SPATIAL ANALYSIS OF CHOLERA IN KADUNA STATE, NIGERIA

BY

Zubaidah Yakubu BELLO

P13SCMT8059

DEPARTMENT OF STATISTICS,

FACULTY OF PHYSICAL SCIENCES,

AHMADU BELLO UNIVERSITY,

ZARIA

FEBRUARY, 2018

SPATIAL ANALYSIS OF CHOLERA IN KADUNA STATE, NIGERIA

BY

Zubaidah Yakubu BELLO

P13SCMT8059

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN STATISTICS

DEPARTMENT OF STATISTICS,

FACULTY OF PHYSICAL SCIENCES,

AHMADU BELLO UNIVERSITY,

ZARIA

FEBRUARY, 2018

DECLARATION

I declare that the work in this dissertation titled "SPATIAL ANALYSIS OF CHOLERAIN KADUNA STATE, NIGERIA" has been carried out by me in the Department of Statistics. The information derived from literature has been duly acknowledged in the text and a list of references provided. No part of this dissertation proposal was previously presented for another degree or diploma at this or any other institution.

_____ _____

Zubaidah Bello Yakubu Date

CERTIFICATION

This dissertation titled SPATIAL ANALYSIS OF CHOLERAIN KADUNA STATE, NIGERIAby Zubaidah Bello Yakubu meets the regulations governing the award of the degree of Master of Science of Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Prof. A. Isah	(signature)	Date
Chairman, Supervisory Committee		
Dr. H. G. Dikko	(signature)	Date
Member, Supervisory Committee		
External Examiner	(signature)	Date
Dr. H. G. Dikko	(signature)	Date
Head of Department		
Prof.Sadiq.Z.Abubakar	(signature)	Date
Dean, School of Postgraduate Studies		

ACKNOWLEDGEMENTS

Firstly, my gratitude goes to ALLAH (SWA) for sparing my life to start and finish this study. A big thanks to my humble supervisor Prof. A. Isah for his patience, motivation and tremendous support. My deepest and sincere gratitude to my dad and my husband for the warm-hearted encouragement throughout these journey. I consider it a deep honor.

Lastly, special thanks to my friends and family and every single person who helped out in one way or the other, I am most grateful.

DEDICATION

This work is dedicated to ALLAH (SWA) that has enriched my life in immeasurable ways and my family who have always supported my pursuit of knowledge.

ABSTRACT

The focus of this research is to analyze the spatial patterns and clusters of cholera epidemic in five local government areas(LGAs) of Kaduna State with the aid of Geospatial techniques. To analyze the spatial pattern and clusters of cholera epidemic in the study area, inventories provided by the Ministry of Health Kaduna State, coordinate of the locations of towns and localities were used for the analysis in ArcGIS 10.3 software environment and R Statistical software. The method employed in this study is Ripley's k-function method to determine spatial clusters of Cholera epidemic, analyzespatial pattern of cholera, identify disease clusters and risks using Geospatial Technology (GIS). The result shows that thereare about I87 locations of patients with Cholera disease cases in the study area from the year 2010 to 2015. Amongst the study area, Zaria Local Government Area, has the highest number of about 71 locations which constitutes about 37.97%, Igabi Local Government with 53 locations constituting about 28.34%, SabonGari Local Government with 19 locations constituting about 10.16%, Kaduna South Local Government having 28 locations constituting 14.97%, and Kaduna North with the lowest locations of about 16 patients (8.56%). It was found that there were about 1,363 reported cases of people with Cholera disease in Igabi LGA, being the area with the highest number and with 217 cases in Zaria which is the lowest. The result showed that four among the five local governmentswere clustered with the exception of Kaduna North for which the disease pattern is random. Moreover, the result revealed that the overall spatial pattern of cholera epidemic in the study area is clustered, and the overall cholera disease risk was more concentrated in Igabi and Kaduna South LGA.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Spatial statistics is the process of extracting or creating new information about a set of geographic features to perform routine examination, assessment, evaluation, analysis or modeling of data in a geographic area based on pre-established and computerized criteria and standards. It is a technique for analyzing spatial data mostly on human scale. Complex issues arise in spatial analysis, many of which are neither clearly defined nor completely resolved. The most fundamental of these are the problems of defining the spatial location of the entities being studied (Scott and Getis, 2008).

Nigeria is a prime area in studying spatial patterns associated with diseases because it is a country where millions of people live in close proximity not only to other people but also to open and unsafe water sources and refuse dumps. It is also a country that is actively engaged in alteration of its aquatic ecosystems, a process often associated with changed disease ecologies (WHO 1993). Cholera is one of the deadliest diseases in Africa (WHO 1993), within 2-3 hours of unset symptoms, a previously healthy person may severely become dehydrated and if not treated may die within 24hours (Sack *et al* 2004; WHO 2010). A link between cholera, phytoplankton blooms and copepod zooplankton has been demonstrated in Asia (Colwell *et al.*, 1996). The African Great Lakes have been suspected to play a role as reservoirs of the bacteria Vibrio cholerae (v.cholerea), while human infection and movement are probably involved in the propagation of the disease inland. During the 19th century, cholera spread repeatedly from its original reservoir or source in the Ganges delta in India to the rest of the world, before receding to South Asia (WHO 2008). Six pandemics were recorded that killed millions of people across

Europe, Africa and the Americas. The seventh pandemic, which is still on going, started in 1961 in South Asia, reached Africa in 1971 and the Americas in 1991. The disease is now considered to be endemic in many countries and the pathogen causing cholera cannot currently be eliminated from the environment, WHO(2008). Regions of the world where cholera is currently prevalent are Africa, Asia and parts of the Middle East. Imported cases occasionally occur in richer countries in travellers returning from endemic areas, (National Travel Health Network and Centre (2007). The disease no longer poses a threat to countries with minimum standards of hygiene, but it remains a challenge to countries where access to safe drinking water and adequate sanitation cannot be guaranteed, (WHO, 2009). The mechanistic basis for a climate-cholera connection involves multiple pathways and the primary transmission from environmental reservoirs initiates seasonal outbreaks of cholera in endemic regions, (Pascual*et al*, 2002).

GIS and epidemiological approaches are helpful tools to control the disease spatially and temporally. GIS is a computer system for capturing, storing, querying, analyzing, and displaying geospatial data (Chang, 2008). The general functions of GIS in health studies are disease mapping and modeling, spatial andspatio-temporal changes analysis and risk assessment, public health care and hospital management. GIS is an integrated collection of computer software and data used to view and manage information connected with specific locations, analyze spatial relationship and model spatial processes. GIS has the capabilities of analyzing the spatial patterns and distribution of disease and its influential environments towards creating an innovative cholera control plan in the country. Spatial epidemiology is an essential approach in understanding of spatial disease risk transmission and pattern particularly disease mapping and descriptive analysis (Chin-Lai, 2009). This study explores spatial analysis toolset in R statistical package, ArcGIS and Excel software for cholera mapping and pattern analysis in Kaduna State.

1.2 Statement of the Problem

The threat of cholera rampaging through Nigeria has been a major concern in the country. Recent global health reports show a continual vulnerability of large populations to infectious diseases in relation to our environments. Infectious disease are complex to control and prevent, leading to questions on how to combat them through novel and creative solutions. Cholera is an epidemic and infectious disease which is of global and public health significance.

Thus, it remains a global threat especially in countries where access to clean safe drinking water and sufficient sanitation cannot be assured. The disease has been a public health burden in Nigeria. Although, health scientist and researchers have embraced the use of spatial data in determining the risk of diseases, little has been done to increase the uptake of spatial techniques in health sciences (Grahamm*et al* ;2004). These research focuses on the use of spatial statistical method to identify the disease pattern and disease risk and Geo-statistical method to map out the disease.

1.3 Significance of the Study

Cholera has been a great concern generally and has impacted negatively on the economies of developing nations. Health workers generally are unable to identify high or risk areas in the areas they operate so as to tailor interventions and do effective health monitoring. Research conducted so far by medical and climate professionals have either lacked knowledge or showed variations on the climate conditions that accompanied the transmission of cholera. The geographical distribution of any major disease forms an important basis for locating appropriate interventions for its control and a means to monitoring their effectiveness. It also provides a possibility for identifying ecological factors with which the disease may be associated.

