainfall) has positive relation with malaria transmission. This can be seen from correlation dynamics for VCI during month of July. To get a robust model I have developed PCR model and have done cross validation using the leave one out method. I used stepwise regression to eliminate unimportant principal components. My model can estimate malaria incidence in this area two months earlier than the main transmission season (post monsoon). Thus AVHRR-based vegetation health indices (VCI and TCI) can be used as a proxy for numerical estimation of malaria cases in Orissa state.

4.5 Modeling for Malaria Transmission in South America

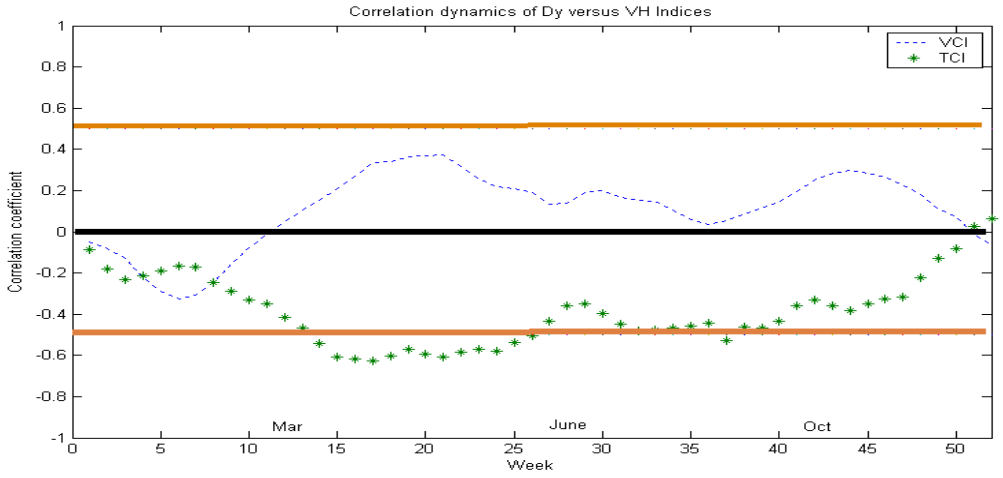
A.



B.

Satellite data collection

C.



(D)

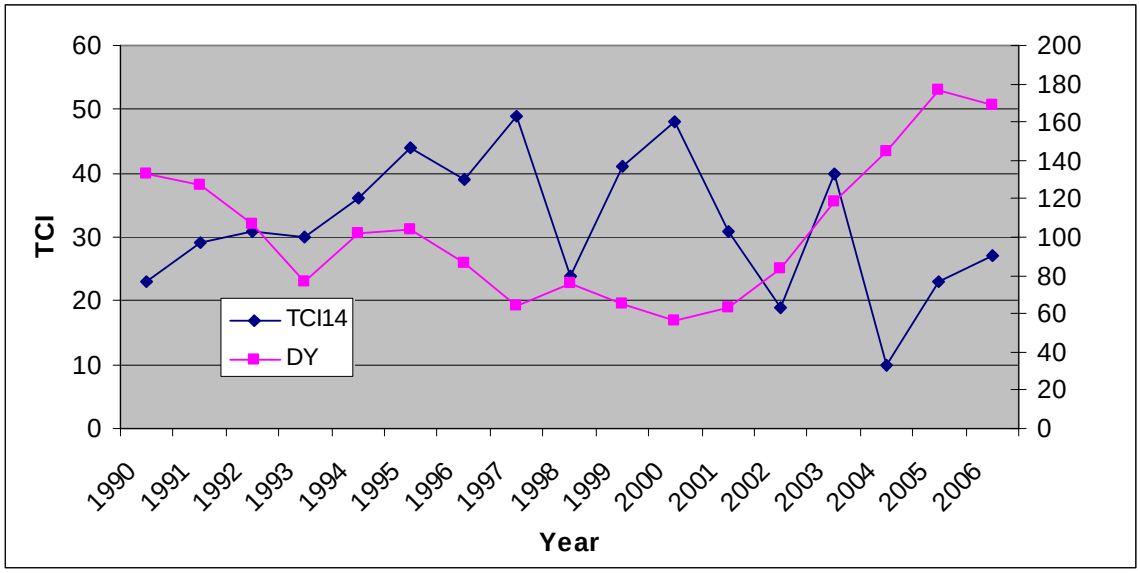


Figure 4.15: (A). Percent of malaria in Rondonia department, Brazil and trend line (B) District Map of Rondonia, rectangular box is pointed where Satellite data have been taken. (C) Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI, (**D**). Detrended value (DY) and Temperature condition Index (TCI week 14) and from year 1990-2006

4.5.1 Trend Analysis

Figure 4.15 (D) shows detrended value (DY) and Temperature condition Index (TCI) for for years 1990-2006, which suggests that TCI (week 14) has strong negative correlation with DY. The malaria time series (Figure 4.15 A.) for Rondonia, Brazil was approximated by linear equation (4.17). The weather-related variations around the trend

were expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 4.18)

$$Y_{trend} = 10557\text{-}5.24^{*} \text{ Year} \quad (4.17)$$

$$DY = (Y / Y_{trend})^{*}100 \quad (4.18)$$

Where Y is actual malaria cases per thousand of population, Y_{trend} is the malaria estimated from trend; Year is in year number; DY is deviation (%) from the trend.

4.5.2 Correlation Dynamics

Investigation included correlation analysis of trend malaria cases (DY) versus VCI and TCI, shown in Figure 4.15(C). When mosquito activity season starts (April-May), correlation for Rondonia increases to maximum (-0.7 for TCI and 0.3 for VCI) at the end of April (week 20-25). There is no significant relation between VCI and malaria incidence.

4.5.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 14 to 20, therefore in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 14 to 20) values. Table 4.22 presents the results of fitting the ordinary least squares (OLS) regression model given by equations (4.19) for Rondonia

$$DY = b_0 + b_1TCI_{14} + b_2TCI_{15} + b_3TCI_{16} + b_4TCI_{17} + b_5TCI_{18} + b_6TCI_{18} + b_7TCI_{19} + b_8TCI_{20} \quad (4.19)$$

Table 4.22. Results of multiple linear regressions (OLS) of DY on the equation (4.19) for Rondonia, R^2=0.75

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	256.07604	58.75465	4.36	0.0018	.	0
TCI14	5.65025	3.92212	1.44	0.1836	0.0344	29.07148
TCI15	-17.94924	11.23983	-1.6	0.1447	0.00505	198.05486
TCI16	21.64502	14.11157	1.53	0.1594	0.00223	447.44866
TCI17	-21.38452	12.32622	-1.73	0.1168	0.00215	464.80709
TCI18	9.6336	7.88253	1.22	0.2527	0.00369	270.82763
TCI19	3.35864	4.10134	0.82	0.434	0.01043	95.89343
TCI20	-5.18695	3.49724	-1.48	0.1722	0.01778	56.24375

Table 4.22 shows regression coefficients of equation (4.19) where one can see that the value of R^2 is large (0.75) and p-values for the coefficients are high, which are not significant at 10% level with very small tolerance and high variance inflation. The R^2 for the equation (4.18) is 0.75 with the largest p value 0.43 which is an indication of collinearity among the predictor variables. In addition, VH indices of neighboring weeks are highly correlated as seen in Table 4.23.

Table 4.23. Correlation Matrix among weekly TCIs Rondonia, Brazil

	TCI14	TCI15	TCI16	TCI17	TCI18	TCI19	TCI20
TCI14	1	0.9554	0.9025	0.8439	0.753	0.6914	0.6555
TCI15	0.9554	1	0.9797	0.9363	0.869	0.8134	0.7766
TCI16	0.9025	0.9797	1	0.9812	0.924	0.8664	0.8287
TCI17	0.8439	0.9363	0.9812	1	0.9741	0.9286	0.8983
TCI18	0.753	0.869	0.924	0.9741	1	0.9829	0.9662
TCI19	0.6914	0.8134	0.8664	0.9286	0.9829	1	0.9885
TCI20	0.6555	0.7766	0.8287	0.8983	0.9662	0.9885	1

To avoid this multicolinearity problem, we used an alternative method of estimation, principal component regression (PCR).

Table 4.24: Eigenvector of correlation matrix Rodonia,Brazil

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
TCI14	0.347361	0.598194	0.575274	0.415168	-0.004472	0.074313	0.112505
TCI15	0.37967	0.373916	-0.021215	-0.666928	-0.109174	-0.366213	-0.35322
TCI16	0.389196	0.210486	-0.471934	-0.207213	0.142152	0.409735	0.592007
TCI17	0.394235	-0.005351	-0.451626	0.442784	0.162841	0.174859	-0.622441
TCI18	0.388827	-0.265032	-0.157958	0.308917	-0.117038	-0.722867	0.349223
TCI19	0.376932	-0.405565	0.207033	-0.078093	-0.713253	0.365365	-0.047505
TCI20	0.367467	-0.472539	0.417263	-0.205718	0.647231	0.076676	-0.021423

4.5.4 Principal Component Regression

Table (4.25) shows the coefficients of principal component regression methodology using predictor variable TCI of week 15 through 20. Stepwise regression suggests that includes only first principal component gives the most stable result for PCR.

$$DY=18.15 -5.91\ TCI_{14} + 17.26\ TCI_{15}-32.88TCI_{16}+49.93\ TCI_{17}-44.07\ TCI_{18} +19.69TCI_{19}-2.87TCI_{20} \quad (4.20)$$

Table 4.25: Results of principal component regression for equation (4.20) Rondonia R^2=0.68, RMSE = 21

	Parameter	Standard		
Variable	Estimate	Error	t Value	Pr > |t|
Intercept	103.12689	7.24462	14.23	<.0001
Prin1	-9.27971	2.97521	-3.12	0.007

Figure 4.16 shows the observed and simulated value form 1990 to 2006 using equation 4.19 with R square value .68

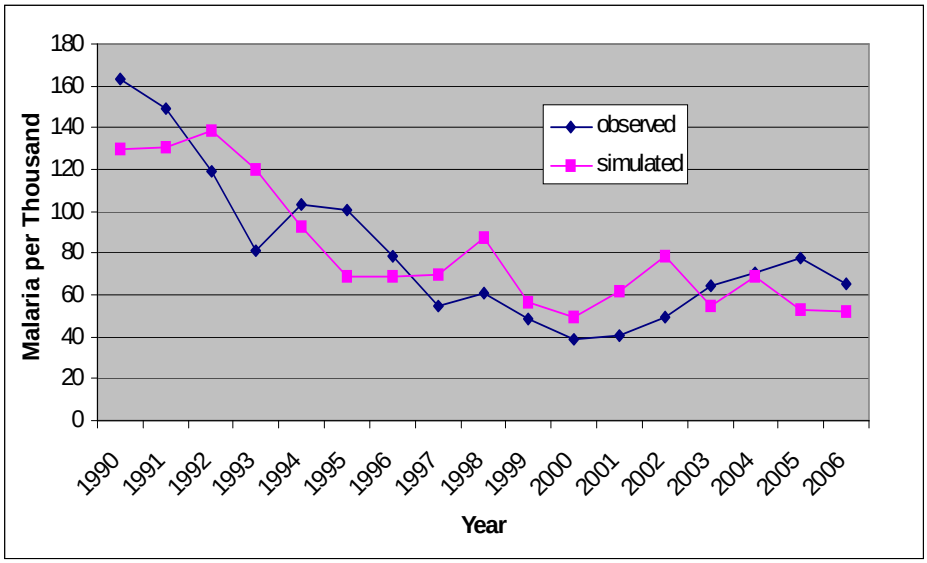
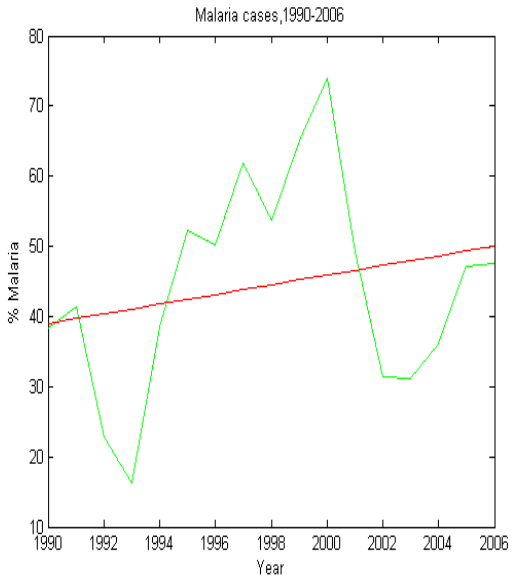


Figure 4.16 .Independently simulated and observed malaria per thousand of population Rondonia (R2 =.68).

