SPATIAL DISTRIBUTION OF MALARIA CASES IN THE FORMER KILIFI DISTRICT 2005 TO 2007

By

EDWARD ABWAO.

"ROJECT SUBMITTED TO THE INSTITUTE OF TROPICAL AND INFECTIOUS DISEASE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN MEDICAL STATISTICS.
(APPLIED STATISTICS)

UNITID
COLLEGE OF HEALTH SCIENCES
UNIVERSITY OF NAIROBI

Declaration

I declare that all assessed work submitted for my degree is the results of my own work. In all other cases, material from the work of others is acknowledged and quotations and paraphrases are suitably indicated.

EDWARD ABWAO

W62/P/8110/2006

Signature.....Date...24.11.2009

This project has been submitted for examination with my approval as the supervisor.

Dr. THOMAS ACHIA

University of Nairobi

School of Mathematics

P.O Box 30197-00100, Nairobi, Kenya

Signature.....Date...24/11/2009

Dedications

This work is dedicated to all members of the Abwao family and especially my mum and my late auncle Joshua.

Acknowledgment

I would like to acknowledge the assistance given to me by my supervisor Dr. Thomas Achia, for his time and patience in seeing me through this study. I would also like to appreciate the time put in by Mr. Lawrence Muthami who took time off his busy schedule to go through the work and gave me his guidance. I would also like to appreciate the efforts of Dr. Mwalili of JKUAT for his very valuable assistance that went along way in completing this study. The assistance, encouragement and friendship of my two classmates, Maina Waweru and Dr. Okumu J. William is also greatly appreciated. They have really been of help. The overall assistance from the course coordinator Dr. Rosemary Nguti and UNITID who ensured that this study was concluded is also appreciated. Above all, I am so grateful to the LORD God, through whose love and grace, has made me the person I am today, and has given me hope for a better tomorrow.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

- AIC- Akaikes Information Criterion
- CSR-Complete Spatial Randomness
- Cumulative Incidence Proportion- This will be referred to as proportion unless there is any ambiguity.
- EDA-Exploratory Data Analysis
- GIS-Geographical Information Systems
- HMIS- Health Management Information System
- IDSR- Integrated Disease Surveillance and Response
- RBM- Roll Back Malaria initiative
- RS- Remote sensing
- VHC- Village Health Committee
- WHO- World Health Organization

Abstract

Background: Malaria prevalence varies between regions and even within the same region; there are variations depending on various Environmental determinants of malaria transmission. Kilifi district is located in the coast province of Kenya which is classified as a malaria endemic zone.The ability to identify areas with high spatial clustering of malaria is of great significance as it will enable targeted intervention to be employed in these areas to fight malaria. Spatial data analysis help identify areas with spatial clusters and also enable one identify trends that may be existing in the wider region. We try to see if there is any difference in malaria cases reported at the different types of health facilities. The study also tries to see if there is any difference in malaria prevalance in the population of under five years and those over five years. This study tries to find out the various sub regions where malaria is more clustered than others in district, based on the number of patients seen and treated of malaria in Kilifi district for the period 2005 to 2008. This study also tries to identify any spatial trends that may be present.

Methods: In this study, we examine the local distribution pattern of malaria using data collected between 2005 and 2007 from all the different types of health facilities located in the divisions within the district. These facilities were point-referenced. Continous spatial data analysis was then used to analyse the data.

Results: Malaria cases varies within the district and this can be seen from the different numbers in the different facilities. Analysis showed a significant association of malaria risk with elevation, humidity, rainfall. The resulting map broadly agreed with expert opinion about the variation of

risk in the country, and further showed marked variation even at local level. High risk areas were in the low-lying ocean shore regions, with the risk reducing as one moves away from the shores towards the high altitude region. Children under five years were at a much higher risk of contracting malaria as compared to those over five years. There was also no difference in the number of patients who had malaria when analysed depending on the type of health facility they visited.

Conclusion: The map provided description of the geographic variation of malaria incidence in Kilifi District, and might help in the choice and design of area specific interventions, which is crucial for reducing the burden of malaria in in Kilifi District and may be applied to the whole country.

Chapter 1

Introduction

Malaria remains a leading cause of morbidity and mortality in Kenya, especially in young children and pregnant women. It accounts for 33.5 percent of outpatient attendances and 19 percent of admissions to health facilities. Malaria is the most important cause of death in children under 5 years of age and is estimated to cause 20 percent of all deaths in this age group. In spite of this situation, malaria is a preventable and curable disease (MoH Kenya 2006 National Guidelines for Diagnosis, Treatment and Prevention of Malaria for Health workers in Kenya).

Malaria burden and transmission patterns vary across the country. Understanding how malaria varies in the community as a result of seasonal or year-to-year changes in climate and environmental factors is important for the planning of national malaria control programs, as it will allow interventions to be adapted to specific sites or times of the year. This is essential for the effective control of the disease.

The aim of this project is to study the spatial malaria patterns in the district of Kilifi based on the number of patients tested and treated of malaria in all government health centres and also study the relationship between the malaria case burden and other diseases in the health facilities.

The analysis of spatial point patterns came to prominence in geography during the late 1950s

and early 1960s, when a spatial analysis paradigm began to take firm hold within the discipline. Researchers borrowed freely from the plant ecology literature, adopting techniques that had been used there in the description of spatial patterns and applying them in other contexts: for example, in studies of settlement distributions (Dacey 1962; King 1962), the spatial arrangement of stores within urban areas (Rogers 1965) and the distribution of drumlins in glaciated areas (Trenhaile 1971). The methods that were used could be classified into two broad types (Haggett et al. 1977). The first were distance-based techniques, using information on the spacing of the points to characterize pattern (typically, mean distance to the nearest neighbouring point). Other techniques were area-based, relying on various characteristics of the frequency distribution of the observed numbers of points in regularly defined sub-regions of the study area (quadrats).

The geographic information system (GIS) and global positioning system (GPS) have been widely applied to health and epidemiology for malaria research and control in most sub-Saharan Africa (White 1972). Spatial point pattern analysis may help to identify high-risk diseases areas, sources of diseases, and high-risk populations (Craig MH 1999). These statistical techniques are based on case events and count data, where known geographic locations (x-y coordinates) of disease cases are commonly represented as points (PAHO 2003). The disease mapping could also play an important role in formulating malaria control activities, evaluating changes in malaria transmission over time and allocating resources to control malaria (GPY Clark 2000, Snow RW 1993) especially in high or persistent local malaria transmission areas (hot spots)(Gill CA 1921).

1.1 Study Area-District Profile

Kilifi District is one of the seven districts in Coast Province of the Republic of Kenya. Kilifi Town is its headquarters, situated 540 kilometres away from Nairobi City. It borders with Malindi to the North West, Taita Taveta to the West and Mombasa and Kwale to the South. It is divided into seven administrative divisions namely; Kaloleni, Bahari, Choni, Kikambala, Ganze, Vitengeni, and Bamba. Kilifi has 36 locations and 108 sub locations. It covers an area of 4779.2sq.km. The land rises gradually from the sea level in the East to an altitude of about 2700 ft in the South Western side.

This district is divided into 3 constituencies namely: Kaloleni (Kaloleni division), Ganze (Ganze, Vitengeni and Bamba divisions) and Bahari (Bahari, Kikambala and Chonyi divisions).

Table 1.1: Climatic profile

Location	Latitudes 3.21 S and 4.28 S, Longitudes 38.43 E and 39.95 E.
Climate	Mainly tropical, with average annual rainfall of 1100 mm along the coastal belt and 400 mm in the hinterland. The major rain season is between March and May.
Rainfall	High - 1100mm Low - 400mm
Temperature Range	High: $26.5^0c - 34^0c$ Low: $22.5^0c - 24.5^0c$ Average: 30^0c
Average relative humidity	60 percent

Table 1.2: **Demography and Population Profile** (2007 figures)

Total Population	665,647
Total No. of Males	322,246
Total No. Females	343,601
Sex Ratio (Female/Male)	111:100
No. of Households	90,311
Household size	6.17

Figure 1.1:

Figure 1.2:

1.2 Objectives of the study

1.2.1 Main Objective

The main objective of the study is to determine the spatial distribution pattern of malaria cases in Kilifi district between the years 2005 and 2008.

1.2.2 Specific objectives

The specific objectives of this study are:

1. To establish the spatial clustering pattern of clinical malaria based on the reported number of new patients diagnosed and treated of malaria in all the government health facilities in the years 2005 to 2008.
2. To describe the overall malaria case burden in the health facilities in Kilifi district between the years 2005 and 2008.
3. To compare the malaria burden between the population under five and the over five for the period 2005-2008.

Expected Outcomes

The expected outcome of this study is the malaria distribution maps representing the cumulative incidence of malaria in Kilifi district for the patients over five years and those who are under five years.

Significance of the study

The study will establish the cumulative incidence of malaria for the different regions of the district. It will be possible to establish malaria hot spots and any spatial relationship in malaria cases between sub regions that neighbour each other. Data will be obtained for timely and targeted interventions for control and management of malaria that will be applied for each specific region and this will be helpfully in the fight against malaria as envisioned in the millennium development goals.

Chapter 2

Literature Review

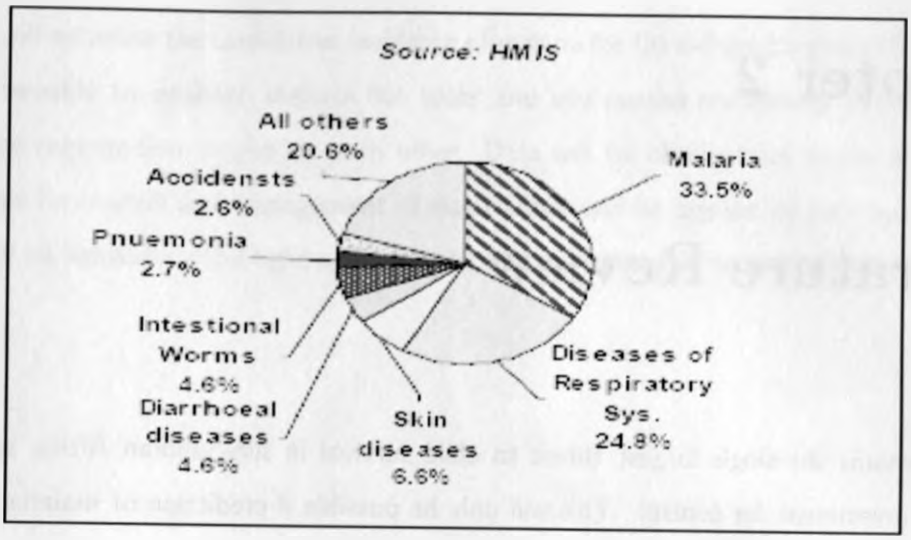
Malaria remains the single largest threat to child survival in sub-Saharan Africa and warrants long-term investment for control. This will only be possible if prediction of malaria occurrence, regional variation and the people at risk could be could be made.