The link between climate and medical data has not been well defined, and health information system have been weak due to the lack of case detection, irregularity in reporting and poor coordination (WHO, 2009).There is the need to draw more attention to hot spots and areas where intervention measures can be tailored to improve monitoring of the occurrence, distribution and control of cholera in different geographical areas and time periods. This can be used by health workers or officials to make appropriate planning and resource allocation in limiting the outbreak. Spatial statistical method is also necessary to analyze the pattern that are associated with the spatial distribution of cholera at different geographical locations in the study area.

1.4 Aim and Objectives

The aim of this study is todetermine the spatial heterogeneity of cholera epidemic in Kaduna state.The specific objectives through which the stated aim will be achievedareto:

(i) Generate eminent spatial patterns in the data set through spatial map.

(ii) Evaluate the degree of spatial clusters of disease locationsand disease risks.

(iii) Create an inventive cholera control plan.

1.5 Definition of Terms

1. **Spatial Analysis**-Technique for analyzing spatial data or human scale mostly geographic data or techniques applied to structures at human scale, most notably in an analysis of geographic data.
2. **Epidemic**- Rapid spread of infectious disease to a large number of people in a given population within a short period of time, or an outbreak or unusually high occurrence of disease or illness in a populated area.

3. **Epidemiology**- The study of the patterns , causes and effects of health and disease condition in a defined population.
4. **Cholera**- Is an acute diarrhea infection caused by ingestion of food or water contaminated with the bacterium vibrio cholera.
5. **Disease**- Can be seen as a disorder from the normal healthy state of the body, soul and mind of human being, which manifest is itself in an abnormal development of the physical, physiological and mental state of the human being concerned.
6. **Threat**- An expression of an intention to inflict pain, injury, evil or punishment. An indication of impending danger or harm.
7. **Contamination**- Is the term describing the state of a person or material on coming in contact with disease pathogen.
8. **Transmission**-The act or process of transferring a disease from one person to another'
9. **Outbreak**- A sudden, violent or spontaneous occurrence especially of disease or strife.
10. **Infection**- The invasion of body tissues by disease causing microorganism, their multiplication and the reaction of body tissues to these microorganisms and the toxins that they produce.
11. **Spatial Dependency**- is a key concept on understanding and analysing spatial phenomena. Such notion stems from what Waldo Tobler calls the first law of geography: "everything is related to everything else, but near things are more related than distant things.''
12. **Risk Factor**- Any attribute, characteristic or exposure of an individual that increases the likelihood of developing a disease or injury.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, an attempt was made to review the related literature relevant to the current research that will enable us to put in proper perspective, what has been done and what needs to be accomplished .

2.2. Conceptual Review literature

This section presents the conceptual literature review on spatial analysis of cholera epidemics.

2.2.1 Cholera

Cholera is an infection of the small intestine that is caused by bacterium Vibrio cholera 01 and 0139 (Riyan 2004 and WHO 2010). Cholera pandemic started in 1961, reaching West Africa and Nigeria in the late 1970.The first recorded cases of cholera in Nigeria occurred in a village near Lagos in December 1970 leading to an important epidemic of 22,931 cases and 2945 deaths with a case fatality rate (CFR) of 12.8% during 1971.Between 1972 and 1990, Nigeria reported few cases. In 1991, 59,478 cases were reported with the death toll of 7,654. The CFR was 12.9% ,the highest rate reported about the country to date. Cholera is endemic in Nigeria (Falade and Lawoyin 1999) and epidemiological features (Utsalo et al 1991; Eko et al 1994; Hutin 2003) have been reported from various parts of the country with investigation on possible sources of the outbreak. Cases started to be registered in January 1991 and among the first affected states were Kano, Akwa-Ibom, Bauchi, Niger and Oyo. By September of the same year the disease have spread to 19 of 21 states including the Federal Capital.Nigeria reported an outbreak of cholera in May 2013, and up until October 2014, a total of 40,608 suspected cholera cases were reported. With 898 deaths, that gave a case fatality rate (CFR) of 1.95%.

The outbreak has experienced a strong upsurge since early 2014, with more than 34,000 cases and 664 deaths reported betweenJanuary 2014 to October 2014. Suspected cholera cases were recorded in 19 of the country's 36 states (51%). Nine of the affectedstates include; Bauchi, Kaduna, Plateau, Kano, Borno, Adamawa, Katsina, Kebbi and Zamfara. Investigations of cholera outbreak in Nigeria have focused on epidemiological features, the probable source of contamination and the risk factors without spatial linkage of health data. However, advances in Geographical Information System (GIS) provides this opportunity and have become an indispensable tool for processing, analysing and visualizing spatial data with domains of environmental health, disease ecology and public health (Kristemann et al 2002). The World Health Organization (WHO) is continuing to support the ministries of health and health authorities in Local Government Areas (LGA's) as they respond to outbreaks.

Furthermore, the future of cholera prevention as well as any other infectious disease lies in the ability to develop effective and accurate climatic models which could serve as early warning systems for infectious disease outbreaks. This shift in containment to prevention in public health measures represents a positive step in the right direction towards minimizing the death rate during outbreaks such as the ones in Zimbabwe.

2.2.2 Socioeconomic and Demographic Risk Factors

Socioeconomic and demographic factors have been reported to significantly enhance the vulnerability of a population to infection and contribute to epidemic spread (Ali *et al* 2002a, 2002b; Borroto and Martinez-Piedra 2000; Sasaki *et al* 2008). Such factors also mandate the extent to which the disease will reach epidemic proportions (Miller 1985; Emch 1999) and also modulate the size of the epidemic (Pascual*et al* 2002, 2006; Hartley *et al* 2005). Population risk factors of cholera include; poverty, lack of development, high population density, low education

and lack of previous exposures (Ackers*et al* 1998; Ali *et al* 2002). The synergy of poverty, high population, density, poor sanitation, poor housing and lack of good water supplies enhance exposure to pathogenetic cholera vibrios. In epidemic prone regions like Africa, cholera outbreaks have been linked to multiple environmental and socioeconomic sources (Acosta *et al* 2001; Sharpiro*et al* 1999).

Cholera diffuses rapidly in environments that lack basic infrastructure with regard to access to safe water and proper sanitation. The cholera vibrios can survive and multiply outside the human body and can spread rapidly in environments where living conditions are overcrowded and where there is no safe disposal of solid waste, liquid waste, and human faeces (Ali et al 2002a,2002b). Root (1997) and Siddique*et al* (1996) have reported that increase in population density can strain sanitation systems, thus putting people at increased risk of contracting cholera.

2.2.3 Factors Influencing Spread of Cholera

Cholera germs are passed in the stools of infected persons and therefore it is widely spread by consuming food or water which has been contaminated by fecal waste/stools. This happens more often in developing countries, and underdeveloped countries the lack clean water supplies for drinking and proper sewage disposal systems and also, practice poor sanitation and poor food hygiene. Once cholera is introduced to a population in a specific location, complex factors may lead to prolonged transmission. Socioeconomic, environmental, demographic and climatic factors enhance the vulnerability of a population to infection and contribute to the epidemic spread of cholera (Acosta *et al* 2001; Hutin *et al* 2003; Reller 2001; Tauxe 1998; St Louis 1990). A hint of these socioeconomic, environment, demographic and climatic factors is as follows;

Poor sanitation;Overcrowding/high population;Lack of clean drinking water; unsafe water supply;Proximity and density of refuse dump;High poverty and low income level;Climatic conditions, direct correlation between cholera and sea surface temperature, Proximity to surface water sources, contaminated drinking water bodies, High migration ;Poor personal hygiene and Poor cooking practice

2.2.4 Spatial Epidemiology

Spatial epidemiology is concerned with the description and analysis of the Geographic, or spatial variations in disease with respect to demographic, environmental, behavioral, socioeconomic, genetic and infectious risk factors. It is the study of spatial distribution of disease incidences and its relationship to potential risk factors. The spread of infectious disease is closely associated with the concepts of spatial and temporal proximity, as individuals who are linked in a spatial and temporal sense are at risk of getting infected, therefore proximity to environmental risk factor is important.

The main objectives of spatial epidemiology analysis are the description of spatial patterns, identification of disease clusters and explanation of disease risk. Geographic data systems include geo-referenced data and attributes,and they point the areas with spatial location. These data are obtained by field surveys, remote sensed imagery or use of existing data generated either by government or organizations or agencies of government such as cadastral, metiological or national census statistics and health organizations.