4.6 Modeling for Malaria Transmission in Ampa Brazil

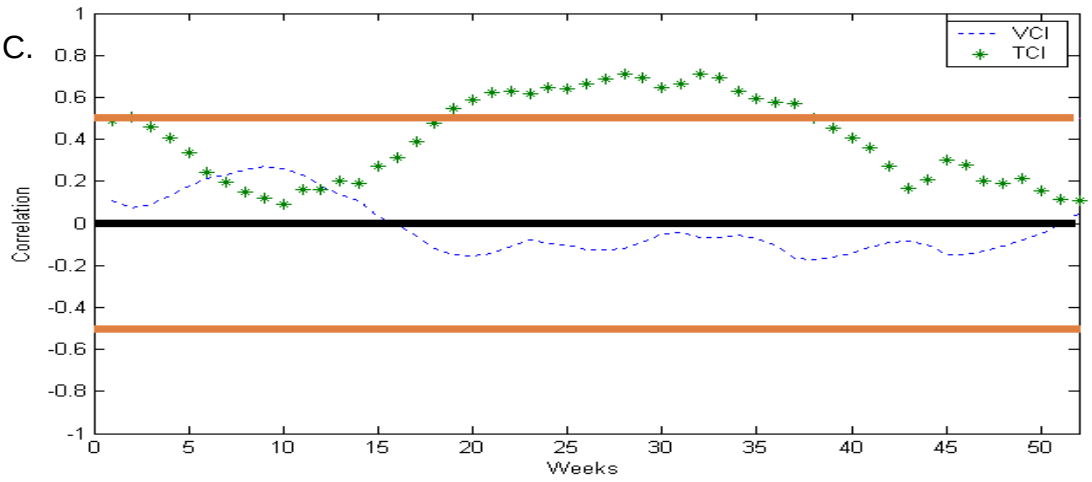
A.



B.

Satellite data Collection

C.



D.

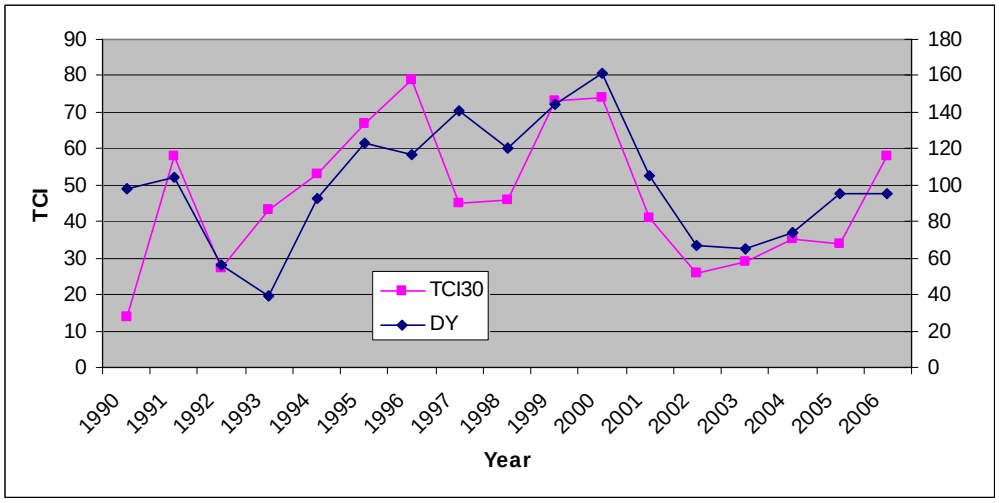


Figure 4.17: A. Percent of malaria in Ampa department, Brazil and trend line B. District Map of Ampa, rectangular box is pointed where Satellite data have been taken. C. Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI, D. Detrended value (DY), Temperature condition Index (TCI week 30) and from year 1990-2006

4.6.1 Trend Analysis

The malaria time series (Figure 4.17A.) for Ampa, Brazil was approximated by the linear equation (4.21). The weather-related variations around the trend were expressed as a ratio of actual percent of malaria cases to the estimated from the trend (Equation 4.22)

$$Y_{trend} = -1331+0.69* \text{ Year} \quad (4.21)$$

$$DY= (Y / Y_{trend})*100 \quad (4.22)$$

4.6.2 Correlation Dynamics

Figure 4.17 (D) shows DY and Temperature condition Index (TCI week 30) for moisture condition index (VCI) for years 1990-2006. From the Figure it can easily be found that TCI (week 30) has strong correlation to DY. Investigation included correlation analysis of DY versus VCI and TCI, shown in Figure 4.17(B). During cooler month (spring and fall) when mosquitoes are less active; correlation is low for both indices. During March and April when mosquito activity season starts, correlation between TCI and malaria incidence (DY) for Ampa increases to reach maximum (0.6 for TCI and -0.2 for VCI) at the end of April (week 20). During month May- August the correlation between TCI and DY is positive and high (more than 0.5) which can be interpreted as when the temperature decreases malaria activities increase. These results are compatible with mosquito's activity for this region.

4.6.3 Regression Analysis

In the statistical analysis we used both bivariate correlations and multiple regressions. The bivariate correlation reveals that DY is significantly related to TCI for weeks 27 to 32. In multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 26 to 32) values. Table 4.26 presents the results of fitting the ordinary least squares (OLS) regressions model approximated by equations (4.23) for Ampa

$$DY=b_0 + b_1TCI_{27}+b_2TCI_{28}+b_3TCI_{29}+b_4TCI_{30}+ b_5TCI_{31}+ b_6TCI_{32} \quad (4.23)$$

Table 4.26. Results of multiple linear regressions (OLS) of DY on the equation (4.22) for Ampa, Brazil R^2=0.75

	Parameter	Standard				Variance
Variable	Estimate	Error	t Value	Pr > |t|	Tolerance	Inflation
Intercept	48.67809	14.86099	3.28	0.0084	.	0
TCI27	1.76272	2.7766	0.63	0.5398	0.00895	111.78403
TCI28	-3.3755	6.51856	-0.52	0.6158	0.00169	590.54694
TCI29	9.26917	9.8519	0.94	0.369	0.00077648	1287.85505
TCI30	-12.96663	11.82338	-1.1	0.2985	0.00058226	1717.44859
TCI31	7.0096	10.71384	0.65	0.5277	0.00075312	1327.80521
TCI32	-0.40819	4.47397	-0.09	0.9291	0.0039	256.37708

Table 4.26 shows coefficients of equation (4.23) and that the value of R^2 is large 0.75. The p-values for the regression coefficients is high, which is not significant at $p<0.05$ level and very small tolerance and high variance inflation. R^2 for the equation (4.23) is 0.75 with p values 0.53 and 0.92 which are clear indication of collinearity among the predictor variables. In addition VH indices of neighboring weeks are highly correlated.

To avoid this multicollinearity problem, we used an alternative method of estimation, principal component regression (PCR), which results in better estimation and prediction than OLS.

Table 4.27. Eigenvector of correlation matrix Ampa, Brazil

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
TCI27	0.404885	-0.475497	0.314737	0.609925	0.355423	-0.112151
TCI28	0.409727	-0.383336	0.277996	-0.273028	-0.655866	0.321232
TCI29	0.414242	-0.236655	-0.23124	-0.612545	0.310519	-0.497285
TCI30	0.415901	0.06531	-0.664051	0.123372	0.173001	0.580214
TCI31	0.40803	0.430536	-0.201695	0.353888	-0.476815	-0.504856
TCI32	0.396401	0.617514	0.537162	-0.194369	0.299683	0.2131

4.6.4 Principal Component Regression

We found simulated percent of malaria time series using predictor variables TCI_{26} through TCI_{32} and VCI_{34} through VCI_{34} for the two principal components (Prin1, prin3) in the PCR methodology.

$$DY=48.37 +1.46\ TCI_{27} + 1.34TCI_{28}-0.77TCI_{29}-2.7\ TCI_{30}-0.7\ TCI_{31}+2.51TCI_{32} \quad (4.24)$$

Table 4.28: Results of principal component regression for equation (4.23) Ampa R^2=0.70, RMSE = 8.4

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	99.92518	4.88944	20.44	<.0001
Prin1	9.68803	2.11064	4.59	0.0004
Prin3	83.1171	30.34406	2.74	0.016

Figure 4.18 shows the simulated and observed malaria

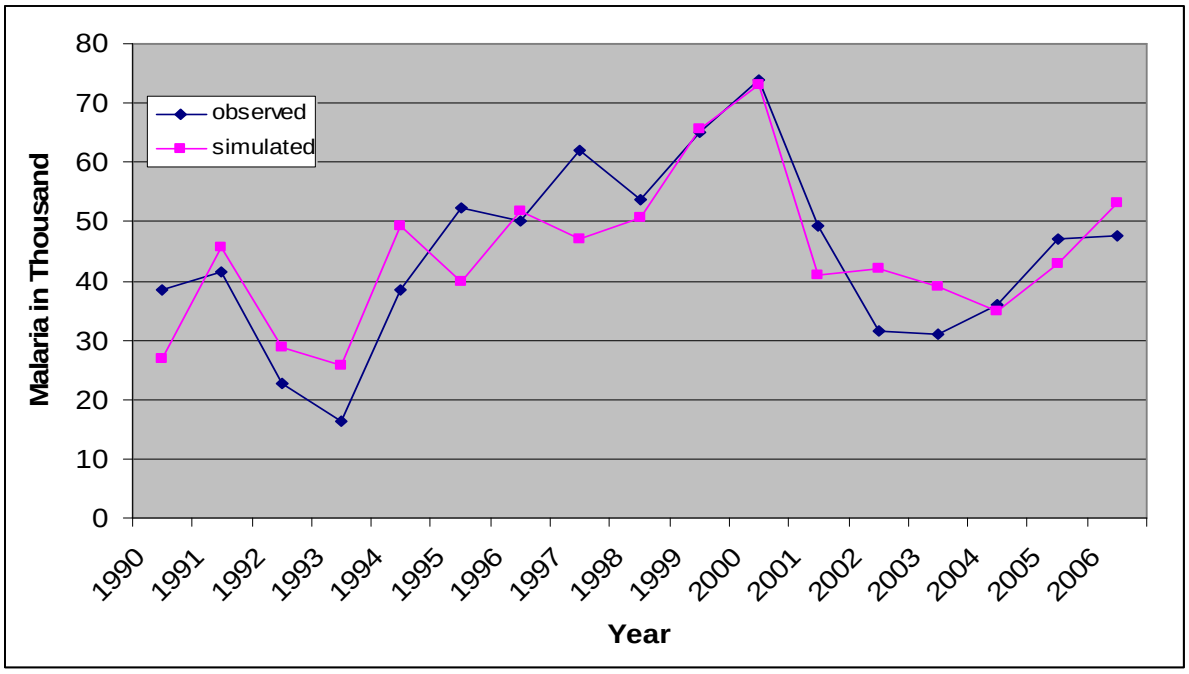
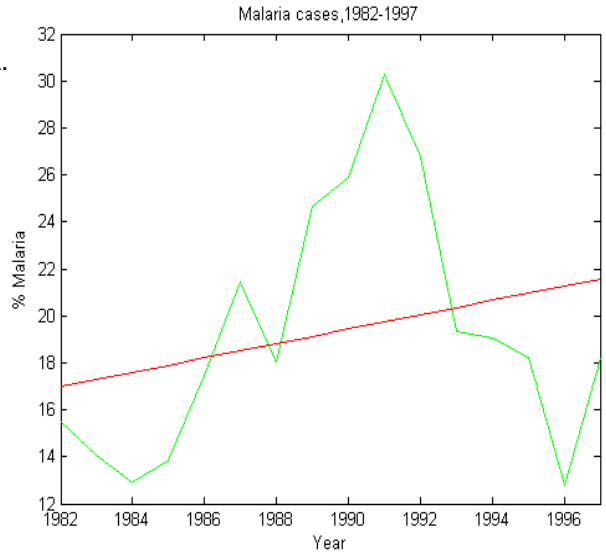


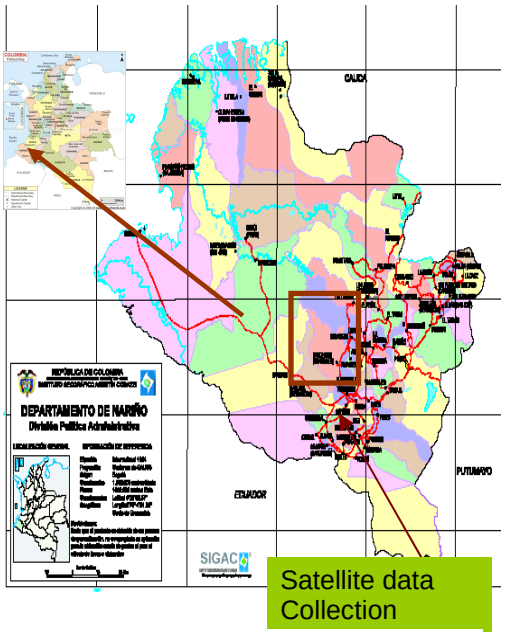
Figure 4.18 Independently simulated and observed malaria per thousand of population Ampa ($R^2 = 0.70$).