Malaria has been prevalent in most parts of the tropical and subtropical world, including Africa, Asia, and Central and South America, for a long time. The disease has been the single most important cause of morbidity in some of these areas. With resources becoming scarce, disease control in the future is likely, therefore, to benefit from concerted efforts, including those based on an understanding of the micro epidemiology of the disease in a given situation.

In Kenya, malaria accounts for 33.5 percent of outpatient morbidity.

The investigation of infectious disease clustering is receiving renewed interest, due to advances in geographical information systems (GIS) and spatial statistics, which allow for the quantification of the degree of clustering of infections. Such approaches have been used to investigate the spatial clustering of dengue (Morrison et al. 1998), LaCrosse encephalitis (Kitron et al. 1997) and sleeping sickness (Fe'vre et al. 2001), but their application to malaria has been limited (Schellenberg et al. 1998; Chadee and Kitron 1999; Ghebreyesus et al. 2003).

9.3 Leading Causes of outpatient Morbidity



Good maps of malaria risk have long been recognized as an important tool for malaria control. The production of such maps relies on modeling to predict the risk for most of the map, with actual observations of malaria prevalence usually only known at a limited number of specific locations. Estimation is complicated by the fact that there is often local variation of risk that cannot be accounted for by the known covariates and because data points of measured malaria prevalence are not evenly or randomly spread across the area to be mapped (I Kleinnschmidt 2000). With the changing environmental and climatic conditions, malaria prevalence is shifting from the traditional areas to new area that did not previously experience malaria endemics.

Malaria is an environmental disease. Anopheles mosquitoes transmit the causative agent, Plasmodium spp., when the environmental parameters (such as water availability, temperature, and humidity) permit. For example, in many parts of the world where temperature is right, malaria transmission is highly seasonal, with peak transmission following the period of peak rainfall (Madeline C Thomson 1999)

Malaria not only poses a risk to survival but also the repeated clinical consequences of infection during early life place a burden on households, health services and ultimately the economic

development of nations (Bloom and Sachs, 1998). Sachs and Warner (1997) have argued that the persistence of endemic malaria in the tropics, and particularly in Africa, is contributory to a perpetual state of depressed economic growth in these regions. These macro-estimates of burden and economic associations provide clear support for a renewed effort aimed at halving malaria mortality by the year 2010, referred to as the Roll Back Malaria (RBM) initiative (Nabarro and Tayler, 1998; WHO, 1998). This optimistic goal has been conceived at a time when existing, affordable therapeutics are rapidly failing, health service provision is breaking down, vaccines seem to be a pipe dream, and poverty, conflict and corruption continue to afflict many African states (Desowitz, 1999).

The resources targeted at malaria control will always be limited when compared with other social sector investment. Furthermore, not all malaria interventions will be equally appropriate for every setting. The challenge for the public-health sector is to decide which interventions would be appropriate where and how these may be tailored to the local epidemiology to achieve maximal health impact for minimal investment. Clearly not all resource allocation is evidence-based but there is an increasing recognition among the malaria research and control communities that mapping risk and the projected benefits of intervention is a fundamental monitoring and decision-informing tool. It is in this context the utility of remote sensing (RS) and geographical information systems (GIS) are evaluated.

An understanding of regional patterns of disease-specific mortality and disability increasingly drives approaches to global public health. Through the integration of high-resolution population and climate probability models of P. falciparum transmission, geographical information systems have been used to define the spatial limits of populations exposed to the risk of malaria infection in Africa.

In Kenya, malaria transmission is acutely seasonal with peaks occurring 2-3 months after the peak rains in April-May, although the extent of the malaria burden varies considerably from year to year (Simon Brooker, Benson Estambale 2004).

The interaction between temperature and rainfall is largely responsible for the seasonal characteristic of malaria transmission. Seasonal variation of infection risk is a common feature of malaria in Sub Saharan Africa and is reflected in intra annual changes in vector densities, entomological inoculation rates and malaria admissions (Christie, 1959; Julvez et al., 1992; Aniedu, 1997; Hay et al., 1998b, 2000).

Understanding how malaria varies in the community as a result of seasonal or year-to-year changes in climate and environmental factors is important for the planning of national malaria control programs since it may allow interventions to be adapted to specific sites or times of the year. This is essential for the effective control of the disease. Assessing the relationship between environmental parameters and malariometric indices in a quantitative manner is fraught with difficulties since the prevalence of malaria may vary considerably within a small area, and the data collected at a limited number of points are not necessarily applicable to a broader region (Bjorkman A 1985). Furthermore, prevalence data collected during a limited period of time cannot describe the seasonal variations that occur even in areas of high endemicity (Snow RW 1993)

Information on population distribution, health services, disease risk and seasonality should remove some of the barriers to providing a credible platform upon which to institute selected or targeted malaria control and prevention. GIS and RS provide a framework to develop high-resolution maps of risk, population and service $delivery$[10].

2.1 The Life Cycle

Mosquitoes of the genus Anopheles were first identified as the vectors of human malaria in 1897 by Sir Ronald Ross (Wernsdorfer and McGregor, 1988). Today, four species of Plasmodium are known to infect humans namely P. falciparum, P. malariae, P. vivax and P. ovale. P. falciparum is the dominant malaria parasite found in the stable endemic areas of Africa (Young, 1976;

Gilles, 1993). The malaria parasite develops in two stages; a sexual cycle that takes place within the mosquito vector and an asexual cycle in the human host (Fujioka and Aikawa, 1999). The haematophagous adult female Anopheles seek vertebrate hosts soon after emergence. The ingested blood is used to support egg production and, following development and subsequent oviposition, the female vector seeks further blood meals to nourish future broods. It is this repeated feeding that facilitates the transmission of parasites between hosts. Infection of the human host begins when sporozoites from an infected mosquito are injected into the blood of a susceptible human during a blood meal. It then takes 0.5-4 hours for the sporozoites to invade host liver cells where they multiply and release as many as 30 000 merozoites, which, in turn, invade red blood cells. This asymptomatic period usually lasts about a week in tropical countries. The erythrocytic asexual development stage follows when the parasite develops from a ring form to a trophozoite that then becomes a schizont, which multiplies to produce 4-32 merozoites. This intracellular multiplication causes red blood cells to rupture, with the resultant release of toxins into the blood, occurring in synchronized 48 hours cycles for P. falciparum, P. vivax and P. ovale and in 72 hours cycles for P. malariae. The bouts of fever associated with malaria correspond with these episodes of toxin release. Continued asexual multiplication with the invasion of further erythrocytes, as well as sexual differentiation, results in the production of macrogametocytes (female) and microgametocytes (male). These are the forms of the parasite infective to the mosquito. The parasite's sexual cycle begins when gametocytes are ingested by a mosquito vector feeding on an infected individual (Beier, 1998). Fertilization of the gametocytes to form ookinetes takes place in the midgut of the mosquito and these lodges in the midgut outer wall as oocysts. Numerous sporozoites develop within the oocysts and, as the oocysts rupture, migrate to the mosquito's salivary gland from where they are injected into the human host during subsequent blood meals. Various aspects of this complex life cycle are affected by climate and are explored below.

2.2 Environmental Determinants of Malaria Transmission

Temperature

Malaria is a disease of tropical and temperate countries between the latitudinal limits of 64^0 North and 57^0 South (Gill, 1921) with prevalence increasing towards the equator. As the parasites require time to develop into infective stages, female anopheles are not immediately infective after feeding. This extrinsic incubation period is temperature dependent and optimum conditions have been defined between 25^0C and 30^0C, with development ceasing below 16^0C and above 40^0C (Russell et al., 1946; Gilles, 1993).

Rainfall

Rainfall provides surface water in which female anopheles can lay eggs. In arid areas where temperatures are usually suitable, malaria transmission occurs only when rainfall provides temporary breeding habitat for vectors. These areas are often classified as malarious near water since transmission outside the rainy seasons typically occurs only along riverbeds, oases and other man-made surface water sites. Studies have demonstrated an association between abundance of An. gambiae s.l. and rainfall (Christie, 1959; White et al., 1972; Molineaux and Gramiccia, 1980; Charlwood et al.,1995). Rainfall effects are often most apparent during epidemics when the rise in malaria cases is often proportional to the amount of precipitation, among other factors (Christophers, 1911; Covell, 1957; Wernsdorfer and McGregor, 1988; Malakooti et al., 1998; Kilian et al., 1999).

Climate Seasonality

The interaction between temperature and rainfall is largely responsible for the seasonal characteristic of malaria transmission. Seasonal variation of infection risk is a common feature of malaria in Sub Saharan Africa and is reflected in intraannual changes in vector densities, entomological

inoculation rates and malaria admissions (Christie, 1959; Julvez et al., 1992; Aniedu, 1997; Hay et al., 1998b, 2000).

Atmospheric Moisture

Early documented reports of human malaria describe its association with humid swamps and marshes (Gill, 1921) and several authors have attempted to define optimum conditions of relative humidity (RH) based on such observations (Wernsdorfer and McGregor, 1988; Gilles, 1993). Gill (1921) was the first formally to investigate the effect of changes in RH on Culex fatigans and the transmission of avian malaria. Higher relative humidity was associated with increased vector longevity and greater frequency of feeding. RH also determined the timing and duration of daytime resting behaviour (Boyd, 1930; Russell et al., 1946; Molineaux, 1988).

Altitude

Altitude has long been a subject of interest among malariologists (Schwetz, 1942; Garnham, 1948; Heisch and Harper, 1949; Covell, 1957; Roberts, 1964; Malakooti et al., 1998). Altitude and temperature are explicitly linked, with every 100 m increase in height corresponding to an approximately 0.5^0C decline in temperature. The use of altitude can be confusing, however, with the limit for malaria transmission variously reported above 2000 m in Ethiopia (Covell, 1957), 1800 m in the Congo (Schwetz, 1942) and 1950 m in Kenya (Garnham, 1948). Use of the phrase highland malaria continues, but it is more clearly thought of as temperature-limited unstable transmission.

2.2.1 Research question

Is there significant difference in the cumulative incidence proportion of the malaria spread in Kilifi District?

Chapter 3

Methodology

The ministry of health has introduced Integrated Disease Surveillance and Response (IDSR) system whereby the Districts are the focus for integrating disease surveillance functions. The districts collect and collate data on a weekly and monthly basis, the number of patients treated of all the diseases in all the health facilities within the district. This data shows; the gender of the patients, whether over five years or under five years, total number of outpatients and in patients. The data also has total number of bed days for the in patients, total number of discharges and deaths from these facilities. At the facilities we got the catchment population per facility. This was obtained from the Village Health Committees (VHC) that are involved in health matters in each facility's catchment area. The VHCs are responsible for a specific village and know the total number of people in these villages. For each division, we get population by adding the total number of people in the villages that form the divion. This is the data that will be analysed.

Kilifi district was chosen as it has a new and better HMIS system that is under trial before it is fully implemented in the whole country. Over time, this system has been used to collect data on all the number of patients seen and treated of all the major conditions in the district.