2.2.5 Statistical Methods for Spatial Epidemiology

The origin of spatial epidemiology dates back to 1855 with the classic epidemiologist of John Snow on cholera transmission. Spatial epidemiology generally comprises of at least three types of study focus, namely; disease mapping, disease clustering and ecological analysis

Disease Mapping: Disease mapping have played a key descriptive role in spatial epidemiology. Disease mapping provides information on a measure of disease occurrence across a geographic space. Disease maps are useful in suggesting hypotheses for further investigation or as part of general health surveillance and the monitoring of health problems. A famous example is the classical epidemiology of John Snow (1855),Mapping of locations of cholera victims, Snow (1972) was able to trace the cause of the disease to a contaminated water source. Surprisingly, this was done 20 years before Koch and Pasteur established the beginning of microbiology (Koch 1884). Disease mapping has long been in the form of plotting the observed disease cases of prevalence. Borronto and Maritnez-Piedra (2000) used Geographical Information System (GIS) to map cumulative incidence rates of cholera in 32 Mexican states.

Chevallier*et al.*, (2004) used cartographic representation of cholera incidence rates to study the spatial distribution of cholera in Ecuador. Raw disease rates yields less precise estimates for small population and vice versa, hence, mapping the raw estimate of disease occurrence can lead to spurious spatial features.

Lawson and Clark (2001, 2002) provide recent views of current appropriate disease mapping methods. Several statistical smoothing techniques have been proposed to filter out the noise (rate variation) caused by population variability (Hansen, 1991), spatial filtering (Lolonis and Rushton 1996), and geo-statistical methods (Oliver *et al* 1998, Goovaert and Jacquez 2004)

Disease mapping are able to provide us a rapid visual summary of complex geographic information. These maps may also identify subtle patterns in epidemic/ health data that are sometimes missed in tabular presentation. Disease mapping aims include;

1. Simple description by showing or displaying a visual summary of geographical risk, for example, the map of Snow (1972).
2. Provide estimates of risk by area to inform public health resource allocation.
3. Hypothesis generation by giving clues to causes of disease and or factors that influenced spread by informal or formal examination of maps.

Geographic Correlation: A significant interest in spatial epidemiology also lies in identifying associated risk factors which enhance the risk of infection called ecological analysis (Lawson *et al* 1999) or geographic correlations (Elliot *et al* 2000).Geographic correlation studies examines geographic disparities across inhabitants in an exposure to environmental variables which may be measured in air, water, or soil. Correlation studies aims at;

1. Examination of the association between disease outcome and explanatory variables, in a spatial setting using regression models.
2. Modeling approaches such as logistic for point data and log-linear models for count data. Correlation studies deals with the association between disease risk and exposures of interest.
3. Hypothesis generation by giving clues to the causes of disease and or factors that influence spread of formal and informal examination maps. The formal is carried out via spatial regression. Geographical weighted regression (GWS) is appropriate for the

assessment of the spatial heterogeneity in the estimated relationship between independent and dependent variables.

Clustering: Cluster analysis provides opportunities for the epidemiologist to understand possible association between demographic and environmental exposures and the spatial distribution of disease (Kulldorff and Nagarwala 1997). Searching for disease clustering involves an assessment of global and local accumulation disease incidence (Lawson *et al*2002; Tango 2010). The focus of global cluster analysis is to determine the presence and absence of clustering in the whole study region. There are numerous methods for testing global clustering including those proposed by Alt and Vach (1991), Tango (1995; 1999; 2000). But the most widely used method for global clustering in epidemiology was proposed by Moran which includes k-nearest neighbours test (Cuzick and Edwards, 1990), Moran's index (Borronto and Martinez 2000), Ripley k-index (Ruiz-moreno*et al* 2007). The main aim is to investigate disease clusters and disease incidence near a point source.

2.2.5 Spatial Analytical Methods and Geographic Information Systems in Epidemiology

Matthew (1990) pointed out that person, place and time are the basic elements of Cholera outbreak investigations and epidemiology. Historically, however, the focus in epidemiologic research has been on person and time, with little regard for the implications of place or space and the pattern even though disease mapping has been done for over a hundred years. The development of geographic information systems (GISs) over the last 20 years has provided a more powerful and rapid ability to examine spatial patterns and processes. This, in turn, has fostered the discussion of such policyrelevant issues as health services and planning (Matthew, 1990) as well as the use of GISs for epidemiologic investigations and disease surveillance.

According to Gesler (1986), GIS when combined with spatial analytical methods, may be helpful in the study of health care and health care delivery. The literature on the geography of health care can be divided into three areas of research: the spatial properties of delivery systems and their accessibility, the concomitant utilization and planning of health care services, and the spatial structure of disease patterns in both static and dynamic form (Keams, 1993). In the last decade, computerization of spatial data, through the use of GIS, has emerged as a tool for health care research and epidemiology (Gesler, 1986)

Scott and Getis (2008) found that spatial statistics comprises a set of techniques for describing and modeling spatial data. In many ways, they extend what the mind and eyes do, intuitively, to assess spatial patterns, distributions, trends, processes and relationships. Unlike traditional (non-spatial) statistical techniques, spatial statistical techniques actually use space – area, length, proximity, orientation, or spatial relationships – directly in their mathematics (Scott and Getis 2008).

2.2.6Analyzing patterns

The analyzing Patterns toolset (Table 2.1) contains methods that are most appropriate for understanding broad spatial patterns and trends in GIS (Mitchell 2005). With these tools the following questions can be answered;

i. Which place is cholera epidemic concentrated most?
ii. Does the spatial pattern of the disease mirror the spatial pattern of the population at risk?
iii. Is there an unexpected spike in pharmaceutical purchases; and
iv. Are new Cholera cases remaining geographically fixed?

Table 2.1: A summary of the GIS tools used in the analyzing patterns

Tool	**Description**
Average nearest neighbor	Calculates the average distance from every feature to its nearestneighbor based on feature centroids
High/low clustering (Getis-Ordgeneral *G*)	Measures concentrations of high or low values for a study area
Spatial autocorrelation (globalMoran's *I*)	Measures spatial autocorrelation (clustering or dispersion) basedon feature locations and attribute values
Multi-distance spatial clusteranalysis (Ripley's *K* function)	Assesses spatial clustering/dispersion for a set of geographic featuresover a range of distances

Source: Adopted from Lauren, and Mark, (2010).

The Ripley' *K* function is a unique tool that looks at the spatial clustering or dispersion of points/features at a series of distances or spatial scales. The dark diagonal line represents the expected pattern, if the features were randomly distributed within the study area. The *X* axis reflects increasing distances. The solid curved line represents the observed spatial pattern for the features being analyzed.

When the curved line goes above the diagonal line, the pattern is more clustered at that distance than we would expect with a random pattern; when the curved line goes below the diagonal line, the pattern is more dispersed than expected. Based on a user-specified number of randomly

generated permutations of the input features, the tool also computes a confidence envelope around the expected line. When the curved line is outside the confidence envelope, the clustering or dispersion is statistically significant.