4.7 Modeling for Malaria Transmission in Narino, Colombia

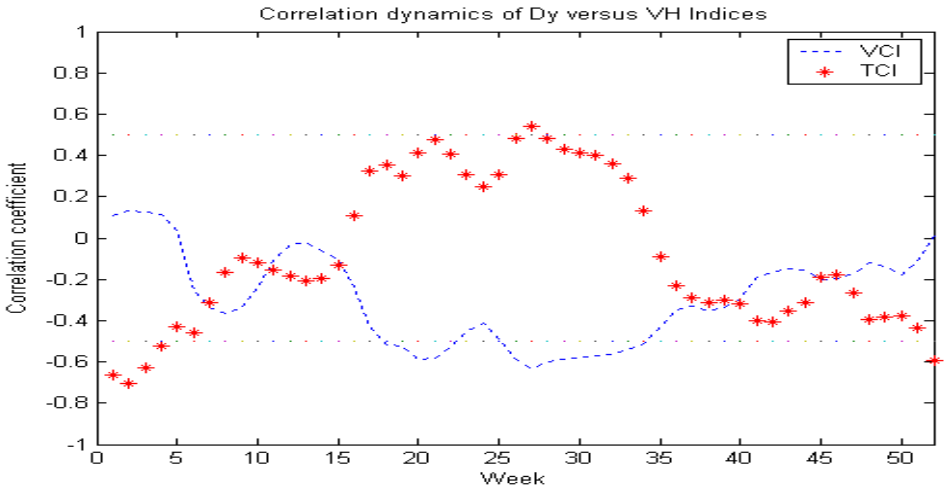
A.



B.



C.



D.

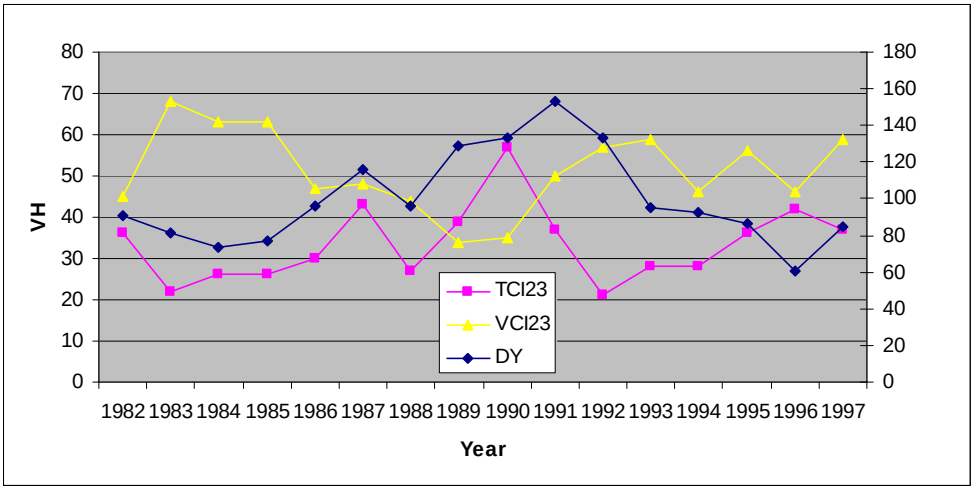


Figure 4.19: A. Percent of malaria in Narino department, Colombia and trend line B. District Map of Narino, rectangular box is pointed where Satellite data have been taken. C. Correlation coefficient dynamics of the percent deviation of malaria from trend versus TCI and VCI, D. Detrended value (DY), Temperature condition Index (TCI week 23) and Vegitation condition Index (VCI week 23) from year 1982-1997

4.7.1 Trend Analysis

The malaria time series (Figure 4.19A.) for Narino, Colombia was approximated by the linear equation (4.25). The weather-related variations around the trend were expressed as a ration of actual percent of malaria cases to the estimated from the trend (Equation 4.26)

$$Y_{trend} = -591.22+0.31* \text{Year} \quad (4.25)$$

$$DY= (Y / Y_{trend})*100 \quad (4.26)$$

Where Y is actual percent of malaria cases, Y_{trend} is the malaria estimated from trend.

4.7.2 Correlation Dynamics

Correlation and regression analysis have been performed using the deviations (DY) and TCI, VCI to investigate the association between them for malaria cases at Narino in Colombia.

Figure 4.19 (D) shows DY and Temperature condition Index (TCI week 23) for moisture condition index (VCI week 23) from year 1982-1997, which has clear indication that TCI(week 23) is positively correlated to variation of malaria(DY) involved for weather condition whereas VCI week 23 demonstrates opposite correlation with DY. Investigation included correlation analysis of trend malaria cases (DY) versus VCI and TCI, shown in Figure 4.19 C. During December and January correlation between TCI and DY is negative and high which can be interpreted as that during these months temperature and malaria activities are inversely related. During the month from May-July correlation between TCI and DY is positive and high which means that during these

months as temperature goes down malaria activities increase and if the temperature goes up the malaria activities go down. It should also be emphasized that correlation between VCI and DY is simply mirror reflection around X axis of TCI which in fact goes in line with the concept developed for vegetation health indices for vegetation stress.

4.7.3 Regression Analysis

The bivariate correlation reveals that DY was significantly related to TCI for weeks 22 to 24 and to VCI for week 22 to 24. Therefore, in multiple regression analysis DY was regressed on the linear combination of TCI (Weeks 22 to 24), VCI (week 22 to 24) values. Table 4.29 presents the results of fitting the ordinary least squares (OLS) regressions model approximated by equations (4.27) for Narino

$$DY=b_0 + b_1TCI_{22}+b_2TCI_{23}+b_3TCI_{24}+b_4VCI_{22}+ b_5VCI_{23} + b_6VCI_{24} \quad (4.27)$$

Table 4.29: Results of multiple linear regressions (OLS) of DY on the equation (4.26) Narino state, R^2=0.79

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	286.91196	67.36509	4.26	0.0021	.	0
TCI22	-0.50381	2.76718	-0.18	0.8596	0.02475	40.40415
TCI23	22.52055	8.22699	2.74	0.0229	0.00296	337.87181
TCI24	-24.02008	8.02042	-2.99	0.0151	0.0033	302.8251
VCI22	-11.16442	4.93544	-2.26	0.05	0.0076	131.53325
VCI23	34.80851	9.1099	3.82	0.0041	0.00213	469.52217
VCI24	-25.77301	6.93837	-3.71	0.0048	0.00339	295.08789

Table 4.29 shows coefficients of equation (4.26) and the value of R^2 is large 0.79. The p-values for the regression coefficients is high, which is not significant at $p<0.05$ level and very small tolerance and high variance inflation. The R^2 for the equation (4.27) is 0.79

with largest p-value of 0.92 which is clear indication of collinearity among the predictor variables. In addition, VH indices of neighboring weeks are highly correlated as seen in Table 4.29.

Table 4.30. Correlation Matrix among weekly TCI and VCI Narino, Colombia

	TCI22	TCI23	TCI24	VCI22	VCI23	VCI24
TCI22	1	0.9225	0.7906	-0.7248	-0.602	-0.4621
TCI23	0.9225	1	0.9517	-0.7052	-0.67	-0.6178
TCI24	0.7906	0.9517	1	-0.6019	-0.6305	-0.678
VCI22	-0.7248	-0.7052	-0.6019	1	0.9644	0.8443
VCI23	-0.602	-0.67	-0.6305	0.9644	1	0.9467
VCI24	-0.4621	-0.6178	-0.678	0.8443	0.9467	1

In order to avoid multicollinearity problem, we used an alternative method of estimation, principal component regression (PCR), which results in better estimation and prediction than OLS. Eigenvector is necessary to recalculate the coefficient of original variables.

Table 4.31. Eigenvector of correlation matrix Narino, Colombia

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
TCI22	-0.390069	0.451506	-0.509188	0.545678	0.289304	0.057056
TCI23	-0.421713	0.407421	0.086808	-0.609458	0.194383	-0.489292
TCI24	-0.402902	0.363214	0.570651	0.111032	-0.40429	0.452022
VCI22	0.420477	0.309851	0.470303	0.102715	0.69248	0.126205
VCI23	0.418162	0.430867	0.088582	0.323764	-0.451281	-0.56848
VCI24	0.394986	0.465372	-0.422514	-0.450661	-0.178562	0.462512

4.7.4 Principal Component Regression

Using PCR methodology, the variables in model equation (4.26) principal component estimators are found and are shown in Table 4.31. We found simulated percent of malaria time series using predictor variables TCI_{22} through TCI_{24} and VCI_{22} through VCI_{24} for four principal components (Prin1,Prin3, prin4 and prin6) .

$$DY=206 +1.39\ TCI_{22} + 23.87\ TCI_{23}-26.35TCI_{24}-7.03\ VCI_{22} +32.44\ VCI_{23}-26.55VCI_{24} \quad (4.28)$$

Table 4.31: Results of principal component regression for equation (4.28) Narino, Colombia R^2=0.74, RMSE = 2.79

	Parameter	Standard		
Variable	**Estimate**	**Error**	**t Value**	**Pr > |t|**
Intercept	99.91562	3.98142	25.1	<.0001
Prin1	-5.2505	1.89497	-2.77	0.0182
Prin3	-11.19929	6.97608	-1.61	0.1367
Prin4	65.83612	30.20639	2.18	0.0519
Prin6	-533.31639	147.71637	-3.61	0.0041

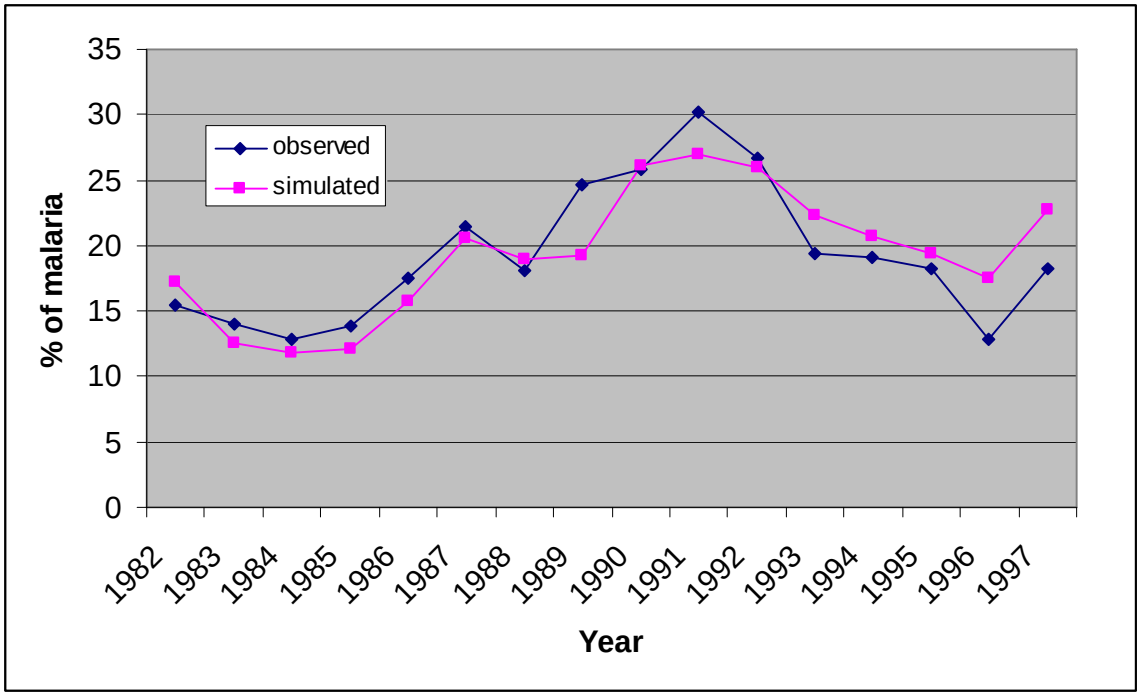


Figure 4.20: Independently simulated and observed % malaria (R^2 = 0.70).

4.7.5 Contribution and Summary

The result of this study showed that AVHRR-based vegetation health indices (VCI and TCI) can be used as a proxy for numerical estimation of number of malaria cases in some regions of South America(Narino in Colombia and Rondonia and Ampa district in Brazil). The number of malaria cases is larger for lower temperature for Ampa, Brazil whereas the relation is simply opposite in the case of Rondonia, Brazil. Strong negative correlation for TCI 15 with malaria explains that malaria proliferation occurs if temperature increases in this region during March-April. A possible reason is that Rondonia is tropical rain forest where almost whole year there is rainfall with very high humidity and average temperature is much lower (high cloud cover) than Ampa. If the temperature increases humidity will decrease and better condition for malaria

transmission will persist. Now, If we look at the ecology and weather condition of Ampa we can see average temperature is higher (which ultimately reduces humidity) than that of Rondonia. Thus, temperature needs to fall which will increase humidity and better condition for malaria transmission will prevail. For malaria proliferation there is optimum range of temperature (27-30 degree Celsius) and humidity (60%-80%). Too much or too little of these two parameters restrict proliferation. For both of the regions, number of malaria cases for the whole year can be detected numerically in first few months of the year. Thus I found that AVHRR-based thermal condition index (TCI) can be used as a proxy for numerical estimation of number of malaria cases in Rondonia and Ampa in Brazil. I also developed model for estimating malaria earlier than the major malaria season.

In case of Narino Department, Colombia both thermal and moisture conditions of months April and May are highly correlated to malaria cases in the year. So our model for malaria based on satellite data in South America would be very helpful for Ministry of Health to take proper prevention measurement for the whole year. Unfortunately, I didn't have access for malaria data after 1997. Whenever new dataset for malaria shall be available we can modify the model to see if the nature of malaria transmission will change.

4.9 Results and Models

The results for all eight regions are given in brief in the following Table (4.32).