Using WHO Health mapper programme, we are able to get the positioning of all the health facilities in the district together with the administrative boundaries (Provincial, District, Division,

Location and sub locations). These are then used to study the spatial distribution of malaria in the district.

A distinction is often made between prevalence and incidence of a disease. Prevalence is a measure of the total number of cases of disease in a population at a point in time, while incidence rate is the occurrence of new cases of disease (incident number) in a population divided by the person-time over a specified period. Thus, prevalence indicates the magnitude of disease burden whereas incidence conveys information about the risk of contracting malaria. In the present study, the measurement of incidence is complicated by changes in the population at risk, since sometimes the same person may report more than once during a month to the health facility. Each episode of malaria roughly lasts for a week and utmost for one month in the presence of recrudescence. In these circumstances, the definition of incidence is usually restricted to the first event reported in that month. Once a person is classified as a malaria case, he or she is no longer liable to become a new case within the same month. Beyond one month, the person reporting and testing positive at a clinic is considered a new case. Therefore, the incidence density (ID) of malaria was calculated by relating the numbers of new cases to the person years at risk, calculated by adding together the periods during which each individual member of the population is at risk during the measurement period. ID is defined as: Number of new cases/Total person years at risk.

The spatial analysis is composed of a set of chained procedures whose aim is to choose an inferential model that explicitly establishes the spatial relationship present in the phenomenon. The initial procedures of analysis include the set of generic methods of exploratory analysis and the visualization of data, in general through maps. These techniques permit the description of the distribution of the variables of study, the identification of observations that are outliers not only in relation to the type of distribution but also in relation to its neighbors, and to explore the existence of patterns in the spatial distribution. Through these procedures it is possible to come up with hypotheses about the observations, in a way of selecting the best inferential model supported by the data.

3.1 Spatial Analysis Processes

The spatial inferential models are usually presented in three forms: continuous variation, discrete variation, and the point processes. The resolution of a spatial problem may involve the utilization of one or the interaction of some or even all of them.

3.1.1 Point processes

Point processes are a set of irregularly distributed points in a terrain, whose location was generated by a stochastic mechanism. The localization of points is the object of study, which has the objective of understanding its generating mechanism. A set of points $(u_1, u_2, \ldots . u_n)$ in a certain region A is considered where events occurred. For example, if the phenomenon under study is homicides that occurred in a certain region, there may be need to verify if there is any geographic pattern for this kind of crime, that is, to find sub-regions in A with greater probability of occurrence.

3.1.2 Continuous processes

The inferential models of continuous variation consider a stochastic process $Z(u), \mu \epsilon A, A \in K$ whose values can be known in every point of the study area. Starting from a sample of one attribute z, collected in various μ points contained in A, $\{z(\mu_a), a = 1, 2, \ldots, n\}$, we aim at inferring a continuous surface of values of z. The estimation of this stochastic process can be done in a completely non-parametric way or from kriging estimators. These classical inferential models of surfaces estimation are denominated geostatistics. Geostatistics uses two types of estimation procedures: the kriging and the stochastic simulation. In kriging, at each point μ_0, a value of the random variable Z is estimated $\bar{Z}(\mu_0)$,using an estimator $Z'(\mu_0)$, that is a function of the data and of the spatial covariance structure $Z'(\mu_0) = f(C, (n))$. These estimators present some important properties: they are not biased and are optimal in the sense that they minimize

the functions of the inferential errors.

3.1.3 Discrete processes

The inferential models of discrete variation concern the distribution of events whose localization is associated to areas delimited by polygons. This case occurs much frequently when we deal with phenomena aggregated by municipalities, quarters or census tracts, like population, mortality and income. In this case, we dont have the exact locality of the events, but value aggregated by area. The objective is to model the pattern of spatial occurrence of the geographic phenomenon under study. In this type of modeling we consider that the geographic space under study, region A, is a fixed set of spatial units. The most used model of distribution considers a stochastic process $Z_i : i = 1, 2, \ldots, n$, composed of a set of random variables. We seek to construct an approximation of the joint distribution of these variables $Z = Z_1, \ldots, Z_n$, where each random variable is associated with one of the areas and has a distribution to be estimated. If the process is stationary, the expected value of Z_i is the global mean of the region and the covariance structure depend only on distance, or on the neighborhood structure between the areas.

3.2 Spatial Data Analysis

Spatial data analysis deals with the situation where observational data are available on some process operating in space and methods are sought to describe or explain the behavior of this process and its possible relationship to other spatial phenomena. Spatial data analysis is involved when data are spatially located and explicit consideration is given to the possible importance of their spatial arrangement in the analysis or in interpretation of results.

The Laplace Principle of probability theory asserts that if there is no information to indicate that either of two events is more likely, then they should be treated as equally likely, i.e., as having the same probability of occuring. That is, if we have a square divided into two parts left and

right and applying this principle to the case of a randomly located point in square, S , there is no reason to believe that this point is more likely to appear in either left half or the (identical) right half. So these two (mutually exclusive and collectively exhaustive) events should have the same probability, $\frac{1}{2}$. But if these halves are in turn divided into equal quarters, then the same argument shows that each of these four occupancy events should have probability $\frac{1}{4}$. If we continue in this way, then the square can be divided into a large number of n grid cells, each with the same probability, $\frac{1}{n}$, of containing the point. Now for any subregion (or cell), $C \subset S$, the probability that C will contain this point is at least as large as the sum of probabilities of all grid cells inside C , and similarly is no greater that the sum of probabilities of all cells that intersect C . Hence by allowing n to become arbitrarily large, it is evident that these two sums will converge to the same limit, namely the fractional area of S inside C. Hence the probability, $Pr(C|S)$ that a random point in S lies in any cell $C \subset S$ is proportional to the area of C.

$$Pr(C|S) = \frac{a(C)}{a(S)} \tag{3.1}$$

since this must hold for any pair of nested regions $C \subset S$ it follows that,

$$Pr(C|S) = Pr(C|R).Pr(R|S) \Rightarrow Pr(C|R) = \frac{Pr(C|S)}{Pr(R|S)} = \frac{a(C)/a(S)}{a(R)/a(S)}$$

This implies that

$$Pr(C|R) = \frac{a(C)}{a(R)} \tag{3.2}$$

hence the square can be replaced by any bounded region, R , in the plane.

This fundamental proportionality result forms the basis for almost all models of spatial randomness.

In probability terms, this principle induces a uniform probability distribution on R, describing the location of a single random point. With respect to any given cell, $C \in R$, it convenient to characterize this event as a Bernoulli (binary) random variable, X(C), where X(C)=1 if the point is located in C and X(C) = 0 otherwise. In these terms, it follows from above that the conditional

probability of this event (given that the point is located in R) must be

$$Pr[X(C) = 1|R] = \frac{a(C)}{a(R)} \tag{3.3}$$

so that

$$Pr[X(C) = 0|R] = 1 - Pr[X(C) = 1|R] = 1 - [a(C)/a(R)]. \tag{3.4}$$

3.3 Complete Spatial Randomness

The basic "reference" or "benchmark" model of a point process is the uniform Poisson point process in the plane with intensity, sometimes called Complete Spatial Randomness (CSR).

Its basic properties are

- the number of points falling in any region A has a Poisson distribution with mean $\lambda area(A)$
- given that there are n points inside region A, the locations of these points are independent and identically distributed and uniformly distributed inside A
- the contents of two disjoint regions A and B are independent.

The uniform Poisson process is often the null model in an analysis.

There are three approaches to testing the CSR hypothesis: the quadrat method, the nearest-neighbor method, and the method of K-functions

3.3.1 Quadrat Method

If there are n points in R, and if we let $a = a(C)$, then the estimated point density λ is given by

$$\lambda = \frac{n}{a(R)} \tag{3.5}$$

then this common Poisson cell-count distribution has the form

$$Pr[N_i = k|\lambda] = \frac{(\lambda a)^k}{k!} e^{-\lambda a}, k = 0, 1, 2 \ldots \tag{3.6}$$

Moreover, since the CSR Hypothesis also implies that each of the cell counts, $N_i = N(C_i) i = 1, \ldots, k$, is independent, it follows that $(N_i : i = 1, .., k)$ must be an independent random samples from this Poisson distribution. Hence the simplest test of this hypothesis is to use the Pearson chi^2 goodness-of-fit test. Here the expected number of points in each cell is given by the mean of the Poisson above, which (recalling that a=a(R)/m by construction) is

$$E(N/\lambda) = a.\frac{n}{a(R)} = \frac{n}{m}$$

Hence if the observed value of N_i is denoted by n_i , then the chi-square statistic

$$x^2 = \sum_{i=1}^{m} \frac{(n_i - n/m)^2}{n/m} \tag{3.7}$$

is known to be asymptotically chi-square distributed with $m - 1$ degrees of freedom, under the CSR Hypothesis. Thus one can test this hypothesis directly in these terms. But since n/m is simply the sample mean, i.e.,

$$n/m = (1/m) \sum_{i=1}^{m} n_1 = \bar{n}$$

this statistic can also be written as

$$x^2 = \sum_{i=1}^{m} \frac{(\bar{n}_{i-n})^2}{\bar{n}} = (m-1)\frac{s^2}{\bar{n}} \tag{3.8}$$

where $s^2 = \frac{1}{(m-1)} \sum_{i=1}^{m} (\bar{n}_{i-n})^2$ is the sample variance. But since the variance of the Poisson distribution is exactly the mean, it follows that $var(N)/E(N) = 1$ under CSR. Moreover, since $s^2/\bar{n}^2$ is the natural estimate of this ratio, this ratio is often designated as the index of dispersion, and used as a rough measure of dispersion versus clustering. If $s^2/n < 1$ then there is too little variation among quadrat counts, suggesting possible dispersion rather than randomness. Similarly, if $s^2/n > 1$ then there is too much variation among counts, suggesting possible clustering rather than randomness.

But this testing procedure is very restrictive in that it requires a rectangular region. More importantly, it depends critically on the size of the partition chosen. As with all applications of Pearsons goodness-of-fit test, if there is no natural choice of partition size, then the results can be very sensitive to the partition chosen. The power of the quadrat test depends on the size of quadrats, and falls to zero for quadrats which are either very large or very small. The power also depends on the alternative hypothesis, in particular on the spatial scale of any departures from the assumptions of constant intensity and independence of points. The choice of quadrat size carries an implicit assumption about the spatial scale.