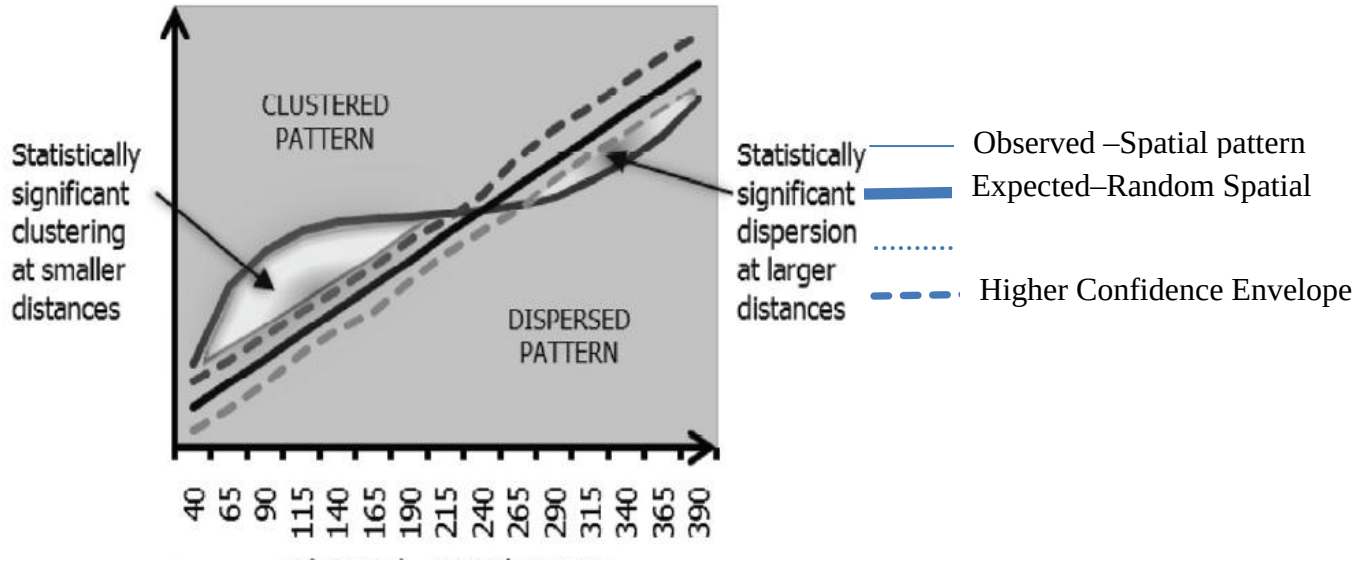

Figure 2.1: Components of the *K* Function graphical output

Source: Adopted from Lauren, and Mark, (2010).

The Ripley's*K* function is useful for comparing different sets of features within the same study area, such as two strains of a disease or disease cases in relation to population at risk. Similar observed spatial patterns suggest similar factors or rather similar spatial processes are at work. A researcher might compare the spatial pattern for a disease outbreak, for example, to the spatial pattern of the population at risk to help determine if factors other than the spatial distribution of population are promoting disease incidents. Wheeler (2007) and Illian*et al.*, (2008) provide examples of additional applications for the tools in the Analyzing Patterns toolset.

Several studies were tested for global clustering to determine the existence of clustering in a studyarea without pinpointing specific locations. The Tests most frequently used in the selectedliterature are Diggle and Chetwynd's bivariate K-function (Bernstein, *et al*. 2004), Mantel-Bailar's test (D'Aignaux, *et al.* 2002), and the Potthoff-Whittinghill method (PW) (Schmiedel, *et al*. 2010). The K-function is a preferred methodbecause it corrects for edge effects and allows for a range of spatial and temporal scales (Diggle, *et al.*, 1995; Houben ,*et al.* 2006; McNally, *et al.* 2008).

2.2.7 Mapping clusters

The tools discussed in *K* function inAnalyzing Patterns toolset are global statistics that answer the question: *Is* there statistically significant spatial clustering or dispersion? Tools in the Mapping Clusters toolset (Table 2.2), on the other hand, identify *where* spatial clustering occurs, and *where* spatial outliers are located:

i. Where are the sharp boundaries of cholera epidemics?

ii. Where do we find anomalous spending patterns in the study area?

iii. Where do we see unexpectedly high rates of cholera epidemics?

Table 2.2: A summary of the tools in the mapping clusters toolset

Tool	**Description**
Cluster and outlier analysis(Anselin's local Moran's *I*)	Given a set of weighted features, identifies clusters of high orlow values as well as spatial outliers
Hot spot analysis (Getis-Ord*i* $G*$)	Given a set of weighted features, identifies clusters of featureswith high values (hot spots) and clusters of features with lowvalues (cold spots)

Used in 36 studies, spatial cluster methods are the most common tool for assessing nonrandom spatial patterns. Many statistical methods have been developed to determine if disease clusters are of sufficient geographic size and concentration to have not occurred bychance (Knox, 1989;Kulldorff, 2006). Global clustering tests evaluate without pinpointing the specific locations ofclusters, whereas local clustering tests specific small-scale clusters and focused clustering assesses clustering around a prefixed point source such as a nuclear installation. These testshave different substantive interpretations and detection methods; the presence of one doesnot imply the presence of the other (Huang and Pickle, 2008). Overall, in this literature, a total of 35 studies used spatial cluster detection methods. Such methods have had to overcome limitations tostatistical power, including the availability of few cases, high variability in the background population density, multiple testing, and size and shape of cluster windows.

Ajoke, *et al.* (2012) carried out a research onCholera Epidemiology in Nigeria: an overview. In Nigeria, since the first appearance of epidemic cholera in 1972, intermittent outbreaks have been

occurring. The later part of 2010 was marked with severe outbreak which started from the northern part of Nigeria, spreading to the other parts and involving approximately 3,000 cases and 781 deaths. Sporadic cases have also been reported. Although epidemiologic surveillance constitutes an important component of the public health response, publicly available surveillance data from Nigeria have been relatively limited to date.

Akyala, *et al.* (2014) investigated Cholera Outbreak in an Urban North Central Nigerian Community-The Akwanga Experience. The method employed includes; a descriptive, active case search and un-matched case control study. Univariate and bivariate analysis using Epi-Info version 3.3 was also employed. The Results show that out of 18 cases patients, 10(55.6%) were male while 8(44.4% were female, of which 40% are from a sub-urban community of KurmiTagwaye and the attack rate was 2/1000 population with two cases fatality. The Age ranged from 1-84 years: mean (34+18) years, age group of 20-29 years were mostly affected with vibrio cholera serotype Ogawa was isolated from stool. The main water source, RafinKurmiTagwaye River was polluted by residents defecation, post-defecation bath and car washing compared to controls, and the case patients were likely to have drank from the river (or 4.56, 95% CL.2.75-18).

According to UNICEF (2015), the number of cholera cases is increasing in the last 3 weeks (w35-37) in the two most active basins (Congo River and Lake Chad Basins). Cholera is spreading through camps sheltering displaced people in Maiduguri, capital of northern Nigeria's Borno state. In the five days up to 15 September, 187 patients were admitted to MSF's centre, two-thirds of them in a serious state.

An outbreak was also reported in Zaria in Kaduna state with 192 cases recorded in two weeks in 2015 (UNICEF, 2015). The call for vigilance in the neighboring countries around Lake Chad has to be maintained. Ministries and agencies involved in the fight against cholera must pursue measures to enhance cross-border cooperation. In early 2015, 13 of 36 states recorded cholera cases, with Anambra, Kano, Rivers and Ebonyi states being the worst affected; during the end of April 2015, 2,108 cases had been reported, with 97 deaths with the CFR rate rising to 4.76%, causing extreme concern (International Federation of Red Cross (IFRC, 2016). According to WHO (2013), the cholera outbreaks are attributed due the fact that Cholera is most likely to be found and spread in places with inadequate water treatment, poor sanitation, and inadequate hygiene. Therefore, cholera outbreaks have been occurring in developing countries for example, Iraq (2007- 2008), Guinea Bissau (2008), Zimbabwe (2008-2009), Haiti (2010),Democratic Republic of Congo (2011-2012), and Sierra Leone (2012).

2.3 Empirical Review of literature

Okabe et al (1992) opined that Perhaps the most important developments in recent years are the applications of K-function analysis to the study of point patterns, and the use of Voronoi polygons to study spatial tessellations. The K-function is the ratio of the sum of all pairs of points within a pre-specified distance, d, of all points to the sum of all pairs of points regardless of distance. The function is adjusted to take into account distances that are closer to the boundary of the study area than to d. The original K-function by Ripley (1977) was modified by Besag (1977) to take into account the need to stabilize variance, and Getis (1984) generalized the formula to include the weighting of points, such that the sum of pairs of points became the sum of the multiples of the weights associated with each member of a pair of points. Diggle (1983) has done much to exploit this formulation to show many new features of patterns. For example,

not only can one easily show the difference between an existing pattern and a random pattern but one can also develop theoretical expectations for other than random patterns. In addition, patterns divided into different point types (marked patterns) can be studied easily. For testing purposes, a confidence envelope of possible outcomes under the hypothesis of say, randomness, is usually constructed by means of a Monte Carlo simulation. Studies of the spatial distribution of vegetation dominate the empirical literature of K-function analysis (Diggle 1983), but the method has been used for the study of human population distribution (Getis 1983) and disease distribution (Morrison et al 1996).Gatrell et al (1996) showed that the K-function can be used as an indicator of time-space clustering, that is, one simultaneously finds pairs of points separated by designated units of time and distances in space. This approach is particularly useful for identifying disease clustering over time.