Table 4.32: Results of all eight regions for correlation dynamics and Principal component Reegression

Site	Regions	Correlation Coefficient and Sensitive period		Principal component Regression	
		TCI	VCI	R square	RMSE
1	Bandarban, Bangladesh	+0.56(July and Sep)	Not significant	0.86	1.78
2	Bikaner, Rajsthan, India	+0.6(June-July)	Not significant	0.74	1.36
3	Tripura, India	-0.8(March)	-0.5(Feb)	0.74	0.58
4	Gujarat, India	-0.8(Feb-Mar)	Not significant	0.82	0.82
5	Orissa, India	+0.6(Feb-Mar)	+0.5 (Sep)	0.70	0.49
6	Rondonia, Brazil	-0.6(April)	Not significant	0.68	21
7	Ampa, Brazil	+0.6(May-June)	Not significant	0.70	8.4
8	Narino, Colombia	+0.5(April-May)	-0.5(April – May)	0.74	2.79

The principal component regression models for all eight regions are given below.

Bandarban, Bangladesh:

$$DY= 86.48- 0.98\ TCI_{32}-0.36\ TCI_{33}+0.61\ TCI_{34}+ 2.20TCI_{35}-1.31TCI_{36} \quad (3.17)$$

Bikaner, Rajsthan, India:

$$DY=18.15 -5.91\ TCI_{26} + 17.26\ TCI_{27}-32.88TCI_{28}+49.93\ TCI_{29}-44.07\ TCI_{30}+19.69TCI_{31} -2.87TCI_{32} - 25.07\ VCI_{34}+52.56VCI_{35}-27.59VCI_{36} \quad (4.4)$$

Tripura, India:

$$DY=129.7 - 0.07\ TCI_{15}-0.08\ TCI_{16}-0.08\ TCI_{17}-0.09\ TCI_{18}-0.09\ TCI_{19}-0.12TCI_{20} \quad (4.8)$$

Gujarat, India :

$$DY=364.12 - 4.29\ TCI_{10}-43.77\ TCI_{11}+123.25\ TCI_{12}-88.52\ TCI_{13}+31.25\ TCI_{14}-1.68TCI_{15}-73.11\ TCI_{16}+83.63\ TCI_{17}-34.30TCI_{18} \quad (4.13)$$

Orissa, India:

$$DY=97.69 + 0.18\ VCI_{3}-1.56\ VCI_{4}+1.12\ VCI_{5}-0.005VCI_{6}-0.78\ TCI_{10}+1.85TCI_{11}+0.3TCI_{12}+0.03TCI_{13}+0.30TCI_{14}+0.03TCI_{15} \quad (4.16)$$

Rondonia, Brazil :

$$DY=18.15 -5.91\ TCI_{14}+17.26\ TCI_{15}-32.88TCI_{16}+49.93\ TCI_{17}-44.07\ TCI_{18}+19.69TCI_{19}-2.87TCI_{20} \quad (4.20)$$

Ampa, Brazil :

$$DY=48.37 +1.46\ TCI_{27}+1.34TCI_{28}-0.77TCI_{29}-2.7\ TCI_{30}-0.7\ TCI_{31}+2.51TCI_{32} \quad (4.24)$$

Narino, Colombia:

$$DY=206 +1.39\ TCI_{22}+23.87\ TCI_{23}-26.35TCI_{24}-7.03\ VCI_{22}+32.44\ VCI_{23}-26.55VCI_{24} \quad (4.28)$$

CHAPTER FIVE

El-Nino Southern Oscillation, Vegetation Health, and Malaria

5.1 Introduction

El-Nino Southern Oscillation (ENSO) is a global climatic phenomenon caused by a rapid increase of sea surface temperature in the tropical Pacific. Many regions across the world have seen recurring malaria epidemics every five to eight years, which may have some association with the ENSO cycle. This association has led a number of researchers to believe that the El Niño Southern Oscillation (ENSO) and malaria epidemics are correlated. Subsequently, in many of the studies, it has been proposed that malaria epidemics in the former British Punjab, Pakistan, Sri Lanka, the highlands of Uganda, Columbia, Argentina, Ecuador, Peru, and Bolivia are somehow linked with ENSO cycles (Kilian *e*t al. 1999). The recent strong 1997-1998 EI Nino and the following 1998-1999 La Nina caused disturbance to weather, environment, economy, and human lives worldwide. The total impact of these events (ENSO) on societies is estimated in billions of dollars and consequences include famine, human health problems, loss of life, property damage, and destruction of the environment. I examine the last 15-year association between monthly sea surface temperature (SST) anomalies in the tropical Pacific and vegetation, health (condition) index and malaria incidences for the South East Asia and Colombia. The ENSO phenomenon mainly affects the coastal areas of tropical and subtropical climate zones. One of the goals of this study was to investigate and to quantify the relationship among variability of ENSO, vegetation health and malaria in South East Asia and South America. This relationship was investigated by the examination of correlations between Vegetation condition index (VCI) and Temperature

condition index and monthly ENSO indices, Sea Surface Temperature Anomalies (SSTA) and Southern Oscillation Index (SOI).

The El-Nino Southern Oscillation (ENSO) refers to an irregular cycle of warming and cooling of the sea surface temperatures (SSTs) in the eastern equatorial Pacific Ocean (Joseph S. D'Aleo 2002). This oceanic warming and cooling named El Nino and La Nina, respectively is accompanied by changes in air pressure patterns above the Pacific Ocean and affects not only the local atmosphere circulation but also has repercussions on the climate worldwide. The recent capability of prediction of El Nino has improved the ability to forecast temperature and rainfall conditions up to a few months in advance. Seasonal forecasts of temperature and rainfall have the potential to provide important societal benefits particularly for the prediction, control and prevention of climate-sensitive diseases such as malaria. Thus a better understanding of relationship between The El-Nino Southern Oscillation (ENSO), the climatic anomalies it engenders, and malaria epidemics could help the world wide increase in incidence of mosquito-transmitted diseases. One of the the purposes of this work is to assess the possibility of using ENSO forecasts incorporating Vegetation Health for improving malaria control.

5.2 Indices to measure ENSO

Since ENSO is coupled with atmospheric and oceanic phenomenon, condition in both the atmospheric and ocean can be used to evaluate the ENSO state and provide a measure of the strength of an event. The three most commonly used measures of ENSO used today are the SOI (southern oscillation index), sea surface temperature anomaly (departure from normal) in the eastern and central Tropical Pacific and the new

multivariate ENSO index (MEI). Monthly ENSO measures can be found on the website. The National Center on Environmental Prediction's Climate Prediction Center provides indices including SOI and sea surface temperature anomalies. Most of the indices are updated monthly and includes historical database.

5.3 Southern Oscillation Index (SOI)

The SOI is the oldest measure and was first designed by Walker in the 1920s and revised later in the 1950s by Willet. SOI compares the pressure (relative to normal) at Darwin, Australia and the central Pacific Island of Tahiti. In El Ninos, pressures are unusually high in the western Pacific (Darwin) and unusually low in the central pacific (Tahiti). (Joseph S. D'Aleo 2002)

5.4 Sea Surface Temperature Anomalies in the Tropical Pacific

If El Ninos are warm ocean water events and La Ninas are cold ocean water events, it is not surprising that sea surface temperature would be useful as a measure of ENSO. The region in the Central Pacific, where the warm and cold plumes of the ENSO cycle is most apparent, stretches from along the northwest coast of South America west along the equator to beyond the International dateline. Region 3.4 is defined as the region 120 to 170 degrees west longitude and 5 degrees north and 5 degrees south. Sea surface temperature anomaly in this region plays a significant role on weather condition in many parts of the world.

5.5 Impact of ENSO on Vegetation Health and Malaria in Tripura

The impact of ENSO was investigated by the examination of correlations between weekly Vegetation condition index (VCI), Temperature condition index and monthly ENSO indices, Sea Surface Temperature Anomalies (SSTA) and Southern Oscillation Index (SOI) for ten years. The results demonstrated the strong negative correlation exists (Figure 5.1) between monthly SSTA and yearly DY (detrended malaria incidence) during the month of January. So temperature rise in Pacific Ocean will affect malaria transmission in Tripura (south east India) which can be detected six months earlier than major malaria transmission period (pre and post monsoon time). Furthermore prediction of ENSO can forecast malaria in Tripura even earlier. I also calculated the correlation between monthly southern oscillation index with yearly DY (detrended malaria incidence) which demonstrates (Figure 5.3) positive correlation in the month of January and February. This result also justifies the correlation dynamics I found with SSTA which is opposite to this phenomena. More analyses is done to see if moisture and thermal conditions are also sensitive to SSTA and SOI as TCI and VCI were found sensitive to malaria incidence. I again looked at the SSTA (month by month) (Figure 5.2 b) and VCI and TCI for 52 weeks and found quite exciting result for VCI and partially good results with TCI. January SSTA is significantly correlated with VCI during the month of March which again justifies the results I found (results of Tripura in chapter four) the sensitivity of malaria with VCI is negative during the same month in Tripura. The contribution of this research is also that I found sensitivity of ENSO (January) to thermal and moisture conditions (May-June) in Tripura area which can also be used for early warning on malaria. Opposite phenomena is found while I compared correlation

between SOI and VCI and TCI. The weather and ecosystem of Tripura can be characterized as one of the regions which is highly vulnerable to ENSO events