3.3.2 Two-Tailed Test of CSR

The standard test of CSR in most software is a two-tailed test in which both the possibility of "significantly small" values of $\bar{d_m}$ (clustering) and "significantly large" values of $\bar{d_m}$ (dispersion) are considered. The upper-tail points, z_α, for the standard normal distribution is defined by $Pr(Z \geq z_\alpha) = \alpha$ for Z N(0,1). It follows that for the standardized mean in (18)

$$Pr(|Z_m| \geq z_{\alpha/2}) = Pr\left(Z_m \leq -z_{\alpha/2} or (z_{\alpha/2} \leq -z_m)\right] = \alpha \tag{3.9}$$

under the CSR Hypothesis. Hence if we estimates point density as above and constructs corresponding estimates of the mean and standard deviation under CSR by

$$\bar{\mu} = \frac{1}{2\sqrt{\lambda}}, \bar{\sigma} = \sqrt{\frac{(4-\pi)}{(m4\pi\lambda)}}$$

We can then test the CSR Hypothesis by constructing the following standardized sample mean:

$$Z_m = \frac{\bar{d_m} - \bar{\mu}}{\bar{\sigma}} \tag{3.10}$$

If the CSR Hypothesis is true, then by (3.9) and (3.10) above, Z_m should be a sample from N(0,1). Hence a test of CSR at the α-level of significance is then given by the rule:

Two-Tailed CSR Test : Reject the CSR Hypothesis if and only if $|Z_m| > Z_{\alpha/2}$

One-Tailed Test of Clustering and Dispersion

Values of $\bar{d}_m$ (and hence z_m) that are too low to be plausible under CSR are indicative of patterns more dispersed than random. Similarly, values too large are indicative of patterns more clustered than random. In many cases, one of these alternatives is more relevant than the other. So the key question here is whether this pattern is significantly more clustered than random. Similarly, one can ask whether the pattern is significantly more dispersed than random. This leads naturally to one-tailed versions of the test above. First, a test of clustering versus the CSR Hypothesis at the $\alpha - level$ of significance is given by the rule:

Clustering versus CSR Test : Conclude significant clustering if and only if $z_m = z_\alpha$

The same applies when testing for dispersion versus the CSR Hypothesis at the $\alpha-$ level of significance where we use the rule:

Dispersion versus CSR Test: Conclude significant dispersion if and only if $z_m > z_\alpha$ In the one-tailed test of clustering versus CSR above, suppose that for the observed standardized mean value, z_m , one simply asks how likely it would be to obtain a value this low if the CSR Hypothesis were true. This question is answered by calculating the probability of a sample value as low as z_m for the standard normal distribution N(0,1) . If the cumulative distribution function for the normal distribution is denoted by

$$\Phi(Z) = Pr(Z \leq z) \tag{3.11}$$

then this probability, called the P-value of the test, is given by

$$Pr(Z \leq z_m) = \Phi(Z_m) \tag{3.12}$$

Unlike the significance level, α, above, the P-value for a test depends on the realized sample value, z_m , and hence is itself a random variable that changes from sample to sample. However, it can be related to α by observing that if $P(Z \leq z_m) \leq \alpha$, then for a test of size α, one would conclude that there is significant clustering. More generally the P-value, $(P(Z \leq z_m))$ can be

defined as the largest level of significance (smallest value of α) at which CSR would be rejected in favor of clustering based on the given sample value, z_m.

Similarly, one can define the P-value for a test of dispersion the same way, except that now for a given observed standardized mean value, z_m, one asks how likely it would be to obtain a value this large if the CSR Hypothesis were true. Hence the P-value in this case is given by

$$Pr(Z \geq z_m) = Pr(Z > z_m) = 1 - Pr(Z \geq z_m) = 1 - \Phi(z_m) \tag{3.13}$$

The corresponding P-value for the general two-tailed test is given as the answer to the following question: How likely would it be to obtain a value as far from zero as z_m if the CSR Hypothesis were true? More formally this P-value is given by

$$P(|Z| \geq z_m) = 2.\Phi(-|z_m|) \tag{3.14}$$

Here the absolute value is used to ensure that -$|z_m|$ is negative regardless of the sign of z_m . Also the factor 2 reflects the fact that values in both tails are further from zero than z_m.

3.4 Continuous Spatial Data Analysis

The key difference between continuous spatial data and point patterns is that there is now assumed to be a meaningful value, Y(s), at every location, s, in the region of interest. For example, Y(s) might be the malaria cases at s or the human population at s.

If the region of interest is again denoted by R, and if the value, $Y(s)$, at each location, $s \in R$ is treated as a random variable, then the collection of random variables

$$Y(s) : s \in R \tag{3.15}$$

is designated as a spatial stochastic process on R (also called a random field on R). Such (uncountably) infinite collections of random variables cannot be analyzed in any meaningful way without making a number of strong assumptions.

There is a clear parallel between spatial stochastic processes and temporal stochastic processes,

$$Y(t) : t \in T \tag{3.16}$$

where the set, T , is some continuous (possibly unbounded) interval of time.

3.4.1 Basic Modeling Framework

Spatial statistical models start by decomposing the statistical variation of random variables, $Y(s)$, into a deterministic trend term, $\mu(s)$, and a stochastic residual term, $\epsilon(s)$, as follows

$$Y(s) = \mu(s) + \epsilon(s), s \in R \tag{3.17}$$

where $\mu(s)$ is taken to be the mean of $Y(s)$ so that by definition,

$$\epsilon(s) = Y(s) - \mu(s) \Rightarrow E[\epsilon(s)] = E[Y(s)] - \mu(s) \tag{3.18}$$

This implies

$$E[\epsilon(s)] = 0, s \in R$$

The above two equations together constitute the basic modeling framework used in the analysis. This framework is a convenient representation of Y(s), and involves no substantive assumptions. Since $\mu(.)$ defines a deterministic function on R, it is useful to think of $\mu(.)$ as a spatial trend function representing the typical values of the given spatial stochastic process over all R , i.e., the global structure of the Y -process. Similarly, since $\epsilon(.)$ is by definition a spatial stochastic process on R with mean identically zero, it is useful to think of $\epsilon(.)$ as a spatial residual process representing local variations about $\mu(.)$, i.e., the local structure of the Y -process.

Within this framework, our basic modeling strategy will be to identify a spatial trend function, $\mu(.)$, that fits the Y -process so well that the resulting residual process, $\epsilon(.)$, is not statistically distinguishable from random noise. To make this strategy precise, we must of course develop appropriate models of random noise (in a manner paralleling the CSR hypothesis)

Suppose that n points are each located randomly in region R. Then the other key assumption of spatial randomness is that the locations of these points have no influence on one another. Hence if for each i= 1,..,n, the Bernoulli variable, $X_i(C)$, now denotes the event that point i is located in region C , then under spatial randomness the random variables $X_i(C) : i = 1, \ldots, n$ are assumed to be statistically independent for each region C . This together with the Spatial Laplace Principle above defines the fundamental hypothesis of complete spatial randomness (CSR), which we shall usually refer to as the CSR Hypothesis.

In terms of the individual variables, $X_i(C)$, the total number of points appearing in C , designated as the cell count, N(C) , for C , must be given by the random sum

$$N(C) = \sum_{i=1}^{n} X_i(C) \tag{3.19}$$

[It is this additive representation of cell counts that motivates the Bernoulli (0-1) characterization of location events above]. Since the expected value of a Bernoulli random variable, X, is simply $P(X = 1)$, it follows (from the linearity of expectations) that the expected number of points in C must be

$$\begin{aligned} E[N(C)|n, R] &= \sum_{i=1}^{n} E[X_i(C)|R] = \sum_{i=1}^{n} Pr[X_i(C) = 1|R] \\ &= \sum_{i=1}^{n} \frac{a(C)}{a(R)} = n.\frac{a(C)}{a(R)} = \frac{n}{a(R)}a(C) \end{aligned} \tag{3.20}$$

Finally, it follows from expression (1) that under the CSR Hypothesis, the sum of independent Bernoulli variables in (2) is by definition a Binomial random variable with distribution given by

$$Pr[N(C) = k/n, R] = \frac{n!}{k!(n-k)!}\left(\frac{a(C)}{a(R)}\right)^k \left(1 - \frac{a(c)}{a(R)}\right)^{n-k}, k = 0, 1, \ldots, n \tag{3.21}$$

For most practical purposes, this conditional cell-count distribution for the number of points in cell, $C \subset R$ (given that n points are randomly located in R) constitutes the basic probability model for the CSR Hypothesis.

3.4.2 Generalized Spatial Randomness

The above notion of spatial randomness is derived from the principle that regions of equal area should have the same chance of containing any given randomly located point. More formally, this Spatial Laplace Principle asserts that for any two subregions (cells), C_1 and C_2 , in R ,

$$a(C_1) = a(C_2) \Rightarrow Pr[X(C_1) = 1/R] = Pr[X(C_2) = 1/R] \tag{3.22}$$

However, simple area may not always be the most relevant reference measure (backcloth). In the example of malaria, if malaria cases are spatially random, then each individual should have the same chance of contracting this disease. So here, the existing population distribution becomes the relevant reference measure.

To generalize this notion of spatial randomness, we need only replace area with the relevant reference measure, say p(C), which may be the number of houses in C or the total population of C. As an extension of the above, we then have the following Generalized Spatial Laplace Principle: For any two sub regions (cells), C_1 and C_2, in R:

$$p(C_1) = p(C_2) \Rightarrow Pr[X(C_1) = 1/R] = Pr[X(C_2) = 1/R]$$

3.4.3 Spatial Dependence

Spatial dependence in a collection of sample data implies that observations at location i depend on other observations at locations j, $j \neq i$. Formally, stated as; $yi = f(yj)$, $i = 1, \ldots, n$.

Note that dependence is among several observations, as the index i can take on any value from $i = 1, \ldots, n$.

3.5 Spatial Autocorrelation

Spatial autocorrelation is when the value at any one point in space is dependent on values at the surrounding points. That is, the arrangement of values is not just random. Positive spatial correlation means that similar values tend to be near each other. Negative spatial correlation means that different values tend to be near each other.

3.5.1 Moran's I

It is often used to measure the spatial autocorrelation of ordinal, interval or ratio data. Morans I-for a spatial proximity matrix W spatial correlation in attribute values x_i is estimated as:

$$I(d) = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_i j(xi-\mu)(xj-\mu)}{(n\sum_{i=1}^{n}(x_i-\mu)^2)(n\sum_{i=1}^{n}\sum_{j\neq 1}^{n} w_i j)}$$

Where $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$ ie the average of x_i over the n locations x_i is the observed value of population at location i and w_{ij} is the spatial weight measure of contiguity and is defined as 1 if location i is contiguous to location j and 0 otherwise.

The expected value and variance of morans I for a sample of size n could be calculated according to the assumed pattern of the spatial data distribution (Cliff and Ord, 1981) For the assumption of a normal distribution:

$$E_n(I) = -\frac{1}{(n-1)}$$

$$VAR_n(1) = \frac{n^2 w_1 - n w_2 + 3w_0}{w_2 0(n2-1)} - E_n^2(I)$$

For the assumption of random distribution:

$$E_r(I) = -\frac{1}{n-1}$$

$$VAR_r(I) = \frac{n((n^2-3n+3)w_1 - nw_2 + 3w_0^2) - K_2((n^2-n)w_1 - 2nw_2 + 6w_0^2)}{w_0^2(n-1)(n-2)(n-3)} - E_r^2(I),$$

where, $K_2 = \frac{n\sum_{i=1}^{n}(x_1-2)^4)}{\sum_{i=1}^{n}(x_i-x)^2)}$, $w_o = \sum_{i=1}^{n}\sum_{j=1}^{n} w_ij$, $w_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_(ij) + w_(ji))^2$, $w_2 = \sum_{i=1}^{n}(w_(i.) + w_(.i))^2$

$w_(i.)$ and $w_(.i)$ are the sum of the row i and column i of the matrix respectively.