Diggle*et al* (1996) wrote on Spatial point pattern analysis and its application in geographical epidemiology. This paper reviews a number of methods for the exploration and modelling of spatial point patterns with particular reference to geographical epidemiology (the geographical incidence of disease). Such methods go well beyond the conventional 'nearest-neighbour' and 'quadrat' analyses which have little to offer in an epidemiological context because they fail to allow for spatial variation in population density. Correction for this is essential if the aim is to assess the evidence for 'clustering' of cases of disease. Diggle et al (1996) examine methods for exploring spatial variation in disease risk, spatial and space-time clustering, and they consider methods for modelling the raised incidence of disease around suspected point sources of pollution. All methods are illustrated by reference to recent case studies including child cancer incidence, Burkett's lymphoma, cancer of the larynx and childhood asthma. An Appendix considers a range of possible software environments within which to apply these methods. The

links to modern geographical information systems are discussed. The study stresses that as far as the K function for a CSR process is concerned, the important point is that the probability of the occurrence of an event at any point in R is independent of what other events have occurred and is equally likely over the whole of R. Thus, for a homogeneous process with no spatial dependence, the expected number of events within a distance d of a randomly chosen event is simply Kernel estimation and, in particular, the use of ratios of kernel estimates was suggested as a means of assessing spatial variation in disease risk. They suggested that second-order properties may be characterized by the K function and that this could be extended to consider different types of event. This led us to consider the important topic of how we could detect spatial clustering in the presence of environmental heterogeneity. They also examined the use of K functions in a test for space-time clustering. Next, they presented two variants of a model in order to assess whether there is evidence for raised incidence of one type of event in the vicinity of a fixed site, such as a potential source of environmental contamination.

Hinman (2006) investigated Spatial and Temporal Structure of Typhoid Fever in Washington. The study include dot maps of typhoid case locations for the entire 1895 epidemic, at two-week intervals for 1906, 1907, and 1908, and monthly intervals for 1909. These point locations are used to construct a Geographic Information System (GIS) displaying the spatial distribution of individual typhoid cases. The creation of this GIS allows for the investigation of urban typhoid at a localized geographic scale. The temporal resolution of the data and supplementary data included in the reports provides an opportunity to explore urban typhoid within yearsand , between years, to compare morbidity to mortality, and to compare the spatial pattern of multiple diseases. This dissertation describes the creation of this GIS and the results of the spatial analyses. Ripley's K-function and the Gi* statistic were used to evaluate spatial clustering

patterns. The findings of this dissertation indicate that typhoid in early twentieth century Washington, D.C. originated from multiple sources whose impact decreased over time. Studies of this type make use of geospatial approaches unavailable when the original data were collected, in order to investigate potential patterns of typhoid fever invisible a century ago. This research helps to provide a better understanding of the historical geography of urban health in general. The Ripley's K-function was used to measure global spatial auto-correlation with the original point data for each year (summer 1906, 1907, 1908, 1909; whole year 1909).The Ripley's K-function was conducted using ClusterSeer2. Ripley's K-function identified global clustering in the typhoid case data, in contrast to the findings of the original PHS reports. Recent advances in local spatial auto-correlation 98 techniques allowed this study to go beyond a global investigation and explore the possibility of local clusters which are hard to distinguish within the complete spatial distribution of cases. The Getis and Ord statistic indicates that clustering occurs at multiple spatial scales which refute the original PHS conclusions that typhoid's distribution was evenly distributed.

Siddiqui et al., (2006) applied Cuzick-Edward's k-Nearest Neighbors test (Cuzick and Edwards, 1990) to evaluate clustering of cholera cases in Pakistan. They found that in July 2002 and June 2003, cholera outbreaks were detected by a diarrhea surveillance system in a village outside Karachi. The result showed that the diversity of statistical analysis techniques applied to epidemiological data in scientific articles is growing, confirming the need for interaction between different areas of knowledge, such as statistics, geography, and the health sciences but still less attention is put towards the used statistics without the use of geospatial techniques. Despite the growing number of studies with a characteristic element of spatial analysis, the

application of the techniques and its continuity in epidemiological studies requires careful evaluation.

Frank and Alfred (2008) conducted a research on spatial and demographic patterns of Cholera in Ashanti region –Ghana. A GIS based spatial analysis and statistical analyses were carried out to determine clustering of cholera using the clustering and outlier analysis (Moran's I). The result showed that high cholera rates were clustered around Kumasi Metropolis (the central part of the region), with Moran's Index = 0.271 and $P < 0.001$.

Diego et al., (2010) examine the spatial clustering of the spatial structure of cholera outbreaks using Ripley's K and L indices and bootstrapping methods to evaluate the occurrence of spatial clustering in the cases during outbreaks using different temporal windows. The spatial location of cases was also against the spatial location of water sources. The results showed that spatial clustering of cholera cases was detected at different temporal and spatial scales. Cases relative to water sources also exhibited spatial clustering.

Shittu et al (2010), assessed the epidemiological features of a GIS supported investigation of a cholera outbreak in Abeokuta. The epidemiological surveillance data showed a total of one hundred and fifteen cases and 11 deaths with case fatality rate of 9.6%. This study had shown that inadequately maintained pipes and clandestine connections are major problems that need to be addressed in the area with pipe borne water in order to prevent future outbreaks of waterborne diseases.

Steven, et al. (2011) researched on Geographic profiling as a novel spatial tool for targeting infectious disease control. The data was re-analyzed from a classic epidemiological studyof the 1854 London cholera outbreak. Using 321 disease sites as input and evaluated the locations of 13

neighborhood water pumps. The Broad Street pump -the outbreak's source- ranks first and was situated in the top 0.2% of the geoprofile. Geographic profiling ranks six of these seven sites in positions 1-6, all in the top 2% of the geoprofile. In both analyses, the method outperformed other measures of spatial central tendency. They suggest that geographic profiling could form a useful component of integrated control strategies relating to a wide variety of infectious diseases, since evidence-based targeting of interventions is more efficient, environmentally friendly and cost-effective than untargeted intervention.

Paul, et al. (2011) conducted a research on spatio-temporal epidemiology of the cholera outbreak in Papua New Guinea in 2009-2011. The spatio-temporal distribution and clustering of the Papua New Guinea cholera outbreak was analyzed using the national cholera database. The Kulldorff space-time permutation scan statistic contained in the software package SatScan v9.2 was used to describe the first 8 weeks of the outbreak in Morobe Province before cholera cases spread throughout other regions of the country. Data were aggregated at the provincial level to describe the spread of the disease to other affected provinces. The results of the spatio-temporal and cluster analyses revealed that the outbreak was characterized by three distinct phases punctuated by explosive propagation of cases when the outbreak spread to a new region. The lack of road networks across mostparts of Papua New Guinea is was said to have had a major influence on the slow spread of the disease during the outbreak.

Based on the literatures reviewed, this study focuses on applying the Ripley's k-function method on cholera data in Kaduna state to show the simple description of the visual summary of the disease risk.

CHAPTER THREE

METHODOLOGY

3.1 The Study Area

The study area is Kadunastate Nigeria whichand liesbetween latitudes $10^{o}20'0"N$ and $11^{o}10'0"N$ of the equator and between longitudes 7°10‘0”E and 8°0’0”E East of Greenwich Meridian. The state shares boundary with Zamfara, Katsina , Kano in the North/West, Bauchi, Plateau in the North/ East and Niger State in the Country. The statealso occupies an area of about $260km^2$; the distant between the Eastern and Western limits of Kaduna is approximately 13.7 km. The state has a total population of 6,066,526 (NPC,2006). The figure has changed now because of increase in population as Kaduna is an administrative centre in the Northern part of the country. The Population is projected as seven million, two hundred and five thousand, three hundred and fifty four. The major tribes in kaduna are mainly Hausa, Gwari and other people of different languages.Five Local Government Areas namely Igabi, KadunaNorth, KadunaSouth, Sabongari and Zaria.

3.1.1Climate

Kaduna experiences tropical continental climate (A_w), of koppen classification. The A_w climate is characterised by wet and dry season in a year.The wet season lasts for six months (May to October) and is dominated by conventional heavy rainfalls with intensities of up to 120mm/hr. The wet season isthe period when the occurrences of the cholera diseases used to be at its peak in the study area. The mean annual rainfall is about 120cm. The rainfall peaks up in July/August.The monthly temperature is between 21^0C and 32^0C, while the annual temperature rangesbetween 11 and 13^0C is typical Coker(2000).