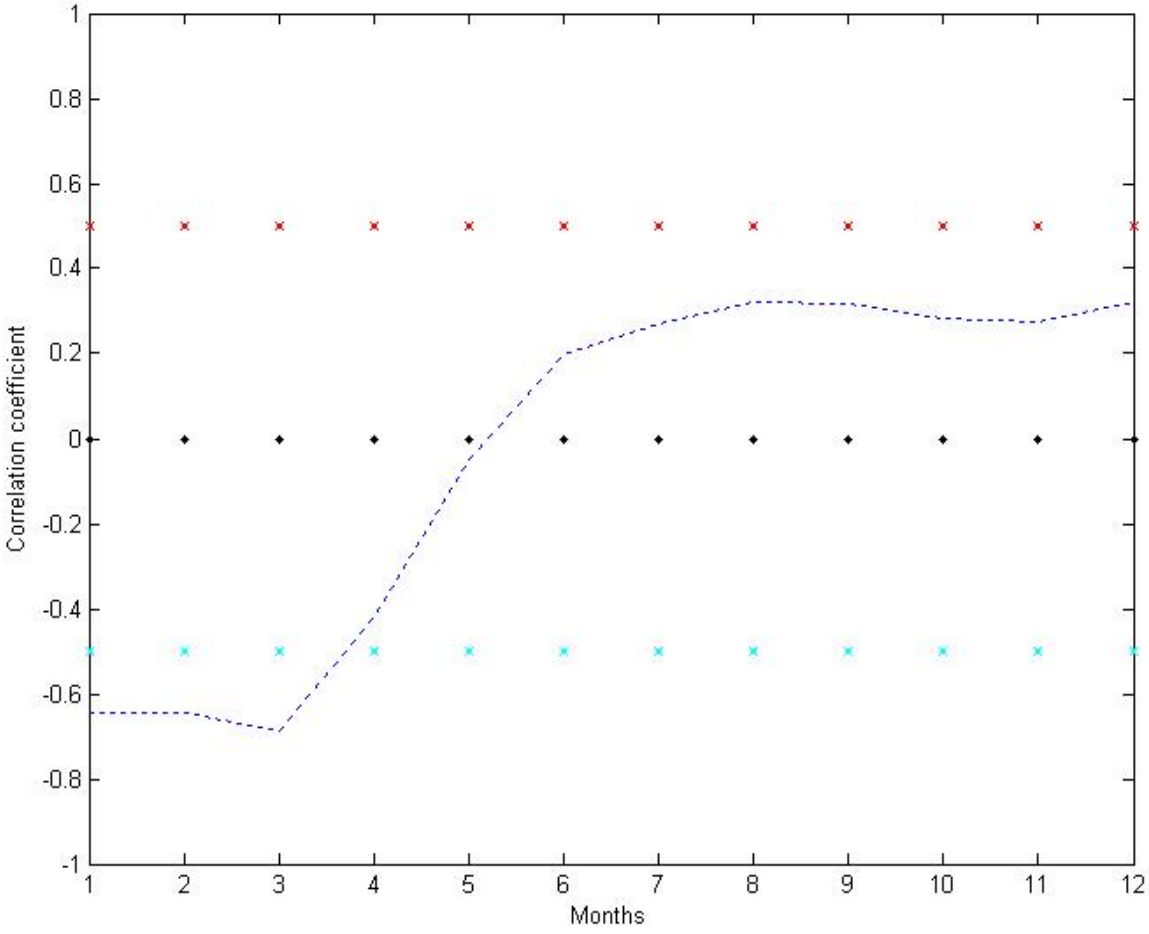
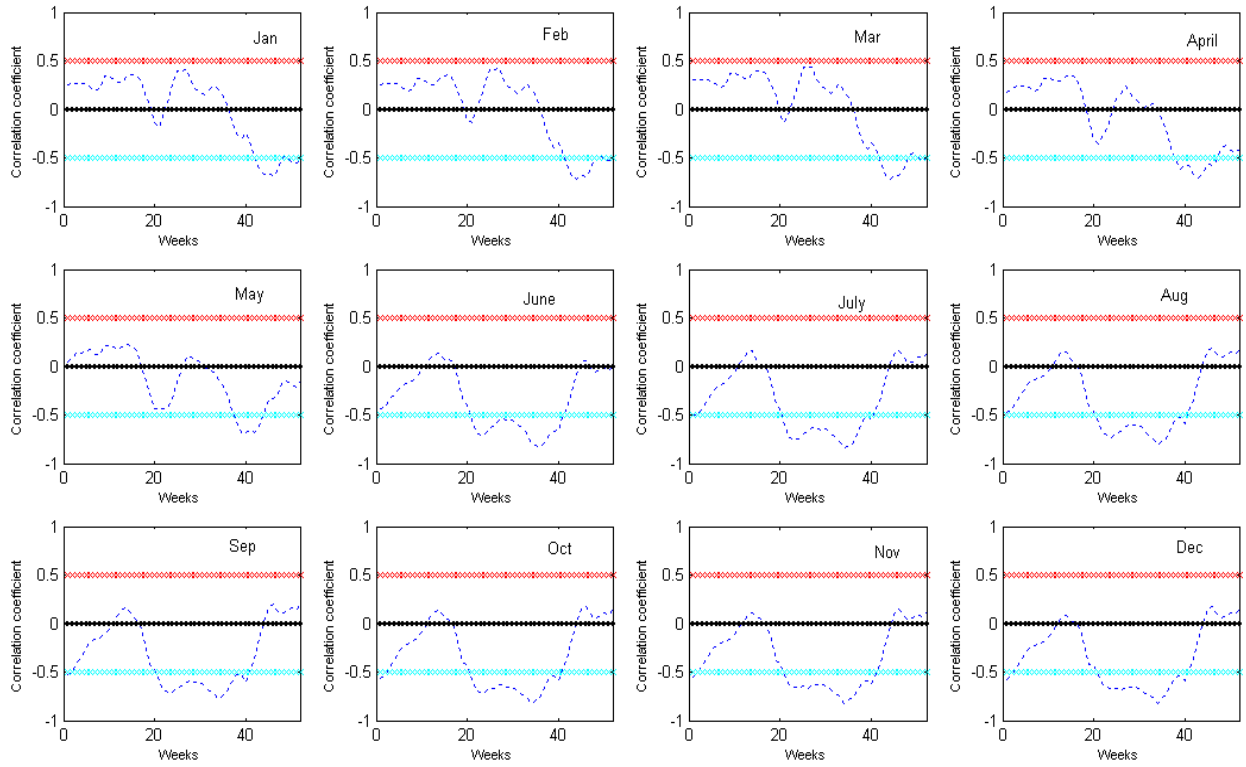
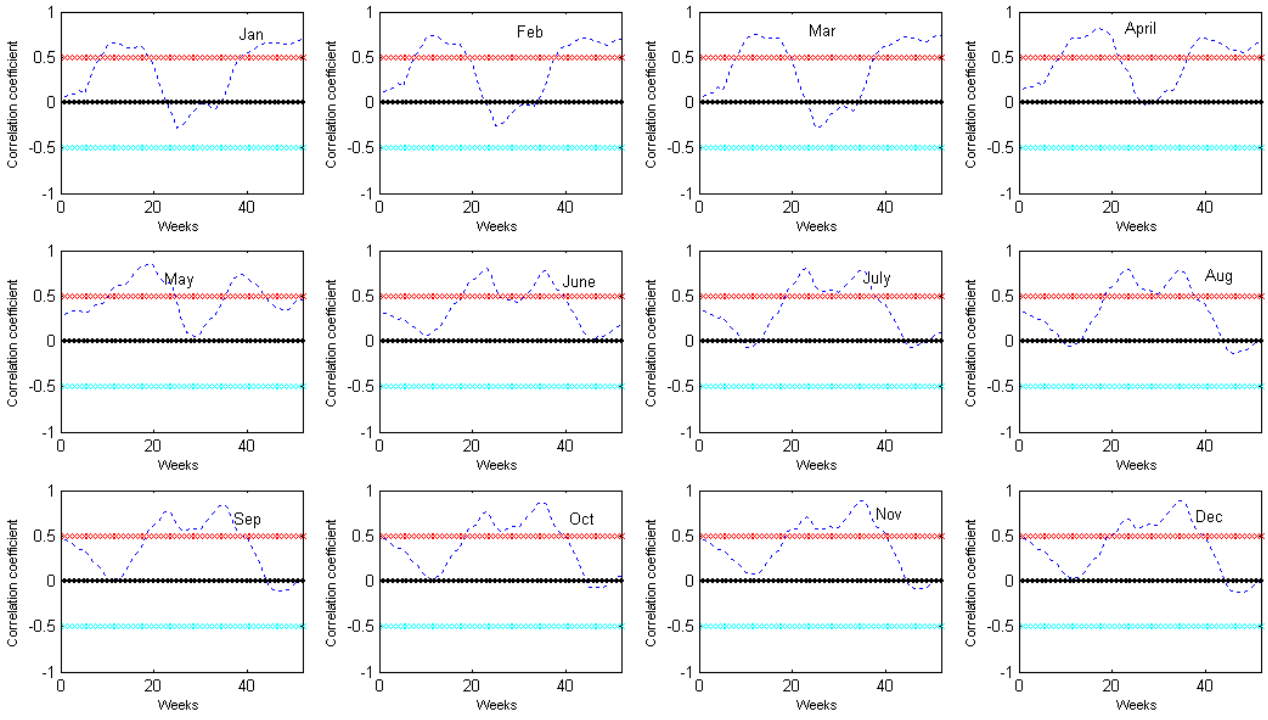


Figure 5.1: Correlation coefficient dynamics of the percent deviation of malaria from trend in Tripura state versus monthly SST anomaly at central pacific zone 3.4 Tripura, India



(a) TCI-SST



(b) VCI-SST

Figure 5.2: a) Correlation coefficient dynamics of TCI for 52 weeks to SST anomaly for each month b) Correlation coefficient dynamics of VCI for 52 weeks to SST anomaly for each month inTripura, India

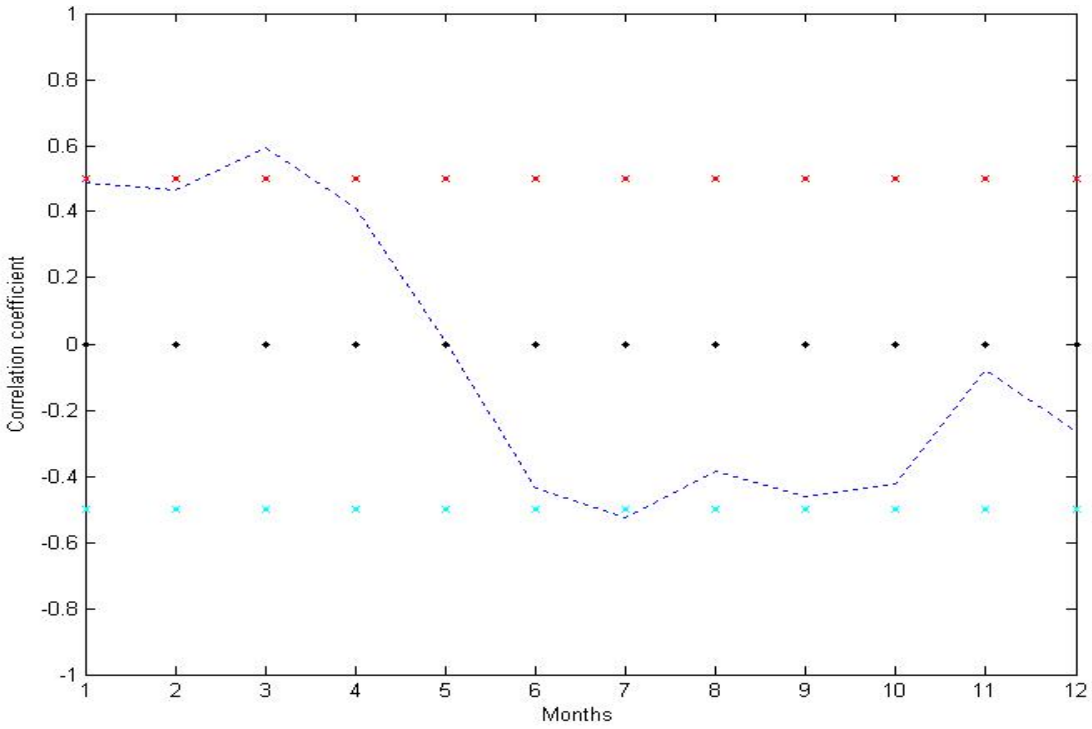
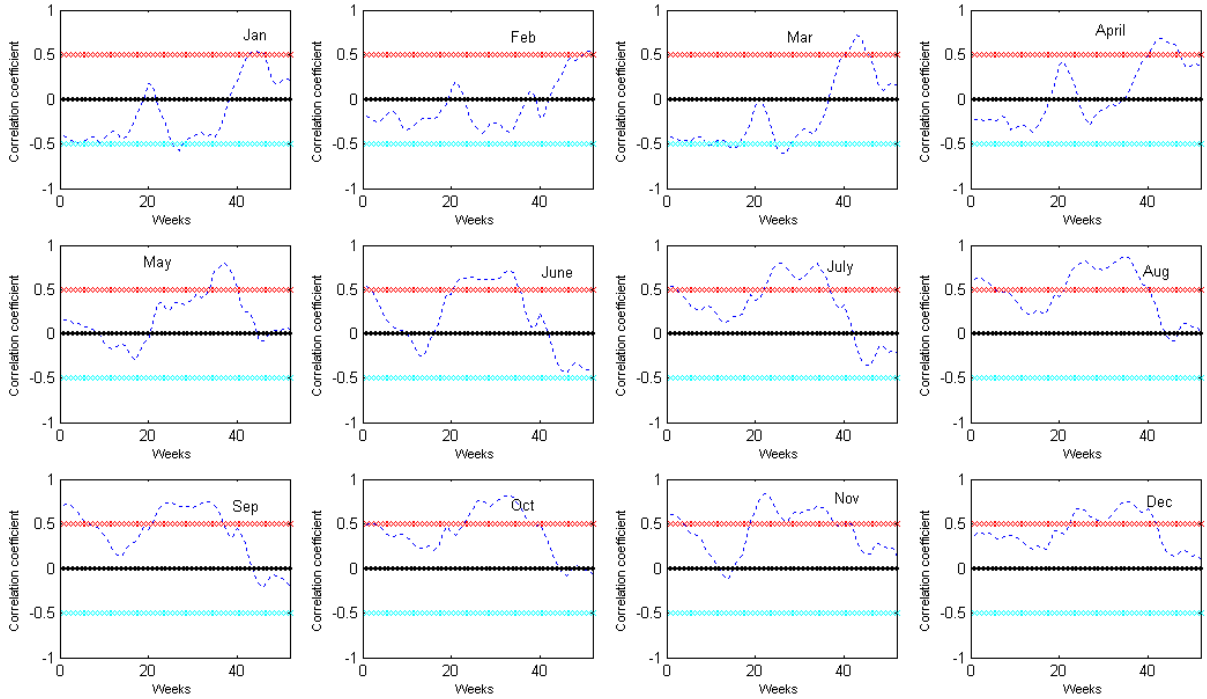
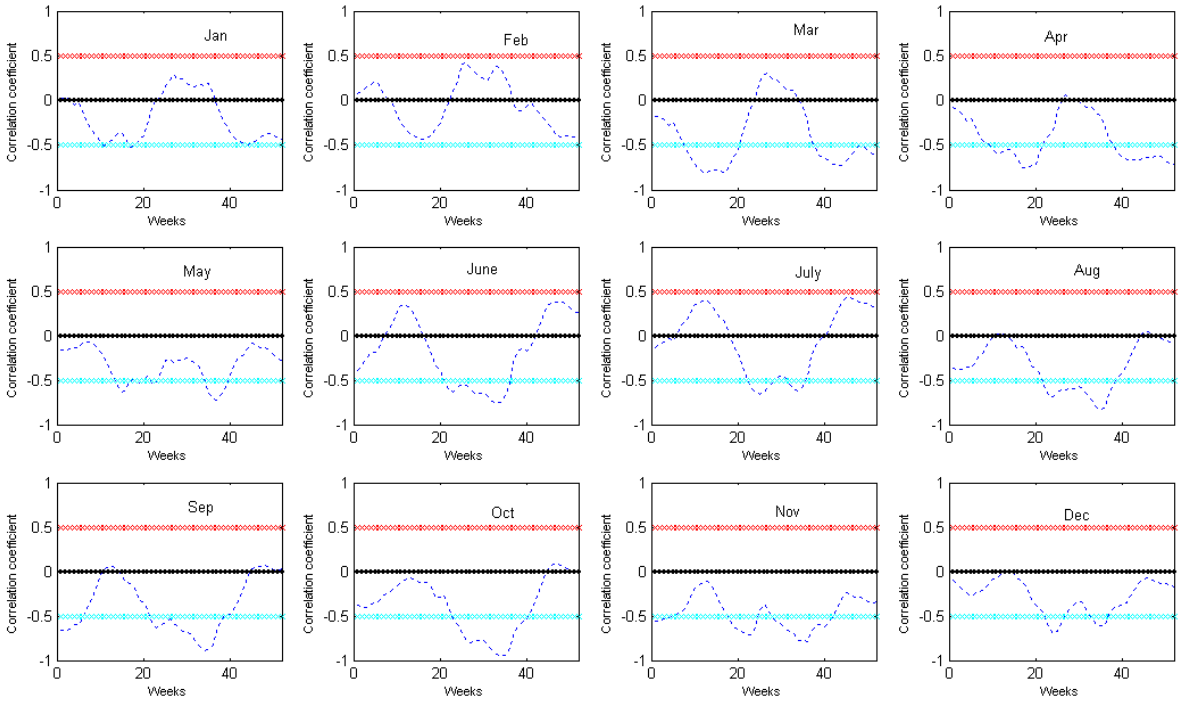