Negative (positive) values indicate negative (positive) spatial autocorrelation. In practice, values greater than 2 or smaller than -2 indicate spatial autocorrelation that is significant at the 5 percent

Moran's I is inversely related to Geary's C, but it is not identical. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation.

Hypothesis Testing

The test of the null hypothesis that there is no spatial autocorrelation between observed values over the n locations can be conducted on the basis of the standardized statistics as follows:

$$Z(d) = \frac{I(d) - E(I)}{\sqrt{VAR(I)}}$$

Morans I is significant and positive when the observed values of locations within a certain distance (d) tend to be similar, negative when they tend to dissimilar, and approximately zero when the observed values are arranged randomly and independently over space

Generalized Spatial Linear Model

A generalized spatial linear model is an extension of the generalized linear models by incorporating spatial dependence into the model.

1. An error distribution for the response variable within the exponential family of distribution.
2. A monotonic link function, g(.), such that:

$g(p_i) = x_i^T + \gamma \sum_{j=1}^{n} c_{ij} y_j,$

where p_i is the mean value of the response variable, is the spatial autocorrelation parameter, c_ij represents the geographical arrangement of data values and $g(.)$ link function.

3.6 Application to the study

The data we have was collected from specific points (health facilities) over time. In order to determine the spatial clustering of malaria cases, we are going to apply the continuous spatial data analysis. This will allow us to see the variation of the cases over time considering the catchment population for each facility and also the two classes of patients; under fives and over fives.

3.6.1 Brief Description of the data

The data used in this study is from the HMIS of Kilifi District. The original data contains the Total number of patients seen from all the Government Health facilities in the district and the condition they were suffering from. This data is collected every day from each of the health facilities and sent to the district hospital for compilation. It is divided into patients under five years and those over five years. The data also contain the division within which the facilities are located. The catchment population for each of the facilities is obtained from Facility Health Committees which represent the community in the running of all these facilities. We have the Village Health committees who are responsible for educating the community on health matters and each VHC is responsible for ten homesteads and they therefore know the total population under their watch.

3.6.2 Definition of variables

- Facility- the name of the facility
- Lat -the latitude of the facility
- Long -the longitude of the facility
- Pop05 -catchment population of the facility for 2005

- Pop06 -catchment population of the facility for 2006
- Pop07 -catchment population of the facility for 2007
- Division- The division in which the facility is based
- Mal04- the number of malaria cases in 2004
- Allcases04- the total number of patient who visited the facility with various conditions in 2004
- Mal05- the number of malaria cases in 2005
- Allcases05- the total number of patient who visited the facility with various conditions 2005
- Mal06- the number of malaria cases in 2006
- Allcases06- the total number of patient who visited the facility with various conditions 2006
- Mal07- the number of malaria cases in 2007
- Allcases07- the total number of patient who visited the facility with various conditions 2007
- Mal08- the number of malaria cases in 2008
- Allcases08- the total number of patient who visited the facility with various conditions 2008
- Mal05/allcases05- the proportion of malaria to all cases in the facility for 2005
- Mal06/allcases06- the proportion of malaria to all cases in the facility for 2006
- Mal0 Mal07/allcases07- the proportion of malaria to all cases in the facility for 2007
- Mal08/allcases08- the proportion of malaria to all cases in the facility for 2008
- Prop05-proportion of malaria cases to the catchment population for 2005
- Prop06-proportion of malaria cases to the catchment population for 2006

- Prop07-proportion of malaria cases to the catchment population for 2007
- Prop08-proportion of malaria cases to the catchment population for 2008
- Divisions 1-Bamba, 2-Kikambala, 3-Kaloleni, 4-Chonyi, 5-Bahari, 6-Vitengeni, 7-Ganze.

Chapter 4

Results

This chapter present summary statistics, exploratory analysis and the results of spatial data analysis. The map shows the incidence density for the patients over five years. We see that there

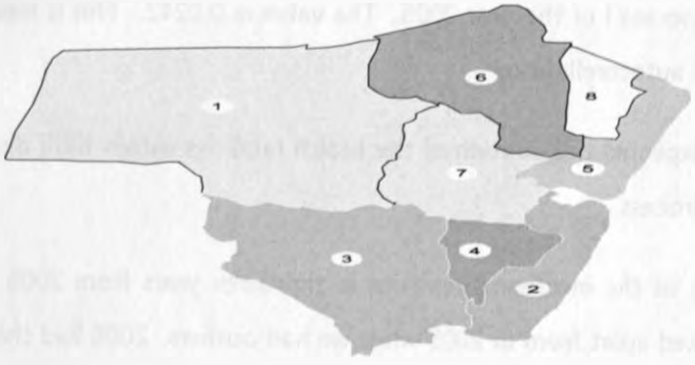


Figure 4.1: The incidence density for patients over five years

is high concentration around the ocean that reduces as one moves eastwards.

This figure also shows the incidence density but this time its for the patients under five years. We see that as one moves northward, the intensity reduces. Regions away from the ocean have less incidence intensity as compared to those close to the ocean.

Figure 4.2: The incidence density for patients under five years

Figure 4.3 above gives us the summary of the variable in the data. It also shows the mean, standard deviation, minimum and maxium values of the variables.

The above gives us the morans I of the year 2005. The value is 0.0242. This is less than one and hence there is no spatial autocorellation.

Figure 4.5 gives us the expected distribution of the health facilities within Kilifi district using the inhomogenous poisson process.

Figure 4.6 boxplot gives us the incidence densities in the three years from 2005 to 2007. The data is normally distributed apart from in 2005 when we had outliers. 2006 had the biggest range of 0.7.

Figure 4.7 shows the boxplot for the incidence density for the year 2005 for the different divisions. Two divisions,Bamba and Kaloleni had outliers.

Figure 4.8 above shows the boxplot for the incidence density for the year 2006 for the different divisions. The data here is generally normally distributed

Figure 4.9 shows the boxplot of the incidence density for the different divisions for the year 2007. Kikambala (Division two)had outliers in this year.

Figure 4.10 shows us the spatial clustering of the malaria cases for the population of the over

five years. This is the figure without incorporating the actual Kilifi map.

Figure 4.11 shows us the spatial clustering of the malaria cases for the population of the over five years. This is the figure does not include the intensity of the health facilities but just shows the the cummulative incidences.

The graphs in figure 4.12 above shows the K,J,G and F functions. They measure spatial clustering. K function measures clustering at multiple different distances, G function gives the probability of an observed point's nearest neighbour appearing at any given distance r. F function gives the probability of a random empty location having a nearest neighbour at a given distance r.

Table 4.1: Summary of the Incidence densities

		Mean	N	SD	Min	Min
I	2005	0.2	37	0.15	0.03	0.62
	2006	0.25	37	0.16	0.04	0.59
	2007	0.23	37	0.15	0.05	0.56
II	2005	0.3	37	0.22	0.03	0.91
	2006	0.36	37	0.2	0.08	0.86
	2007	0.37	37	0.21	0.08	0.75

Table 4.1 gives us a summary of the incidence densities for the two ages: over fives and the under fives when compared for the different years. The under fives have the highest mean of 0.3,0.36 and 0.37 for the years 2005, 2006 and 2007 respectively.

Table 4.2: Anova Table for the different years

Year	Source	DF	SS	MS	F	Sig
2005	Between Groups	1	0.166	0.166	4.632	0.035
	Within Groups	72	2.573	0.036		
	Total	73	2.739			
2006	Between Groups	1	0.216	0.216	6.720	0.012
	Within Groups	72	2.317	0.032		
	Total	73	2.533			
2007	Between Groups	1	0.365	0.365	11.111	0.001
	Within Groups	72	2.368	0.033		
	Total	73	2.733			

Table 4.2 shows the difference in the different years. We see that there is significant difference in the years as the significance ranges from 0.001 in 2007 to 0.035 in the year 2005.

In table 4.3 we get analysis by type of health facility visited by patients. We can see that there is no significant difference in the malaria incidence density as the p values for all the three years are above 0.05. When we compare the malaria to total morbidity we see significant difference

ranging from 0.01 to 0.00012.

Table 4.4 shows us the two AIC for the two possible models. We pick the one with the smallest AIC of -207.721

Table 4.3: Analysis by type of Health Facility

	Year	District Hosp.	Health Centre	Dispensary	Clinic	P-value
Incidence density	2005	0.201	0.634	0.632	0.202	0.13
	2006	0.190	0.587	0.726	0.335	0.086
	2007	0.144	0.580	0.593	0.272	0.102
Malaria to total morbidity	2005	0.255	0.3292	0.365	0.343	0.013
	2006	0.461	0.297	0.335	0.320	0.01
	2007	0.508	0.299	0.281	0.261	0.00012

Table 4.4: Model Selection

		Intercept	X	Y	Log (Lambda)	AIC
Fit 1 (Non stationary poisson process)	Fitted coefficient for trend	-48.21	1.32	0.08		-207.72
	Fitted regular parameters (theta)	-48.21	0.13	0.08		
	Fitted exp (theta)	0.00	0.04	0.001		
Fit 2 (Stationary poisson process)	Uniform intensity	46.25				-207.085
	Fitted regular parameters (theta)				3.834	
	Fitted exp (theta)				46.25	