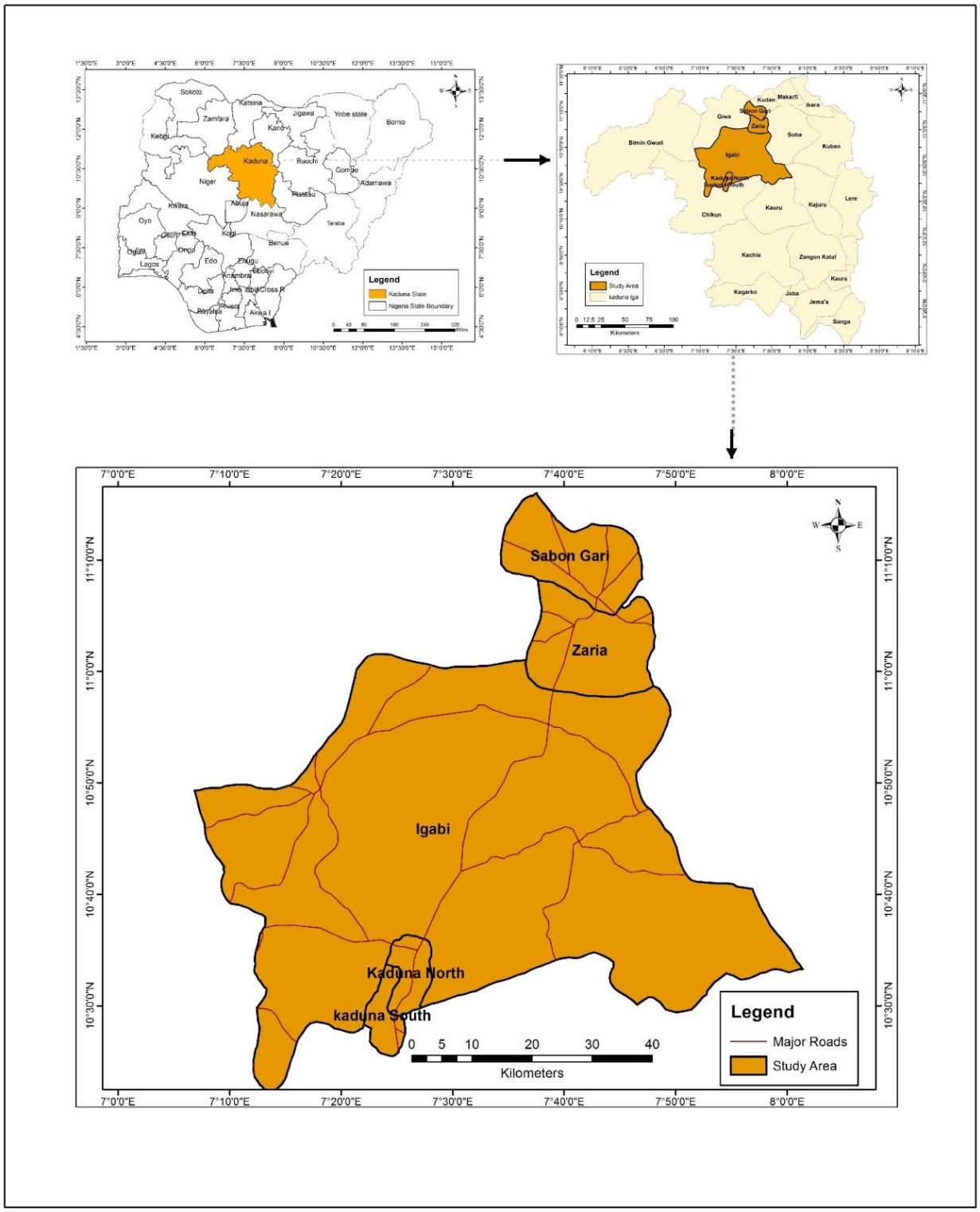

Figure 3.1: map of the Study Area.

Source: Modified from Administrative Map of Kaduna State, 2015

3.1.2Relief and drainage

The morphology of the area on which Kaduna is built comprises of a gentle undulating terrain at about 600m extending up to 50km outward but more to the north and south-east where it terminates at the foot of Jos Plateau. It is drained by river Kaduna. The river takes its sources from the Kajama hills on the Jos Plateau and flows for about 210km before reaching Kaduna town. It cuts across the city dividing it into Kaduna North and South areas. Beyond Kaduna, the river flows for about 100km into Shiroro dam in Niger State. From Shiroro, there is another 200km flow before it finally discharges into River Niger on its North-Central shores at Pategi (Sulaiman, 2001).

3.1.3Soil and vegatation.

The soil of the study area is mainly lithosols developed over the thoroughly weathered solid geology,The regolith is overlain by a thin layer of Aeolian drift (Buchanan and Pugh 1962, Udo 1978). Along Kaduna River and some of its tributaries are alluvial soils and swaps soils (Vertisol, locally called Fadama Soil). On the upland areas, lateritic iron stone form a capping and erosion has dissected it to form mesas and buttes with secondary detrital lateritic surfaces on the lower slopes.The highly weathered soils which are lateritised by loss of silical are considered to be a zonal soil (Coker, 2000). These soils are developed under the prevailing climatic condition with enough rainfall the soil aggregates are usually very small, particularly in wet conditions where the water holding capacity is good but this primarilydepends on the structural composition (Mallo, 1998).

The vegetation is characterized with the tree stratum dominated by *Isoberlina doka*. It is a parkland savannah made up of tall grasses as the lower stratum wiithin the city, the natural vegetation has been replaced by exotic species, which are observable in the parks, garden and

road sides while, some of the trees found within the city includes Eucalyptus (*Eucalytus globules*), medium size tree with large red flower (*Jacaranda minosaefolia*), timber tree (*Gmelia arborea*), and Mango (*Mangeifera indica*) Other includes Gold Mohur (*Delonix regia*), Guava *(Psidiumquajava)*, Margosa (*Azadirachata indica*), and so on. Food crops such as maize, sorghum, sugarcane and groundnut among several others are also grown (Coker, 2000).

3.1.4 Socio-economic activities.

Several Socio-economic activities are common within the Kaduna metropolis, and these activities are classified into; primary, seconadry and tertiary. The major activities in the area includes farming , quarryingon the banks of River kaduna as well as fishing. These activities produce vegetables, cereals, fruits and fish for sale in several markets within the metropolis. Three quarrying areas are well known and they are located at Malali, Badarawa and Tundu wada. These companies account for more than eighty percent of the crushed aggregates for civil engineering purposes in Kaduna Metropolis. Many industries are involved in reprocessing various finished products. The companies vary in scale; some are medium to large scale located within the industrial layout of Kakuri. They are petroleum refining, textile industries, soft drinks bottling (Coca cola products, Pepsi, etc) brewing, automoblie assembling (PAN), Defence Industries Corporation (DIC), Flour Mills, and so on. All these industries enjoy ready market within the metropolis.

Tertiary services are undertaken by Government and privately owned companies, parastatals, finance houses, legal chambers, architects, estates manager, and so on which are distributed all over Kaduna municipal. They offer various services that are in accordance with their areas of specialization. It is interesting to mention here therefore, that most of these services are located within the Central Business Districts (CBD) of the city.

3.1.5Ecological effects

In Kaduna state, the role of humanity and solid wastes are both contributing to the causes and effects of ecological changes within the metropolis. These include Climate change: Global warming, Global dimming, Fossil fuels, Greenhouse gas, Urban Heat Islands and Flooding. Roads and agriculture land uses cut across all category of slopes. However, agricultural land uses it least in the part of the metropolis because the area is designed for industries and its associated road linkages. Gulling pose the greatest threat to the survival of buildings and currently the most active in the areas where ground has been severely scared (Mallo, 1998).

3.2 Spatial Analysis

The spatial method used in this research is the Ripley'sK-function method . However, K-function method was used in this research because of its flexibility and application in a wider scale.

3.2.1 Ripley's K-function

Among much kind of statistical methods that examine the distribution of point on a plane, probably the K-function method is one of the frequently used methods in literature (Cressie 1991). K-function method overcomes the first limitation of the nearest neighbor distance method. It can distinguish various types of point distributions not distinguishable by the nearest neighbor distance. The method indicates the average number of points within a certain distance t from a point density. It is, therefore, a function of the distance *t*.K-function counts the number of points located in the circular region of radius t from every point in location(S), and divides it by the number and density of points for standardization.In contrast to the nearest neighbor distance, the K- function shows a large value when points are clustered.