Figure 5.3: Correlation coefficient dynamics of the percent deviation of malaria from trend versus monthly SOI anomaly in Tripura, India



(a) SOI-TCI



(b) SOI-VCI

Figure 5.4: a) Correlation coefficient dynamics of TCI for 52 weeks to SOI for each month b) Correlation coefficient dynamics of VCI for 52 weeks to SOI for each month in Tripura, India

5.6 Impact of ENSO on Malaria in Bandarban, Bangladesh

The impact of ENSO was investigated by the examination of correlations between weekly Vegetation condition index (VCI) and Temperature condition index and monthly ENSO indices, Sea Surface Temperature Anomalies (SSTA) and Southern Oscillation Index (SOI) for 14 years in Bandarban district Bangladesh which is around 1000 meter above sea level. The results demonstrated nominally negative correlation (Figure 5.5 A) between monthly SSTA and yearly DY during the month of July which is only two months before the malaria season in this area. So during the time of la Nina temperature fall in pacific ocean will affect malaria transmission in Bandaraban . However prediction of strong la Nina can forecast malaria couple of months earlier. I also calculated the correlation between monthly southern oscillation index with yearly DY (detrended malaria incidence) which does not demonstrate (Figure 5.5 B) good correlation prior to malaria season. More analysis is done to see if moisture and thermal conditions are also sensitive to SSTA as TCI and VCI were found sensitive to malaria incidence. I again looked at the SSTA month by month (Figure 5.5 C) and VCI and TCI for 52 weeks. January SSTA is significantly correlated with TCI during the month of April. This dynamics of correlation demonstrates a significant pattern of four month lead correlation dynamics for other months. The contribution of this research is also that I found

sensitivity of ENSO (Januaary) to thermal and (April-June) in Bandarban area which can also be used for early warning on malaria.

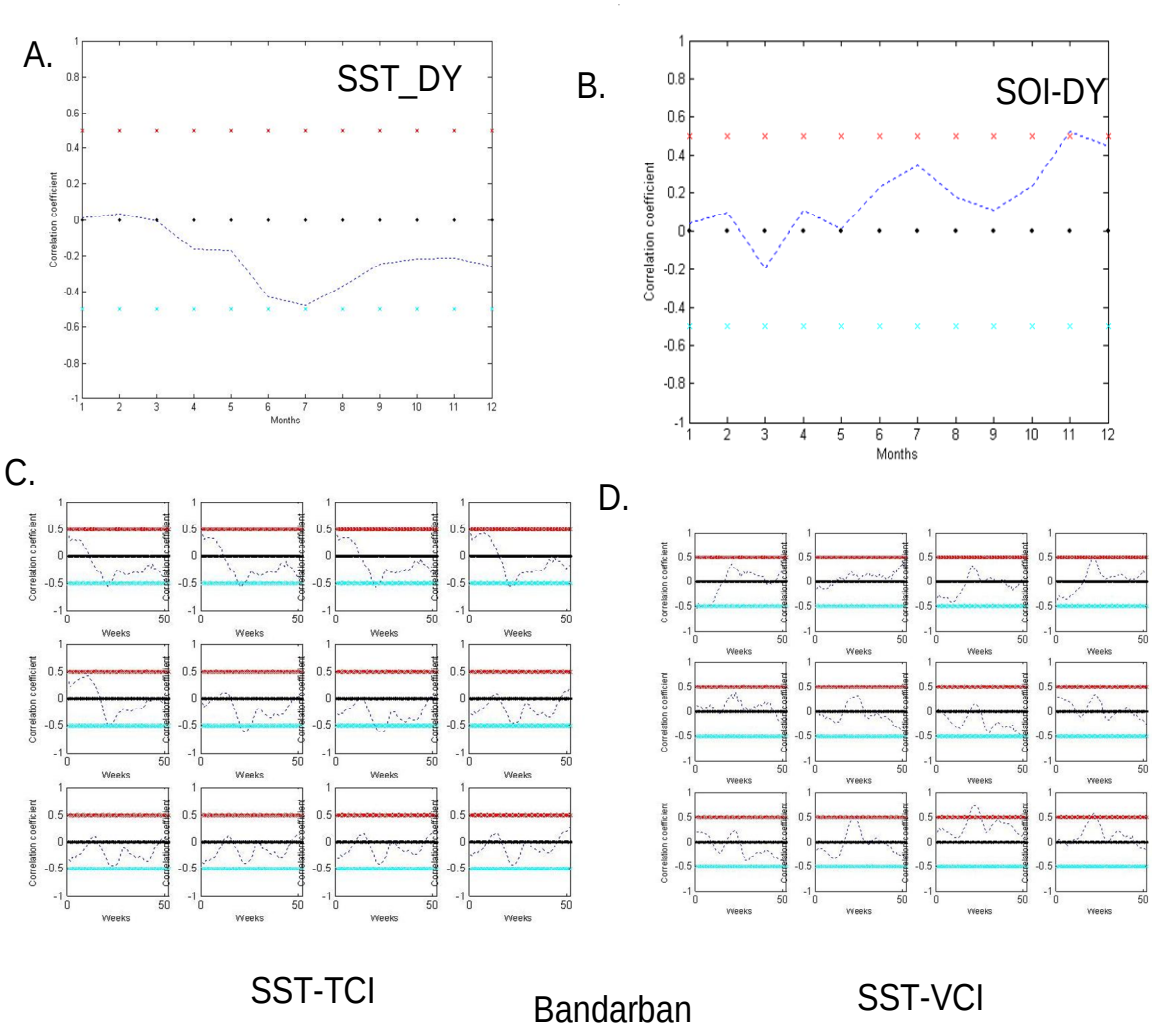


Figure 5.5: Correlation coefficient dynamics of the percent deviation of malaria from trend in Tripura state versus monthly SST anomaly at central pacific zone 3.4 Tripura, India

5.7 Impact of ENSO on Malaria in Colombia

Similar characteristics have been found in Narino, Colombia as Bandarban, Bangladesh. They share similar characteristics in their ecosystem, for example both are situated near costal area, both experience heavy rainfall during rainy season and both are 1000 meters above sea level.

The impact of ENSO was investigated by the examination of correlations between weekly Vegetation condition index (VCI) and Temperature condition index (TCI) and monthly ENSO indices, Sea Surface Temperature Anomalies (SSTA) and Southern Oscillation Index (SOI) for 1982-1997 at Narino, Colombia which is around 1000 meter above sea level. The results demonstrated poor correlation (Figure 5.6 A) between monthly SSTA, SOIA and yearly DY. However we calculated correlation using the same parameters for another department (Choco) (Figure 5.6 C) in Colombia, malaria prone area and 2-3 meter altitude from sea level and close to Narino. In Choco the correlation dynamics showed that malaria situation is sensitive to SSTA at the pacific and it could be detected as early as January for the whole year. Malaria transmits over the whole year (perennial) in this region. Correlation dynamics of SST-TCI (Figure 5.7 A), SOI-TCI (Figure 5.7 C), SST-VCI (Figure 5.7 B), SOI-VCI (Figure 5.7 D), also shows a significant sensitivity of ENSO signal four months earlier. In conclusion, we found there are some areas particularly costal regions that are sensitive to ENSO signal which can be used for malaria early warning system for making better decision for eradicating vector borne diseases like malaria.

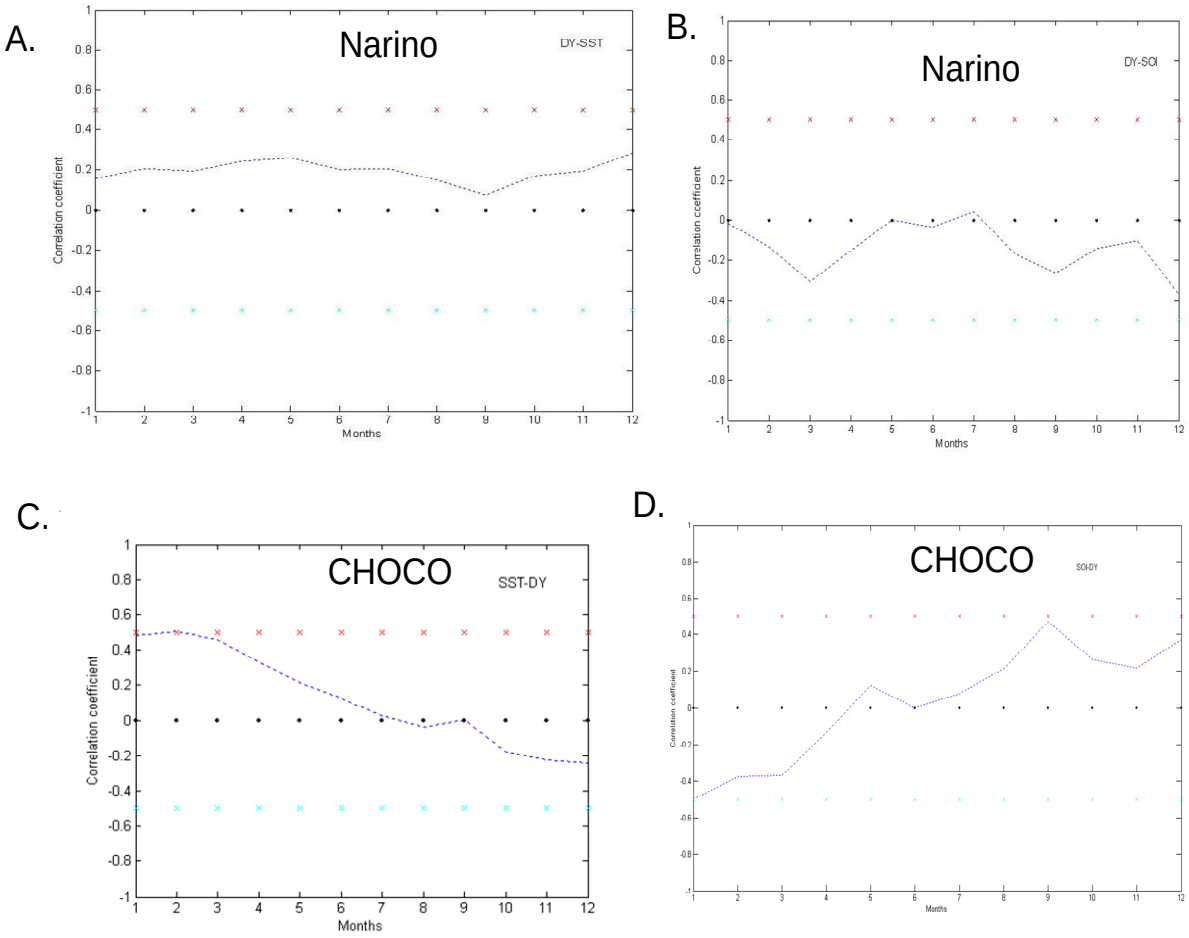


Figure 5.6: Correlation coefficient dynamics of the percent deviation of malaria from trend in Tripura state versus monthly SST anomaly at central pacific zone 3.4 Tripura, India