Figure 4.3: Summary statistics

Report

Division		Pop05	Pop06	Pop07	mal05	Allcases05	mal06	Allcases06	mal07	All cases 07
1	Mean	12944.25	13333	13746	912.174869	4812.8	1212.44672	7542	1407.76348	7602.75
	Std. Deviation	13626.367	14035	14470	518.682675	4047.7	1028.42223	6168	1437.79986	6622.03
	Minimum	2945	3033	3127	309.7828889	1566	306.3231536	1770	315.664490	1833
	Maximum	32510	33485	34523	1558.59541	10720	2679.14412	16272	3493.83312	16907
2	Mean	17213.00	17729	18279	1485.79540	8173.8	1724.94492	10398	1645.42431	11140.0
	Std. Deviation	14841.631	15287	15761	672.738005	4062.5	731.924367	5071	1001.56000	7339.45
	Minimum	5151	5306	5470	131.5436325	447	484.2589674	2425	317.233614	2318
	Maximum	47924	49362	50892	2325.07773	13900	3071.57038	17415	3415.93553	22025
3	Mean	19394.33	19976	20595	1569.89588	8518.3	1956.04444	10024	1825.01185	10174.7
	Std. Deviation	19671.506	20262	20890	1154.35490	6085.3	1159.90041	4161	934.090000	4428.51
	Minimum	2436	2509	2587	170.2708020	679	293.8708689	3266	730.774762	3627
	Maximum	58132	59876	61732	4478.18008	23590	3899.30156	16536	3308.34560	17912
4	Mean	24119.50	24843	25613	2234.50643	10824	2276.31782	13049	2111.21537	12170.5
	Std. Deviation	20100.924	20704	21346	152.436389	256.68	824.598541	4711	183.543396	3581.50
	Minimum	9906	10203	10519	2126.71763	10642	1693.23860	9718	1981.43059	9638
	Maximum	38333	39483	40707	2342.29524	11005	2859.39704	16380	2241.00015	14703
5	Mean	25600.50	26369	27186	2133.70092	13871	3283.74071	15373	2890.88318	15892.8
	Std. Deviation	28014.328	28855	29749	1912.32453	12649	3452.40166	13711	2897.39044	13950.0
	Minimum	6166	6351	6548	461.8078272	3029	487.4745485	3336	351.386303	3650
	Maximum	67205	69221	71367	4840.33310	31822	8258.92102	34170	7005.67898	35340
6	Mean	11865.75	12222	12601	771.727897	5744.5	1358.52511	8401	1268.37695	8895.75
	Std. Deviation	8643.877	8903	9179.3	455.696049	3484.0	707.302472	2672	502.797177	2784.75
	Minimum	3283	3381	3486	417.7433613	2591	636.8017170	4912	574.801556	4776
	Maximum	19308	19887	20504	1438.54018	10724	2332.22940	11420	1747.31504	10712
7	Mean	15559.00	16026	16523	1318.53536	6176.0	1595.72884	10178	1818.75009	13004.0
	Std. Deviation	1996.870	2057	2120.6	102.223217	2210.4	554.492768	294.9	951.066674	1728.17
	Minimum	14147	14571	15023	1246.25263	4613	1203.64325	9969	1146.24439	11782
	Maximum	16971	17480	18022	1390.81809	7739	1987.81444	10386	2491.25578	14226
Total	Mean	18071.57	18614	19191	1475.33534	8310.6	1896.21570	10421	1806.40504	10872.2
	Std. Deviation	16800.979	17305	17841	1013.87575	6081.0	1414.28440	5966	1265.21024	6623.32
	Minimum	2436	2509	2587	131.5436325	447	293.8708689	1770	315.664490	1833
	Maximum	67205	69221	71367	4840.33310	31822	8258.92102	34170	7005.67898	35340