Ripley's K function is a tool for analyzing completely mapped spatial point process data that is, data on the locations of events. These are usually recorded in two dimensions, but they may be locations along a line or in space. Here, the study will only describe K(t) for two-dimensional spatial data. Completely mapped data include the locations of all events in a predefined study area. Ripley's K function can be used to summarize a point pattern, test hypotheses about the pattern, estimate parameters and fit models. The subsequent discussions follow from Doguwa (1988;p 73-86)

K-function is mathematically defined by;

$K(t)=(\sum\sum \delta ij\ (t))/n\lambda$(3.1)

The numerator is the total number of points within a certain distance t from cholera point density.

n: Number of points

x_i: Locational vector of point i

λ: Point density (=n/A, A: the area of the region)

t: Distance parameter

δ_{ij}: Binary function defined by

$$\delta_{ij}(t) = \begin{cases} 1, & if|xi - xj| \leq t \\ 0, & otherwise \end{cases} \qquad (3.2)$$

K-function describes the degree of spatial clustering at the scale represented by the distance parameter*t*.A small *t* implies that we are discussing the point distribution at a small scale, in other

words, in a large spatial extent. If K(t) shows a large value for a large t, we say that points are globally clustered, or to be exact, points are clustered at the scale of t.Since both the K-function use distance between points, it is often called ‘distance methods’.

To evaluate the degree of spatial clustering of points, we again consider the distribution of points under Complete Spatial Randomness (CSR).

The expectation of K-function of points under Complete Spatial Randomness (CSR) is

$E[K(t)] = \pi t^2$ (Cressie, 1991) (3.3)

Comparing *K(t)* and its expectation, we can classify point distributions into one of three categories:

K(t) > πt^2: Points are clustered

K(t)= πt^2: Points are randomly distributed

K(t)<πt^2: Points are dispersed

Standardization of K-function

To compare K-function values among different t values, we standardize the K-function (Besag 1991)

$$L(t) = \frac{\sqrt{K(t)}}{\pi} - t \qquad (3.4)$$

The standardized K-function is called *L*-function.

Point distributions are then classified as below.

L(t)> 0: Points are clustered

L(t) =0: Points are randomly distributed

L(t)< 0: Points are dispersed

Suppose that an observed point pattern can be considered as a realization of a spatial point process model which is stationary and isotropic. The property of the underlying spatial point process elicited directly or indirectly, by the second order methods is the second moment cumulative function $k(t)$ introduced by Ripley (1976, 1977). This function $k(t)$ has attracted considerable attention and is defined through the relation.

$$\lambda^2 Ak(t) = E\{\text{Number of ordered pairs of distinct objects. the first in D, the second within distinct t of the first but not necessarily in D}\} \quad (3.5)$$

Here D is the sampling window, typically a rectangular window with sides a and b ($a \le b$), diagonal of length d and area A. The parameter λ is the intensity of the process defined as the mean number of objects per unit area. An equivalent definition is provided by the relation

$$\lambda k(t) = E\{\text{Number of other objetcs within distance t of an arbitrary object of the process }\}$$

$$= \int_0^{2\pi}\int_0^{t} \{\lambda_2(r)/\lambda\} r dr d\theta \quad (3.6)$$

Where $\lambda_2(r)$ is the second order intensity function of the stationary isotropic process and r is the distance between any two objects of the process.

The usual reference point for object patterns is the random object pattern or Poisson forest, for which $k(t) = \pi t^2$. Thisresult suggests the use of the statistic $L(t)$ defined by

$$L(t) = [k(t) / \pi]^{1/2} \qquad (3.7)$$

Besag, in the discussion of Ripley(1977) suggested the use of a square-root scale to linearize the plot of $k(t)$ against t for a Poisson process. This also acts as a variance stabilizer of objects at short distances and hence k(t).

In the case of a Poison forest $L(t) = t$, for all values of t. For a real object pattern, a plot of $L(t)$ against t can be used to detect the nature of any departurefrom randomness. For clustered or aggregated patterns, there are an abundance of objects at short distances from one another and hence $L(t) > t$, whereas with repulsion between the objects (regularity), there will be deficit of objects at short distances and hence $L(t) < t$. The relations $L(t) > \text{t}, and L(t) < t$ are for small t, asymptotically $L(t) \to t$

Although this chapter is ostensibly concerned with the estimators of $k(t)$, it will be noted that these estimators are not obtained directly but are derived from the estimators of either $\lambda^2 Ak(t) or \lambda k(t)$ the left hand side of equations (3.5) and (3.6) . Furthermore, equation (3.7) involves $\{k(t)\}^{1/2}$ rather than $k(t)$ itself. An optimal estimator of $k(t)$ need not be optimal estimators of $\{k(t)\}^{1/2}$ and we therefore examine the properties of the existing estimator of $k(t)$ in the content of $L(t)$.

3.3 The Estimators of K(t)

3.3.1 The Naive Estimator

A naïve approach consists of counting the number of observable pairs of objects, lying within the window D, that are less than a distance t apart. The resulting estimator, $K_1^{\wedge}$ (t) is given by the equation (3.8) below.

$$\lambda^2 \hat{K}_1(t) = \frac{\sum_{i \neq j}^{n} \sum^{n} 1_{ij}}{A} \qquad (3.8)$$

Where $1_{ij} = 0$ if the distance between the i^{th} object and the j^{th} object is more than t, and $1_{ij} = 1$, otherwise. Usually the value of λ^2 would be unknown, in which case it can be replaced by the estimate $n(n-1)/A^2$ which is unbiased for a Poisson process.

The resulting estimator $\hat{K}_1(t)$ is evidently negatively biased, since it fails to take proper account of neighboring objects lying undetected outside D but in its vicinity (Doguwa,1988;p 73)

Now $E\{A\lambda^2 \hat{K}_1(t)\} = E\{\sum_{i \neq j}^{n} \sum^{n} 1_{ij}\}$ (3.9)

And Ohser et al (1985) showed that for a Poisson process, the right-hand side of (3.9) is.

$$E\{\sum_{i \neq j}^{n} \sum^{n} 1_{ij}\} = \lambda^2 \int_0^t \gamma_D(r) dK(r) \qquad (3.10)$$

Where γ_D (r) = E{ (area of $D \cap D + y$)} and y is uniformly distributed on the boundary of the circle center the origin and radius r. The function γ_D (r) can be interpreted in the following way:

Consider the object X_j situated at a distance r from the arbitrary objet X_i. Suppose that it is known that the line $X_i X_j$ makes an angle θ to some reference direction. A moment's consideration reveals that the position of X_j is restricted, because it is known that X_i lies within the window. If θ is known then it is an easy matter to calculate the area of the restricted region within which X_i must lie. However, a different value of θ would lead to a different position for the restricted region and this need not have the same area. If the process is isotropic, then all the values of θ are equally likely and we then require γ_D (r), the average area of the corresponding equally likely restricted regions (Doguwa, 1988; p 74).

For our rectangular window,Ohser et al (1985) deduce that for small enough r, γ_D (r) is approximately equal to $A - r(2a + 2b)/\pi$. Therefore,

$$E\{\sum_{i \neq j}^{n}\sum^{n} 1_{ij}\} = \lambda^2 \int_0^t \gamma_D(r) dK(r)$$

$$\approx \lambda^2 \{A \int_0^h dK(r) - [(2a + 2b)/\pi] \int_0^t r dK(r)\}$$

However for a Poisson process, $K(r) = \pi r^2$ and

$dK(r) = 2\pi r dr$ $E\{\sum_{i \neq j}^{n}\sum^{n} 1_{ij}\} = \lambda^2 \int_0^t \gamma_D(r) dK(r)$ K(r) = πr^2, and dK(r) = 2π rdr. Hence

$$E\{\sum_{i\neq j}^{n}\sum^{n}1_{ij}\} \approx \lambda^2 A\pi t^2 - 4\lambda^2 t^3(a+b)/3$$

Thus:

$E\{\hat{K}_1(t)\} = \pi t^2 - 4t^3(a+b)/3ab$, for small enough t.

3.3.2 Border edge correction

A simple procedure that avoids the problem inherent in equation (3.8) is to consider only those objects within a variable inner window D_0 , which shrinks as t increases.