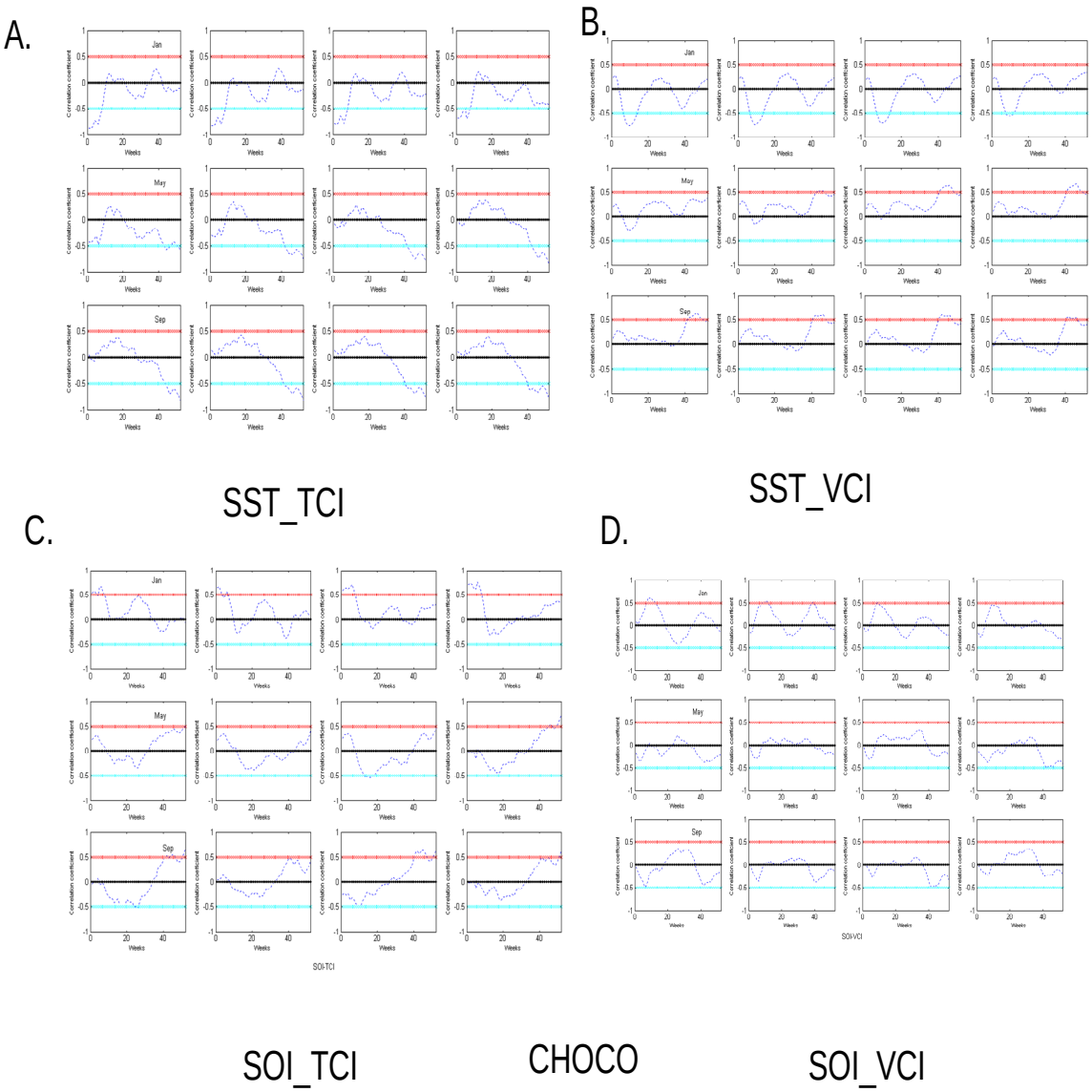


Figure 5.7: Correlation coefficient dynamics of the percent deviation of malaria from trend in Tripura state versus monthly SST anomaly at central pacific zone 3.4 Tripura, India

CHAPTER SIX

Conclusion

Malaria forecast model can provide important information for government health agencies. Malaria early warning is currently based on field surveys and meteorological data, although official forecasts does not exist for many areas where meteorological data are not easy to collect regularly. This methodology is typically expensive and time consuming.

In addition, over the years, large area assessment of malaria has been based on complex in-situ observations such as weather and climate data, ecological properties, technological and other factors using statistical methods. This approach provides quite reasonable large area prediction. However, one of the principal shortcomings of this approach is that in situ observations are not dense enough to adequately represent the volatile nature of spatial environmental patterns. An additional problem with the availability of in situ observations often appeared due to economic, political and social situations.

In order for malaria prediction system to be economical, more efficient methods for data collection analysis are necessary. Since satellites cover large areas in a short amount of time at a relatively low cost, remote sensing is a potential approach to provide reliable, timely and cost effective predictions.

The contributions of this research are the following: I explore the sensitivity of regional ecosystem condition to malaria events for the period of available satellite and

malaria data set. I have found malaria transmission of Bandaban in Bangladeh, Tripura, Orissa, Bikaner (Rajsthan) Gujarat in India , Rondonia, Ampa in Brazil, and finally Narino in Colombia are sensitive to satellite data of thermal (TCI) and moisture condition index derived from AVHRR. I have been able to identify some features of ecosystem in malaria transmission, for example, malaria at Bandarban in Bangladesh, Orissa, and Bikaner in India, Ampa in Brazil, Narino in Colombia exhibit strong positive correlation with thermal condition index (TCI) at least six weeks earlier than the beginning of major transmission season starts. This phenomenon explains the ecosystem, weather condition and transmission behavior of malaria vectors in those areas. In Bandarban, Bangladesh larger amount of rainfall will create favorable condition for mosquito to breed more since the area is forest fringe lots of places are there for water to get stagnant. On the other hand, the weather of Orissa, and Bikaner are very dry and hot during summer which will not permit malaria vector to survive long. So more rainfall (more cloud) and lower temperature is necessary in these areas for malaria transmission. An extra feature can probably be mentioned in Orissa where two malaria vectors dominate two different seasonal cycles which we could see from two mutually exclusive peaks in the correlation dynamics, one with TCI (April-May) and another is VCI (August-September). The average temperature of Ampa Brazil is also hot and moderately humid. Lower temperature and higher humidity will increase malaria transmission which we found in our correlation dynamics (positive relation with TCI). Narino, Colombia exhibits almost similar phenomena as Bandarban, Bangladesh which is a kind of high altitude with forest fringe.

We found strong negative correlation between malaria with TCI in Tripura, Gujarat in India and Rondonia in Brazil . Tripura is forest fringe low altitude from sea level and has very high humidity (almost 90%) during March and April which is a strong barrier for malaria transmission. Since increasing temperature will decreases humidity for the same amount of moisture content, it is very clear why TCI and malaria in this area exhibits negative correlation. Rondonia in Brazil is tropical rain forest, the temperature is moderate, humidity and rainfall are very high. Higher temperature (lower TCI) helps drop the humidity and spread malaria easily.

I have developed remote sensing data-based models to compute malaria cases for those above mentioned regions using principal component regression after I have found that the correlation dynamics goes along in line with environmental features in those areas. These eight models of eight different regions of distinct ecological and environmental features can measure the number of malaria cases much prior to major malaria season in the respective area area. The forecasts were highly accurate, as judged by the percentage of malaria variations explained by the forecast model. Prediction for most malaria cases relied on satellite data measurements well before transmission season, allowing longer lead times than existing procedures. The results of cross–validation leave one out analysis suggests malaria can be predicted with fairly high accuracies based on remote sensing data.

Since our models can predict near about two months before malaria season we try to develop another alternative to predict malaria for long term which will give us additional conformation for malaria transmission in decision making. I also analyzed the impact of ENSO signals on malaria by examining the correlations between Vegetation condition

index (VCI) and Temperature condition index and monthly ENSO indices, Sea Surface Temperature Anomalies (SSTA) and Southern Oscillation Index (SOI). The results demonstrated strong correlation between malaria incidence and SSTA during the month of January. This promising result gives an opportunity to forecast malaria even long before the malaria season starts. A six months prior shifting pattern of good correlation dynamics between SSTA (month by month) and VCI and TCI for 52 weeks have been found. More strong relationship is found using SOI index for ENSO signal as compared to Nino 3.4 index. I found the weather and ecosystem of Tripura (India), Choco (Colombia), Bandarban (Bangladesh) can be characterized as one of the regions which is highly vulnerable to ENSO events. Some areas do not show at all any significant relations between VH and ENSO events. So we can not depend on this result to a great extent. Therefore we incorporate both satellite based sensitivity and ENSO results for short term and long term prediction of malaria.

This current research was limited to a few areas of four countries. A relevant question is how other regions can be modeled and whether other variables such as amount of biomass or fractional vegetation would improve forecast accuracies. The models developed in this study are promising for forecasting malaria based on satellite measurements. Such forecast could be of great relevance to management of vector borne diseases.

Reference

Alberto, G., Angel O., Michel V. H. and Armando A.J., 2007, Forecasting malaria incidence based on monthly case reports and environmental factors in aruzi,Burundi,1997 –2003, Malaria Journal, 6 :129

Anne, E. J., Ulrika U. W., Andrew P. M., Ian M. H. and Alexandre S. G.,2007, Climate prediction of El Niño malaria epidemics in north-west Tanzania, Malaria Journal ,6 :162

Ashwani, K., Neena V., Tanu J. and Aditya P. D., 2007, Burden of Malaria in India: Retrospective and Prospective View, American Journal of Tropical Medicine and Hygiene77(Suppl 6), 69–78

Barrera, R., Grillet, W.E., Rangel, Yadira, Berit, Jesus and Ache, A., 1999, Temporal and spatial patterns of malaria reinfections in north-eastern Venezuela. American Journal of Tropical Medicine and hygiene. 61:784-790.

Bi, P., Tong, S, Donald, K. and Parton, K., 2003, Climatic variables and transmission of malaria: A 12 year data analysis in Shuchen County, China. reference to the carries of malaria and their control. Publication of the South African Institute for Medical Research. 4: 1-179.

Brockwell, P. J. and Davis, R. A., 2000, Introduction to time series and forecasting (New York: Springer).

Campbell, N. A., Simon, E. J. and Reece, J. B., 2004, Essential biology with physiology (San Francisco, CA: Pearson/Benjamin Cummings).

Chatterjee, S., Hadi, A. S. and Price, B., 2000, Regression analysis by example (New York: Wiley).

Conway, E. D., 1997, An Introduction to Satellite Image Interpretation, The Johns Hopkins University Press, MD, 242.

Crippen, R. E., I 1990, Calculating the Vegetation Index Faster, Remote Sensing of Environment, 34, 71-73.

Curran, P. J., 1984, Principles of remote sensing (London ; New York: Longman).

Collins, W., 1978, Remote sensing of crop type and maturity. Photogrammetric Engineering and Remote Sensing, 44, pp. 43-55.

Dhingra N, Dhillon GPS, Lal S., 1998, Process Indicators for Malaria Control, Journal of Communicable Diseases, 30:209-228.

Didan, K. and Huete, A., 2004, Analysis of the global vegetation dynamic metrics using MODIS Vegetation Index and land cover products. Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. IEEE International.

Domenikiotis, C., Spiliotopoulos, M., Tsiros, V. & Dalezios, N. R., 2004, Early cotton yield assessment by the use of the NOAA AVHRR derived Vegetation Condition Index (VCI) in Greece. International Journal of Remote Sensing, 25(14), 2807-2819.

Dreaper, N.R. And Smith, H., 1981, Applied Regression Analysis (New York: Wiley).

Dutt, A. K., Akhtar, R, and Dutta, H.M.,1980, Malaria in India with particular Reference to two West Central States, Social Science and Medicine, Vol.14D, 320;

Elias M. and Rahman M, 1987, The ecology of malaria carrying mosquito Anopheles Philippinensis Ludlow and its relation to malaria in Bangladesh. Medical Research Council Bulletin, Bangladesh 13: 15-28.

Emmanuel M., Stephen M., Noboru M., Li Li, Chen-chieh Feng , Ling, B. , Uriel K., Cindy S., Louisa B., Guofa Z., Andrew K. Githeko and Guiyun Yan, 2006, Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands, Malaria Journal, 5 :13

Faiz Ma., Yunus Eb, Rahman Mr, Hossain Ma, Pang Lw, Rahman Me And Bhuiya S.N, 2001, Failure of national guidelines to diagnose uncomplicated malaria in Bangladesh, American Journal of Tropical Medicine and Hygiene 67, 396 -399

Freund, Rudolf J. and Ramon C. Littell, 2000, SAS System for Regression (Third edition). Cary, NC: SAS Institute.

Gates, D. M., Keegan. H. I., Schleter, J. C. and Weidner, V. R., 1965, Spectral properties of plants, Applied Optics, 4, 11-20.

Ghose, G., Rahul Ray Maj K, Banerjee Lt Col A, 2006 , Epidemiological Investigation of Forest Malaria among GREF and Army Personnel, Medical Journal Armed Forces India, 62:1

Gutman, G., 1999, On the use of long-term global data of land reflectance and vegetation indices derived from the Advanced Very High Resolution Radiometer. Journal of Geophysical Research-Atmospheres, 104: 6241-6255.

Gutman, G.G., 1991, Vegetation indices from AVHRR data: An update and future prospects, Remote Sens. Environ.35, 121-136.

Gunst, R.F. and Mason, R.L., 1980, Regression Analysis and its Application: a Data oriented Approach (New York: m. Dekker).

Greena, R.M., and S.I. Hay, 2002, The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications. Remote Sensing of Environment, 79, 166–175.

Hayas, M. J. and Decker, W. L.,1996, Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. International Journal of Remote Sensing, 17, 3189-3200.

Ingrid, V.F. Broek, V.D., Wardt, S.V.D, Talukder, L. Chakama, S, Brockman,A. Nair, S. and Tim. C, 2004, Drug resistance in Plasmodium falciparum from the Chittagong hill Tracts, Bangladesh, Tropical Medicine and Internatinal Health, 9: 680-687

Jensen, J. R., 2000, Remote Sensing of the Environment: an Earth Resource Perspective (New Jersey: Prentice Hall).

Jean G. , Ousmane T., Nadine D. , A lassane Dicko, Stéphane Ranque , Loic Forest, Jacques Demongeot and Ogobara K Doumbo, 2009, Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali, Malaria Journal**,** 8 : 61

Jones, P. D., Jouzel, J. and Bradley, R. S., 1996, Climatic Variations: oring Mechanisms of the last 2000 years (Berlin ; London: Springer).

Joseph S. D'Aleo , 2002, The ORYX resource guide to El Nino and La Nina, Oryx Press, Westport, Connecticut

Jiang, Z., Huete, A. R., Chen, J., Chen, Y., Li, J., Yan, G. & Zhang, X., 2006, Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. Remote Sensing of Environment, 101(3), 366-378.