Figure 4.4: Moran's I

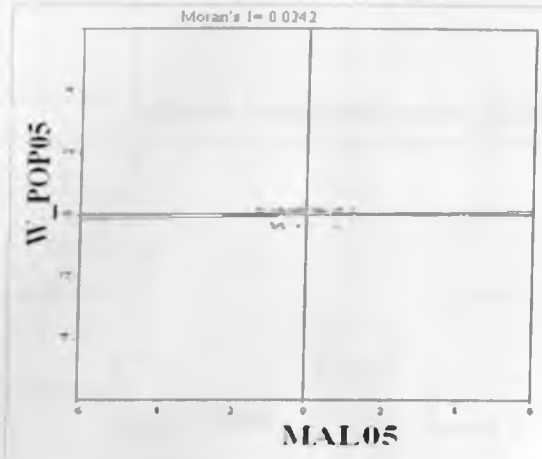


Figure 4.5: Inhomogeneous poisson figure

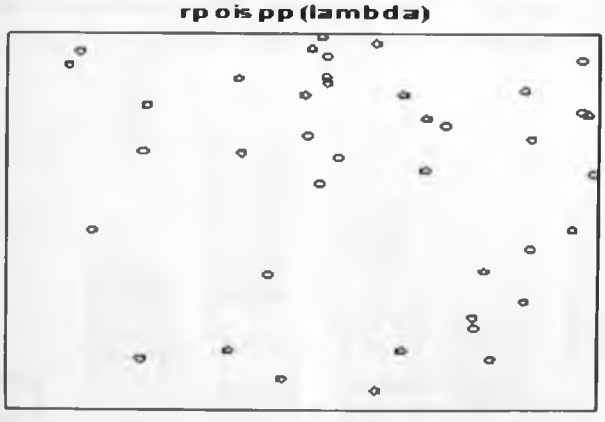


Figure 4.6: Box-plot for all years incidence density

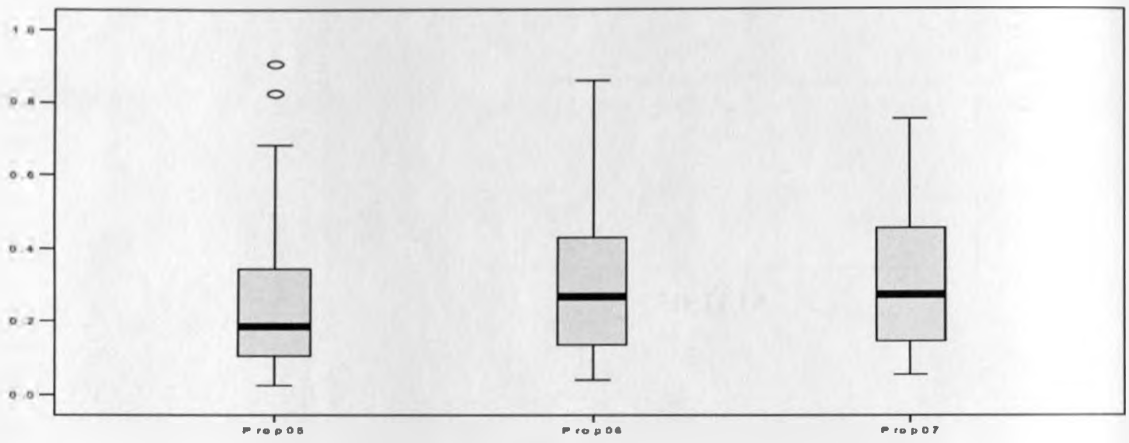


Figure 4.7: Box-plot for 2005 incidence density

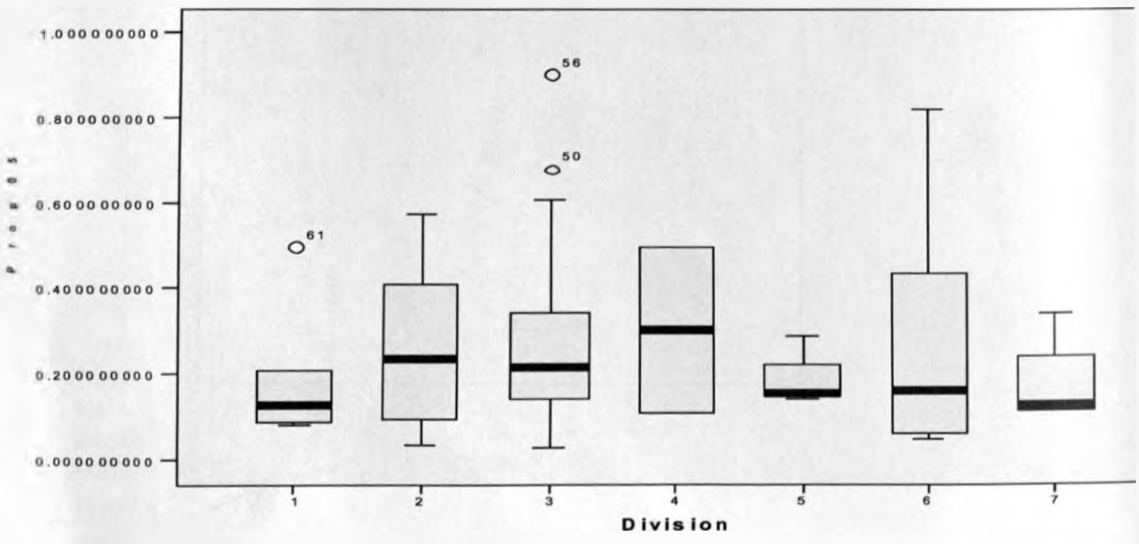


Figure 4.8: Box-plot for 2006 incidence density

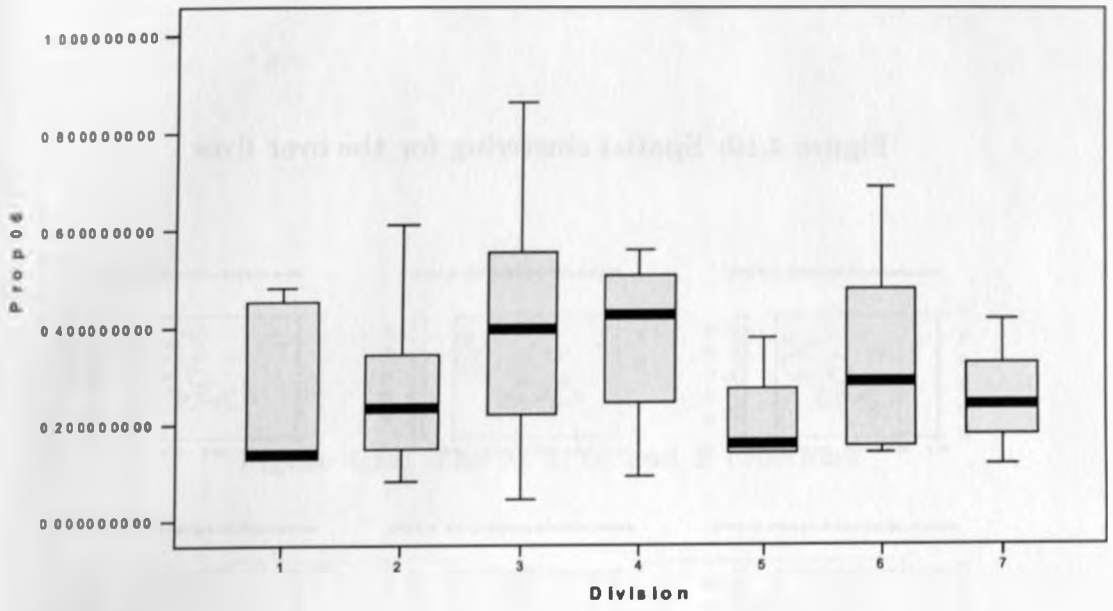


Figure 4.9: Box-plot for 2007 incidence density

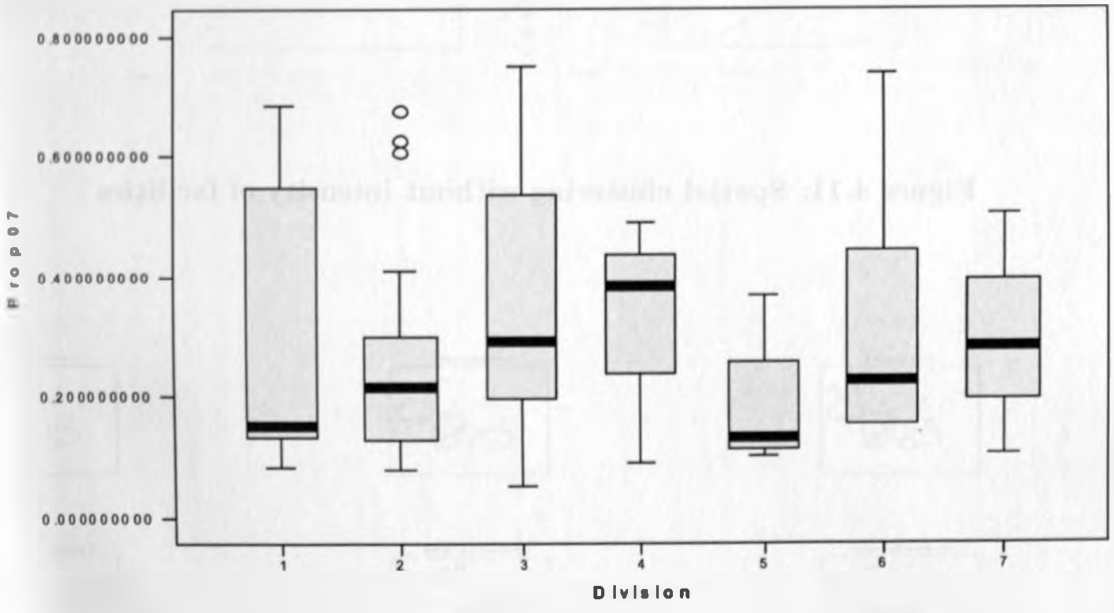


Figure 4.10: Spatial clustering for the over fives

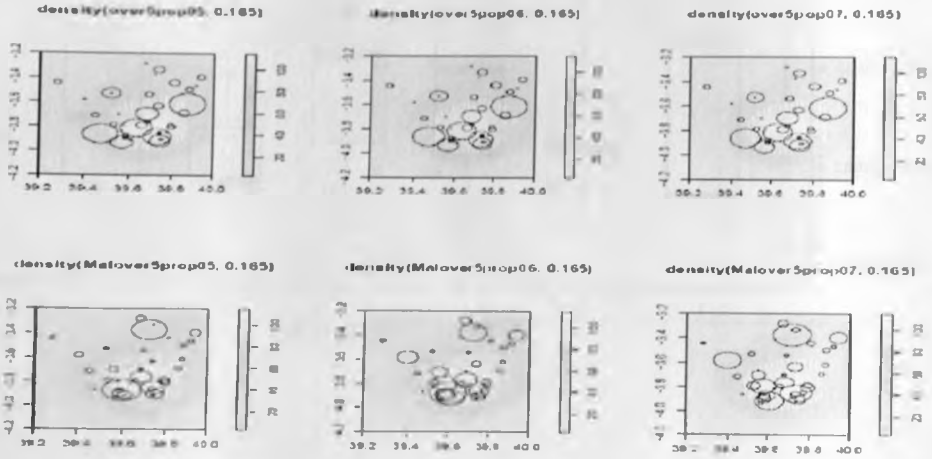


Figure 4.11: Spatial clustering without intensity of facilities

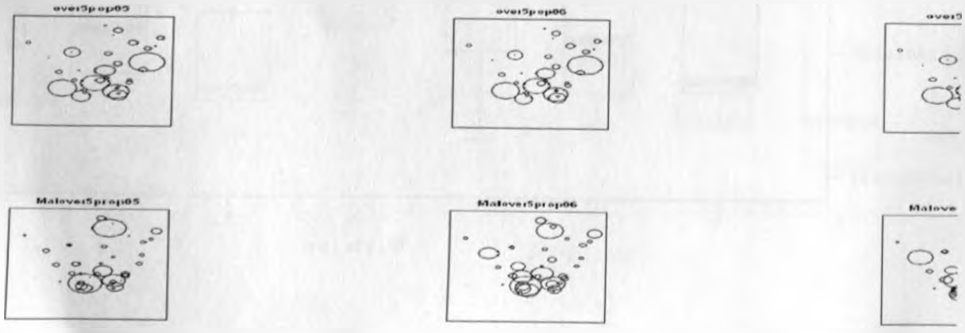
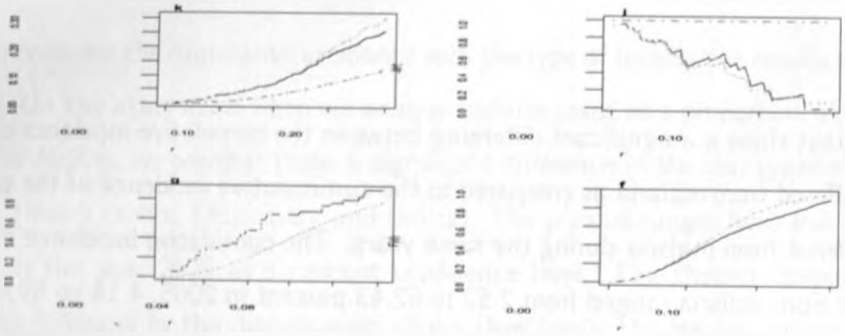


Figure 4.12: The K, J, G, and F functions



Chapter 5

Discussion and Conclusion

The results show that there is a significant difference between the cumulative incidence of children under five who suffered from malaria as compared to the cummulative incidence of the population over five who suffered from malaria during the same years. The cumulative incidence of the over fives who suffered from malaria ranged from 2.52 to 62.43 percent in 2005, 4.14 to 59.43 percent in 2006 and 5.26 to 57.57 percent in 2007. The cumulative incidence of the children who suffered from malaria ranged from 3.39 to 90.66 percent in 2005, 8.00 to 85.87 percent in 2006 and 7.75 to 74.97 percent in 2007. The children under five years suffered highest in 2007 when the mean incidence was 36.996 percent while those over five years suffered the highest in 2006 when the incidence was 25.27 percent. This can be seen in table 4.1.

When we compare the malaria incidences in the three years; 2005,2006 and 2007, we see from table 4.2 that there was a significance difference in the malaria incidences in these three years with a pvalue of less than 0.05.

Figure 4.3 gives us a summary of the data in all the seven divisions showing the mean, standard deviation, minimum and maximum values.

The total malaria cases in comparison with the total diseases cases in the district for patients over

five years ranged from 25.47 percent in 2008 to 32.89 percent in 2005. This gives an average of 29.6 percent which compares to the national average of about 30 percent.

The malaria cases in comparison to the total outpatient morbidity in the district for the patients under five ranged from 30.36 percent in the year 2007 to 37.41 percent in the year 2005. This has an average of 33.39 percent. This compares well with the country's average of 30 percent. There is a significant difference in the proportion of malaria cases to total outpatient morbidity for the patients less than five years as compared to those patients over five years in the two years of 2005 and 2006 were the p-values were 0.006 and 0.048 respectively at 5 percent confidence.

This study therefore shows that there is a significant difference between the proportions of malaria cases to the total outpatient attendance for those over five as compared to those less than five years for the twin years of 2006 and 2007.

When we consider the cumulative incidence with the type of facility, the results are not significantly different. On the other hand when we analyze malaria cases as a proportion of the total morbidity per type of facility, we see that there is significant difference in the four types of facilities (District Hospital, Health centre, Dispensary and clinics). The p-value ranges from less than 0.001 in 2007 to 0.013 in the year 2005 at 5 percent confidence level. The District hospital has the highest percentage followed by the dispensaries, clinics then finally the Health centres.

From figure 4.6, we see that on average, the incidence density of malaria in the whole of kilifi district was on the increase from 2005 to 2007. In 2005, there were two outlier values of 0.82 and 0.92.

Considering the incidences in 2005 per division as can be seen in figure 4.7, we see that divisions one and three had outlier values at 0.52 and 0.9 respectively. Divisions one and four did not have normal distribution of their data. Division five had the smallest interquartile range while division six had the biggest. Divison four had the highest median at 0.28 while division one had the smallest at 0.1.

In 2006, we see that the data was more normally distributed except for division one and six. Division four had the highest median 0.42 while division one had the lowest at 0.16. Division three had the biggest range of 0.81 while division six had the least.

In 2007 as can be seen from figure 4.9, the data was normally distributed in all the divisions apart from division six. Division two had two outliers while the rest divisions did not have any outliers. Division three had the biggest range of 0.7. Division six had a range of 0.35.

The intensity of the study area, which is the measure of number of points per square unit, is 46.3. This takes into account the total number of points mapped which were 37 health facilities.

The study also shows that there is clustering of the population around the big towns like Kilifi and Mariakani. The population is also clustered along the coast line and along the highway.

The K function, the J function and the G function, all give us a value less than one. This indicates that there is spatial clustering of malaria cases. F function measures the empty space between cases and this is also less than one which is indicative of spatial clustering. There is therefore no CSR. We therefore reject H_0 that there is complete spatial randomness in malaria cases in Kilifi district and conclude the alternative hypothesis that there is spatial clustering of malaria cases in kilifi district.

For the over five population, malaria clustering is seen in regions four and six of our map which are Chonyi and Vitengeni divisions respectively. The specific areas are around, Bwagamoyo, Lenga, Dida, Mgamboni, Ribe, Junju and Giryama dispensaries. We also see that the malaria cases seem to be more towards the coastline and reduces as one moves towards west of Kilifi district. This may be due to the increase in altitude as one moves away from the coastline. It may also be due to the reduced amount of rainfall towards west of the district (average of 400mm per year) as compared to the areas along the coastline which receives an average rainfall of about 1100mm.

In the case of the population under five years, the analysis shows that the clustering is more towards the south of the district as compared to the north of the district. The clustering is

much more than for the population over five years. The regions with the highest concentration of malaria cases are two, three and four which are Kikambala, Kaloleni and Chonyi divisions respectively. The specific locations include areas around Dida, Mryachakwe, Vitengeni, Kizingo, Giryama, Jibana, Makanzani, Gotani and Mgamboni. The same reasons for the clustering of the over five could also be the ones causing this type of outcome. As one moves away from the ocean, the cases reduces.

To therefore control malaria cases in these areas with clusters, the VHCs, The public health officers and the medical personnel should focus on these areas in their fight to reduce malaria cases.

Moran's I, a test of spatial autocorrelation gives us a value of 0.0242 for 2005, 0.0467 for the year 2006 and 0.0562 which are all less than two. This is indicative of lack of spatial autocorellation.

In model selection we see that the best model is the Non stationary poisson process as it has the smallest AIC. We can predict the intensity at a given point by fitting the equation

$$\lambda_0(x, y) = exp(-48.21 + 1.32x + 0.08y).$$

5.1 Short-comings and recommendations

This study used health facilities and divisions as the reference points for studying clustering. In order to get better results one can use area reference data and maybe considering the location or sub location area. The reasons for clustering in the above divisions were not also explored. Further studies should be carried out to explain the reasons for the pattern seen, i.e., the clustering that is reducing as one moves from south of the district towards the north, and also the reduced clustering as one move from the coastline towards the west of the district.

Chapter 6

References

1. A Spatial Statistical Approach to Malaria Mapping. I Kleinnschmidt, M Bagayoko, GPY Clark, M Craig and D Le Sueur. 29, 2000.

2. Predicting Malaria Infection in Gambian Children from Satellite Data and Bed Net Use Surveys: The importance of Spatial Correlation in the Interpretation of Results. Madeleine C. Thomson, Stephen J Connor, Umberto D'Alessandro et al. 1999, American Journal Medicine and Hygiene, pp. 2-8.