Kilian AHD., Langi P, Talisuna A, Kabagambe G (1999) Rainfall pattern, El Nino and malaria in Uganda. Transactions for the Royal Society of Tropical Medicine and Hygiene, 93: 22-23.

Kondrachine A. V., 1992, Malaria in WHO Southeast Asia Region; Indian J.Malariol.29 129 .160

Kidwell, K. B., 1997, Global Vegetation Index user's guide (Camp Springs MD: US Department of Commerce, NOAA, National Environmental Satellite Data and Information Service, National Climatic Data Center, Satellite Data Services Division).

Kochar, D. K., Sirohi, P., Kochar, S.K.,. Budania, M.P and Lakhotia, J.P., 2007, Dynamics of malaria in Bikaner, Rajasthan, India (1975–2006), J Vector Borne Diseases 44, December, 281–284

Kogan, F. and Sullivan, J., 1993, Development of global drought-watch system using NOAA/AVHRR data. Advancements in space Research 13(5):219-222.

Kogan, F.N.,1995, Droughts of the late 1980s in the United State as derived from NOAA polar orbiting satellite data, Bulletin of the American Meteorological Society ,76:655-668.

Kogan, F.N., 1997, Global drought watch from space, Bulletin of the American Meteorological Society,78, 621-636.

Kogan, F.N.,2000, Global drought detection and impact assessment from space . In Drought: A Global Assessment (D.A. Wilhite,Ed) V0l,I, Hazard and Disaster Series, Routledge, London and New York, 196-210.

Kogan, F. N. and Zhu, X., 2001, Evolution of long-term errors in NDVI time series:1985-1999. Advances in Space Research, 28(1), 149-153.

Kovats, SR., 2000, El Nino and health, Bull WHO 78, 1127-1134

Kaufmann, R. K. and Zhou, L., 2000, Effect of orbital drift and sensor changes on the time series of AVHRR vegetation index data. IEEE Transactions on Geoscience and Remote Sensing, 38(6), 2584.

Kriegler, F. J., Malila, W. A., Nalepka, R. F. & Richardson, 1969, Preprocessing transformations and their effects on multispee: recognition. Sixth International Symposium on Remote Sensing of Environment. University of Michigan, Ann Arbor, MI

Kumar A,Valecha N,Jain T and Dash A P 2007 Burden of Malaria in India:Retrospective and Prospective View; American Journal of Tropical Medicine and Hygiene 77 69 -78

Los, S.O. Justice, C.O. and Tucker, G.J.,1994 , A global 1 by 1 degree NDVI data set for climate studies derived from GIMMS continental NDVI data. International Journal of Remote Sensing.15: 3493-3518.

Liang, S., 2004, Quantitative remote sensing of land surfaces (Hoboken, NJ: Wiley).

Liu, W. T. & Kogan, F., 2002, Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. International Journal of Remote Sensing, 23(6), 1161-1179.

Lillesand, T. M., Kiefer, R. W., 1994, Remote Sensing and Image Interpretation, 3rd edition. Wiley. New York

Mouchet, J., Manguin, S., Sircoulon, J., Laventure, S., Faye, O., Onapa, A.W., Carnevale, P. Jul vez, J. Fontenille, D. 1998. Evolution of malaria in Africa for the past 40 years: Impact of climatic and human factors. Journal of the American Mosquito Control Association. 14, 121-130.

Myneni, R. B., Tucker, C. J. Asrar, G. & Keeling, C. D., 1998, Interannual Variations In Satellite-Sensed Vegetation Index Data From 1981 To 1991. Journal of Geophysical Research, 103, 6145-6160.

Nanda, N., Yadav, R.S., Subbarao, Sarala K., Joshi, Hema and Sharma, V.P.,2000, Studies on Anopheles fluviatilis and Anopheles culicifacies in relation with malaria in forest and deforested riverine ecosystems in northern Orissa, India. J Am Mosq Control Assoc 16(3), 199

Nagpal B and Sharma V, Indian Anophelines, New Delhi, 1995, 416-423.

Oaks SC Jr., Mitchell VS, Pearson GW, and Carpenter CCJ, Eds., 1991, Malaria: Obstacles and Opportunities. A report of the Committee for the Study on Malaria Prevention and Control: Status Review and Alternative Strategies. Division of International Health, Institute of Medicine. Washington, DC: National Academy Press.

Operational Manual for Implementation of Malaria Programme 2009, Ministry of Health and Family Welfare Government of India New Delhi

Patz JA, Strzepek K, Lele S, Hedden M, Greene S, Noden B, Hay SI, Kalkstein L, and Beier JC,1998, Predicting key malaria transmission factors, biting and entomological inoculation rates, using modeled soil moisture in Kenya. Tropical Medicine and International Health, 3, 818-827.

Pampana E, 1969, A Text Book of Malaria Eradication, Oxford University Press, London, Uk, 17-63.

Pietro C., Tewolde G., Malanding J., Patricia M. G., Marc L., Shashu G., Andom O., Anthony G. B., Michael B., John del C., Stephen J. Connor, Issac Fesseha, Eugene P. Brantly, and Madeleine C. Thomson, 2007**,** Malaria Stratification, Climate, and Epidemic Early Warning in Eritrea, American Journal of Tropical Medicine and Hygiene ,77(6_Suppl), 61-68

Price, J. C., 1988, An update on visible and near infrared calibration of satellite instruments. Remote Sensing of Environment, 24(3), 419-422.

Privette, J. L., Fowler, C., Wick, G. A., Baldwin, D. & Emery, W. J.,1995, Effects of orbital drift on advanced very high resolution radiometer products: Normalized difference vegetation index and sea surface temperature. Remote Sensing of Environment, 53(3), 164-171.

Poveda, G., Roja, W., Quinones, M.L., Velez, I.D. , Mantilla, R.I., Ruiz, D. Zuluaga, J.S. and Rua, G.L. 2001. Coupling between annual and ENSO timescales in the malaria-climate Association in Colombia. Environmental Health Perspectives. 109,489-493.

Póvoa, M.M., Wirtz, R.A, Lacerda, R.N.L., Miles, M.A., Warhurst, D., (2001), Malaria Vectors in the Municipality of Serra do Navio, State of Amapá, Amazon Region, Brazil, 96(2),179-184,

Prata A. J., Caselles V., Coll C., Sobrino J. A.,Ottlé C.,1995, Thermal remote sensing of land surface temperature from satellites: Current status and future prospects , International Journal of Remote Sensing, 12, 175 - 224

Rao, C. R. N. and Chen, J., 1995, Inter satellite calibration linkage for the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-7,-9,-11 spacecraft. International Journal of Remote Sensing, 16, 1931-1942.

Rao, C. R. N. and Chen, J., 1999, Revised post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft, International Journal of Remote Sensing, 20, 3485- 3491.

Rao, T. R, 1984, The Anophelines of India Delhi:Malaria Research Centre

Ramesh, C. D., Sharmila P. and Aditya P. Dash, 2008, Climate change and malaria in India: Interplay between temperatures and mosquitoes, Regional Health Forum –- 12:1

Rosenberg, R. and Maheswary, N., 1982, Forest Malaria in Bangladesh, I.Parasitology, American Journal of Tropical Medicine and Hygiene, 31**:** 175 –182

Rosenberg, R. and Maheswary N, 1982, Forest Malaria in Bangladesh, II.Transmission**,** American Journal of Tropical Medicine and Hygiene , 31: 183–191.

Russel, Pf. West, L.S, Manwell, R.D, and Macdonald, G. 1963, Practical Malariology. Oxford University Press, London, 750-757.

Rahman, A., Kogan F, Roytman, L., 2006 Analysis of malaria cases in Bangladesh with remote sensing data. American Journal of Tropical Medicine and Hygiene 74 (1), 17–19

Raymond H. Myers, 1986, Classical and Modern Regression with Applications, Dexbury Press, Boston Massachusetts

Ryan, T. P., 1997, Modern Regression Method, John Willey & Sons, Inc.

Sinha A.K., 2005, Malaria, APH publishing corporation, New Delhi

Simoniello, T., Cuomo, V., Lanfredi, M., Lasaponara, R. & Macchiato, M., 2004, On the relevance of accurate conection and validation procedures in the analysis of AVHRR-NDVI time series for long- term monitoring. Journal of Geophysical Research Atmospheres, 109, 20101-20113

Sharma, VP., 1996, Re-emergence of malaria in India. Indian Journal of Medical Research, 103:26-45.

Sachs, J. and Malaney, P., 2002, The economic and social burden of malaria, Nature, 415, 680-685.

Simon, I. H., Robert W. S., Devid, J.R., 1998, Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data, Transactions of the Royal Society of Tropical Medicine and Hygiene, 92(1),12-20.

Srivastava, A., Nagpal, B.N., Sexena, R. and Subbarao, S.K., 2001, Predictive habitat modelling for forest malaria vector species *An. Dirus* in India – A GIS-based approach, Current Science,80(9), 1129-1134.

Srivastava, Hc And Yadav, Rs, 2000, Malaria Outbreak in a Tribal Area of Gujarat State, India, Southeast Asian Journal of Tropical Medicine Public Health, 31(2) 219-224.

Tucker, C. J., Vanpraet, C., Boerwinkel, E. and Gaston, A., 1983, Satellite remote sensing of total dry matter production in the Sencgalese Sahel. Remote Sensing of Environment, 13(6), 461-474.

Tucker, C. J., Vanpraet, C. L., Sharman, M. J. & Van 1ttersum, G., 1985, Satellite remote sensing of total herbaceous biomass production in the senegalese sahel: 1980-1984. Remote Sensing of Environment, 17(3), 233-249.

Teklehaimanot, H.D., Lipsitch, M., Teklehaimanot, A. and Schwartz, J. 2004a, Weather-based predictions of Plasmodium falciparum malaria in epidemic- prone regions of Ethiopia II. Weather-based prediction systems perform comparably to early detection systems in identifying times for interventions, Malaria Journal, 3:44.

Teklehaimanot, H.D., Lipsitch, M., Teklehaimanot, A. and Schwartz, J. 2004b. Weather-based predictions of Plasmodium falciparum malaria in epidemic- prone regions of Ethiopia I. Patterns of lagged weather effects reflect biological mechanisms, Malaria Journal. 3:41.

Thomson, M. C., Doblas-Reyes, F. J., Mason S. J., Hagedorn, R. S., Connor, J., Phindela, T., Morse, A. P., and Palmer, T. N., 2006, Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature, 439, 576-579.

Thomson, M. C., Connor SJ, Alessandro UD, Rowlingson B, Diggle P, Cresswell M, Greenwood BM, 1996, A spatial model of malaria risk in the Gambia: predicting the impact of insecticide treated and untreated bed nets on malaria

infection. Proceedings of the Spatial information Research Centre's 10th Colloquim.

Thomson, M., Connor, S., Neill, K. O and Meert, J. P., 2000, Environmental Information for Prediction of Epidemics, Parasitology Today, 16(4), 137-138.

Takahashi K and Arakawa H., 1981, Climates of Southern and Western Asia. Amsterdam: Elsevier Scientific Publishing Company.

Willem, T., , Paulo de Tarso R. Vilarinhos , Petra Schneider and Fatima dos Santos , 2005, Effects of environmental change on malaria in the Amazon region of Brazil, Environmental Change and Malaria Risk: Global and Local Implications, Wageningen UR Frontis, 2005, XXII p.139

Xiang, G., Huete, A. R. and Didan, K., 2003, Multisensor comparisons and validation of MODIS vegetation indices at the semiarid Jomada experimental range. IEEE, Transactions on Geoscience and Remote Sensing, 41(10),2368-2381.

Yang, H.M. and Ferreira, M.U., 2000, Assessing the effects of global warming and local social and economic conditions on the malaria transmission. Journal of Public Health 34:21-222.

World Health Organization, 1999, The World Health Report, 1999: Making a Difference. WHO, Geneva.

World Health Organization, Malaria epidemics: forecasting, prevention, early detection and control – from policy to practice.

World Health Organization: 2003, WHO: Africa Malaria Report. Geneva: World Health Organization

World Malaria Report, 2009, http://www.who.int/malaria /world_malaria_report_2009 /en/index.html

Publications and Presentations

Atiqur Rahman, Leonid Roytman, Nir Y. Krakauer, **Mohammad Nizamuddin** and Mitch Goldberg, "Use of Vegetation Health Data for Estimation of Aus Rice Yield in Bangladesh" Sensors 2009, 9(4), 2968-2975;

Mohammad Nizamuddin, Leonid Roytman, and Felix Kogan, 2009, "Satellite-observed sensitivity of weather condition for predicting malaria vector distribution in Bandarban district, Bangladesh." Proceedings of SPIE, the International Society for Optical Engineering, vol. 7312, [73120T.1-73120T.7] (Presented)

Atiqur Rahman, **Mohammad Nizamuddin**, Leonid Roytman, and Ubydul Haque "Application of AVHHR-based vegetation health indices for malaria vector assessment of Chittagong division, Bangladesh", Proceedings of the International Conference on Statistical Sciences (ICSS, 2008), ISBN 984-300-002873, pp. 334-340 (Presented)

Mohammad Nizamuddin, Leonid Roytman, and Felix Kogan, "Early prediction of malaria in forest hills of Bangladesh using AVHRR based satellite data" American Meteorological Society (AMS) 89TH annual meeting, Phoenix, Arizona 2009 (Presented)