3. Spatial Clustering of Malaria and Associated Risk Factors During an Epidemic in highland Area of Western Kenya. Simon Brooker, Sian Clarke, Joseph Kiambo Njagi, Benson Estambale, Eric Muchiri et al. 2004, Tropical Medicine and International Health, pp. 757-766.

4. Different Malaria Control Activities in an Area of Liberia: Effects on Malariometric parameters. Bjorkman A, Hedman P, Wilcox M, Diamant I, Rombo L, et al. 1985, Ann Tropical Medicine Parasitiology, pp. 239-246.

5. Periodicity and Space-time Clustering of Severe Childhood Malaria on the Coast of Kenya. Snow RW, Armstrong-Schellenberg JRM, Peshu N, Forster D, Mwangi I, Waruiru C, et al. 1993, Trans R Soc Trop Med Hyg, pp. 386-390.

6. McCullagh P, Nelder JA. Generalised Linear Models. London : Chapman and Hall, 1989.

7. A Climate-based distribution model of malaria transmission in sub-saharan Africa. Craig MH, Snow, Le Sueur D. s.l. : Parasitology Today, 1999, Vol. 15.

8. PAHO. Status of malaria programmes in Americas (based on 2002 data). Report of the 44th directing council. Washington D.C : Pan American Health Organisation (PAHO)/ World Health Organisation (WHO), 2003.

9. The global distribution and population at risk of malaria: past, present and future. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. s.l. : Lancet Infectious Diseases, 2004, Vol. 4.

10. Earth Observation, Geographic Information Systems and Plasmodium falciparum malaria in Sub-Saharan Africa.S.I. Hay, J.A. Omumbo, M.H. Craig and R. W. Snow

11. Bloom, D.E. and Sachs, J.D. (1998). Geography, demography, and economic growth in Africa. Brookings Papers on Economic Activity, 207295.

12. Gill, C.A. (1921). The role of meteorology in malaria. Indian Journal of Medical Research 8, 633693.

13. Russell, P.F., West, L.S. and Manwell, R.D. (1946). Practical Malariology. Philadelphia: W.B. Saunders.

14. Gilles, H.M. (1993). Epidemiology of malaria. In: Bruce-Chwatts Essential Malariology (H.M. Gilles and D.A. Warrell, eds), 3rd edn, pp. 124163. London: Edward Arnold.

15. Christie, M. (1959). A critical review of the role of the immature stages of anopheline mosquitoes in the regulation of adult numbers, with particular reference to Anopheles gambiae. Tropical Diseases Bulletin 56, 385399.

16. White, G.B., Magayuka, S.A. and Boreham, P.F.L. (1972). Comparative studies on sibling species of the Anopheles gambiae Giles complex (Dipt., Culicidae): bionomics and vectorial

activity of Species A and Species B at Segera, Tanzania. Bulletin of Entomological Research 62, 295317.

17. Molineaux, L. and Gramiccia, G. (1980). The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa: Geneva, World Health Organization.

18. Kilian, A.H.D., Langi, P., Talisuna, A. and Kabagambe, G. (1999). Rainfall pattern, El Nio and malaria in Uganda. Transactions of the Royal Society of Tropical Medicine and Hygiene 93, 2223.

19. Hay, S.I., Rogers, D.J., Toomer, J.F. and Snow, R.W. (2000). Annual Plasmodium falciparum entomological inoculation rates (EIR) across Africa. I. Literature survey, internet access and review. Transactions of the Royal Society of Tropical Medicine and Hygiene 94, 113127.

20. Hay, S.I., Snow, R.W. and Rogers, D.J. (1998b). Prediction of malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Transactions of the Royal Society of Tropical Medicine and Hygiene 92, 1220.

21. Wernsdorfer, W.H. and McGregor, I. (1988). Malaria: Principles and Practice of Malariology. Oxford: Oxford University Press.

22. Molineaux, L. and Gramiccia, G. (1980). The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa: Geneva, World Health Organization.

23. Malakooti, M.A., Biomndo, K. and Shanks, D. (1998). Re-emergence of epidemic malaria in the highlands of western Kenya. Emerging Infectious Diseases 4, 671676.

24. Roberts, J.M.D. (1964). The control of epidemic malaria in the highlands of Western Kenya. Part II. The campaign. Journal of Tropical Medicine and Hygiene 67, 191199.

25. Cliff A, Ord J, (1981), Spatial Processes (Pion, London).

26. MoH Kenya (2006).National Guidelines for Diagnosis, Treatment and Prevention of Malaria for Health workers in Kenya).

27. Kitron U, Michael J, Swanson J and Haramis L (1997) Spatial analysis of the distribution of LaCrosse encephalitis in Illinois, using geographical information systems and local and global spatial statistics. American Journal of Tropical Medicine and Hygiene 57, 469475.

28. Morrison AC, Getis A, Santiago M, Rigau Perez JG and Reiter P (1998) Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico 19911992. American Journal of Tropical Medicine and Hygiene 58, 287298.

29. Fe'vre EM, Coleman PG, Odiit M, Magona JW, Welburn SC and Woolhouse MEJ (2001) The origins of a new Trypanosoma brucei rhodesiense sleeping sickness outbreak in eastern Uganda. The Lancet 358, 625628.

30. Schellenberg J, Newell JN, Snow RW et al. (1998) An analysis of the geographical distribution of severe malaria in children in Kilifi District, Kenya. International Journal of Epidemiology 27, 323329.

31. Chadee DD and Kitron U (1999) Spatial and temporal patterns of imported malaria cases and local transmission in Trinidad. American Journal of Tropical Medicine and Hygiene 61, 513517.

32. Ghebreyesus TA, Byass P, Witten KH et al. (2003) Appropriate tools and methods for tropical microepidemiology: a case-study of malaria clustering in Ethiopia. Ethiopian Journal of Health Development 17, 18.

Appendix

```
plot(c(-12,12),c(-12,12),type="n")
polygon(r2,border =1, col =(terrain.colors(12))[5])
polygon(r7,border =(terrain.colors(12))[1], col =(terrain.colors(12))[1])
polygon(r1,border =1, col =(terrain.colors(12))[1])
polygon(r3,border =1, col =(terrain.colors(12))[5])
polygon(r4,border =1, col =(terrain.colors(12))[6])
polygon(r5,border =1, col =(terrain.colors(12))[3])
polygon(r6,border =1, col =(terrain.colors(12))[6])
polygon(r8,border =1 , col ="white")
inputing the data.
kil1=read.table("F:/kilifiunder5yrs.csv",h=T,sep=",")
kil1
attach(kil1)
x=long
y=lat
plot(x,y) ploting the facility locations
library(spatstat)
data=ppp(x,y,c(39.2,40),c(-4.2,-3.2),marks=Prop07)
plot(data) ploting the cummulative incidence for the year 2007 for the under fives
summary(data)
plot(density(data,0.165))
plot(data,add=TRUE,cex=0.05)ploting the density of the points
contour(density(data,0.165),axes=FALSE)
data1=ppp(x,y,marks=Pop07,c(39.2,40),c(-4.2,-3.2))
summary(data1)
plot(data1)ploting the catchment population
```

```
plot(density(data1,0.165))

contour(density(data1,0.165),axes=FALSE)

Q=quadratcount(data,nx=4,ny=3)

plot(Q)ploting the quadrant count of the facilities

w=quadrat.test(data,nx=4,ny=3) Chi square testing for the difference in the number of points per quadrat

w

plot(data)

plot(w, add = TRUE, cex = 2)ploting the cummulative prevalence on the quadrant counts

plot(data)

plot(Q, add=TRUE, cex=2)ploting the catchment population on the quadrat count

k=Kest(data1) the k function for the data

plot(k)

hist(data1 x, nclass = 25) ploting the number of facilities per longitude.

data=(data1)

emp = distmap(data)

plot(emp, main = "Empty space distances")

plot(data, add = TRUE)

emp = distmap(data1)

plot(emp, main = "Empty space distances")

plot(data1, add = TRUE)

plot(data data=(data1)

plot(data)

Fc = Fest(data)

Fc

plot(Fc)plotting the f function

par(pty = "s")

plot(Fest(data))

plot(Fest(data), hazard r, main = "Hazard rate of F")
```

```
Gc = Gest(data)
Gc
plot(Gc)plotting the Gfunction
fit1=ppm(data1) getting the model of the observed data
summary(fit1)
AIC(fit1)
fit2=ppm(data1,x+y)Plotting a model with longitudes and latitudes
summary(fit2)
AIC(fit2)
fit4=ppm(data1,1, covariates = list(Prop07,Pop07))
AIC(fit4)
summary(fit4)
fit5=ppm(data1,1, covariates = list(mal05,Prop07,Pop07))
AIC(fit5)
fitnull=ppm(data1,1)
AIC(fitnull)
summary(fit5)
fit7=ppm(data1,x+y)
fit7
plot(fit7,how="image")
predict(fit7,type="trend")
predict(fit7,type="cif",ngrid=256)
coef(fit7)
vcov(fit7)
sqrt(diag(vcov(fit7)))
Under5pop=ppp(x,y,c(39.2,40),c(-4.2,-3.2),marks=div)
plot(Under5pop)
Malunder5prop07=ppp(x,y,c(39.2,40),c(-4.2,-3.2),marks=Prop07)
```

```
plot(Malunder5prop07)
summary(Malunder5prop07)
combined=ppp(x,y,c(39.2,40),c(-4.2,-3.2),marks=Prop07,Pop)
plot(combined)
summary(Under5pop)
plot(density(Under5pop,0.175))
plot(Under5pop,add=TRUE,cex=250)
plot(density(Malunder5prop07,0.165))
plot(Malunder5prop07,add=TRUE,cex=0.05)
plot(Under5pop,add=TRUE,cex=0.05)
contour(density(Malunder5prop07,0.165),axes=TRUE)
data3=ppp(x,y,marks=Prop06,c(39.2,40),c(-4.2,-3.2))
summary(data3)
plot(data3)
plot(density(data3,0.15))
contour(density(data3,0.15),axes=FALSE)
Q=quadratcount(data3,nx=4,ny=3)
plot(Q)
plot(X)
plot(Q, add = TRUE, cex = 2)
plot(density(data3,10))
plot(data3,add=TRUE)
j=Jest(data3)
plot(j)
g=Gest(data3)
plot(g)
plot(data)
plot(w, add = TRUE, cex = 2)
```

```
plot(data)

plot(Q, add=TRUE, cex=2)

plot(fit1,how="image")

predict(fit1,type="trend")

fit2=ppm(data1,x+y)

summary(fit2)

AIC(fit2) (for inhomogenous poison model with an an intensity cordinated)

predict(fit1,type="cif",ngrid=5)

coef(fit1)

fit8=ppm(data,slope,covariates=list(slope=typ))

fitnull=update(fit9,1)

anova(fitnull,fit9,test="Chi")
```