Thus, for the case $t = t'$, the inner window is a centrally located rectangle of dimensions $(a-2t')$ by $(b-2t')$. The effect is that the positions of any objects up to a distance t' outside the inner window are known. The resulting estimator $\hat{K}_2(t)$ is given by the equation.

$$\lambda^2 \hat{K}_2(t) = \frac{\sum_{i}^{n_0}\sum_{j}^{n}1_{ij}}{A_0} \qquad (3.11)$$

Here A_0 is the area of D_0 and n_0 is the number of objects in D_0. If λ^2 is unknown, it can be estimated by the estimate $\frac{n_0(n-1)}{AA_0}$. Evidently, $\hat{K}_2(t)$ is unbiased for all values of t that are less than a/2 and providing that the inner window is not empty of objects. To see this, we obtain the expected value of (3.11) conditional on the number n_0 of objects in D_0 as,

$= [\sum_{i}^{n_0}\sum_{j}^{n} E[1_{ij}]$ for a Poisson process

$$\sum_{i}^{n_0}\sum_{j}^{n}\int_{0}^{t}\{\int_{D\cap\partial b(x_i,r)} dV\}dr / A \text{ where } r = d(x_i, x_j)$$

$$= n_0(n-1)\int_{0}^{t} 2\pi r dr / A$$

$$= n_0(n-1)\pi t^2 / A$$

Where V is a one dimensional Lebsesgue measure on $\partial b(X,r)$, the boundary of the circle centered at X and radius r. Now

$$E\{\hat{K}_2(t) / N(D_0) = n_0\} = n_0(n-1)\pi t^2 / AA_0\lambda^2 = K(t)$$

However, the rapidly decreasing size of D_0 inevitably leads to a rapid increase in the variance of this estimator (Doguwa, 1988; p 75-76).

3.3.3 Ripley's Estimator

Ripley (1976, 1977) suggested considering every ordered pair of objects (X_i, X_J) within the sampling window and assigning a scaling factor S_{ij} to each pairing. The scaling factor is defined by,

$(S_{ij})^{-1}$ = The proportion of the perimeter of the circle, centered on object X_i and passing through object X_j, which lies within the sampling window D

$$S_{ij}^{-1} = \frac{V[D\cap\partial b(X_i, r)]}{2\pi r} \qquad (3.12)$$

Note that S_{ij} is not necessarily equal to S_{ji}. We shall refer to object X_i as the central object and object X_j as the distant object.

Ripley's estimator $\hat{K}_3(t)$ is given by the equation,

$$\lambda^2 \hat{K}_3(t) = \frac{\sum^{n} \sum_{i \neq j}^{n} 1_{ij} S_{ij}}{A} \tag{3.13}$$

Usually the value of λ^2 will be unknown, the which case we suggest replacing it by the estimate, $n(n-1)/A^2$ which is unbiased for a Poisson process. For a Poisson process, $\hat{K}_3(t)$ is unbiased for values of t for which $2t < d$. These follows since;

$$E\{\lambda^2 A \hat{K}_3(t) / N(D) = n\} = E\{\sum^{n} \sum_{i \neq j}^{n} S_{ij} 1_{ij}\}$$

$$= \sum^{n} \sum_{i \neq j}^{n} E\{S_{ij} 1_{ij}\}$$

$$= \sum^{n} \sum_{i \neq j}^{n} \int_0^t \{ \int_{D \cap \partial b(x_i, r)} S_{ij} dV \} dr / A$$

$$= \sum^{n} \sum_{i \neq j}^{n} \int_0^t S_{ij} V[D \cap \partial b(X_i, r)] dr / A$$

But $S_{ij} V[D \cap \partial b(X_i, r)] = 2\pi r$, for all i and j, provided that,

$$V[D \cap \partial b(X_i, r)] \neq 0 \tag{3.14}$$

However (3.14) is satisfied if and only if $2t < d$. Hence

$$E\{\lambda^2 A\hat{K}_3(t) / N(D) = n\} = n(n-1) / A\int_0^t 2\pi r dr \text{ for } 2t < d$$

$$E\{\hat{K}_3(t) / N(D) = n\} = K(t)$$

Ripley's estimator is limited to values of t for which $2t < d$. The restriction occurs because the formulation of $\hat{K}_3(t)$, assumes that each of the objects in D can assume the role of a central object. However if $2t < d$, then there will be a 'null region' in the interior of D containing objects for which this central role has been removed, because the entire perimeter of the circle of radius t centered on one of these objects lies outside the sampling window.

Theoretical value of K(t) for a Poisson Process

In order to study the properties of the various estimators of $k(t)$,a random process was considered.. The shape of the window will have an effect on the performance of the estimators, and therefore, for each process type we use a unit square window and/ or a rectangular window with a length/breadth ratio of 10.

The Poisson process or Poisson forest incorporates two basic properties, those of uniformity whereby the objects exhibit no tendency to occupy particular regions of the plane, and independence, whereby the locations of given objects are determined without reference to that of any other. These two basic properties imply that, the second order intensity function $\lambda_2(t) = \lambda^2$ for $t > 0$, and from equation (3.6)

$$\lambda k(t) = \int_0^{2\pi} \int_0^t \{\lambda r dr d\theta$$

So,

$k(t) = \pi t^2$ for a Poisson Forest (3.15)

3.4 Dataused

The types of data used in this research are polygon and point data. The polygon data is the administrative map of the study area, thelocation of the cholera cases (point data) and inventory data for the cholera epidemic which was used in populating the locations of the cholera patient's settlements. These type of data used in this research is mainly the secondary type. The secondary data is the base line data list of all cholera outbreaks in Kaduna state from 2010 to 2015. The cholera outbreak inventories comprises information on the locations of the cholera cases in terms of wards and localities, number of people affected, age and sex respectively. The data was obtained from Kaduna State Ministry of Health.

3.4.1Procedurefor data analysis and presentation of results

The software used is Google earth image 2016 and the R statistical Package. It was used to update digital map in ArcGIS 10.3 environment. The Microsoft-excel software was used to compile cholera data and ArcGIS 10.3 software and R statistical package was used as the engine for the data analysis and presentation of results.

To analyze the spatial pattern and cluster of cholera epidemic in the study area, inventories provided by theMinistry of Health Kaduna State, coordinate of the locations of towns and

localities were used for the analysis in ArcGIS 10.3 software environment and R Statistical Package. Among the inventory data collected for the cholera epidemics in the study area was in Microsoft-excel format, the data was then converted in to Microsoft query format in excel. The query was done to extract the name of wards, name of localities and the cases of cholera with respect to the localities. These aids in locating the geographical locations of the localities and their coordinates on Google Earth Pro software. This was done in order to integrate the spatial (location of the localities and settlement) in the study area with the non-spatial data (the attributes of the Cholera disease i.e. the cases) so as to link both the spatial and non-spatial data because the data collected from the Kaduna State Ministry of Health were not referenced to geographic locations.

The data from the excel was imported into ArcGIS 10.3 software environment using the Add data tool option as a layer and add in to the GIS interface as shapefile. The shapefile was later converted in to feature classes and this was done in preparation of the data for further analysis. The data was analyzed using the ArcCatalog window which is an ArcGIS software extensionthat is used in carrying out the analysis. Then Spatial Analysis tool and K-Function were selected and clicked for the spatial analysis of cholera epidemics. For each of the analysis conducted, a report was generated and the outputs of the results were presented in theform of Figures, Maps and ArcMap documents.

In R Statistical software, For each Local Government, co-ordinate of the window in latitudes and longitudes are collected. Also, for each of the locations in the window, co-ordinates are converted to distances in kilometers relative to the co-ordinates of the window.

Let d be the diagonal of the window in kilometers and r the radius of interest also in kilometers. Then for each location, say x_i, the Euclidean distance is calculated from every other location, say $x_j \ \forall \ i \neq j$. Our interest is in Euclidean distances that are less than or equal to a particular r where $\frac{d}{100} \leq \mathbf{r} \leq \frac{d}{4}$. Then using the border edge corrections, the spatial pattern is analyzed.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Spatial Distributions

Figure 4.1 shows the location of the five selected Local Government Areasthat were highly affected by the cholera epidemicand the locations of the cholera cases in the areas. It also shows the patterns of the spatial distribution of cholera epidemic in the study area which comprises five local Government areas of Kaduna State, Namely; Kaduna South, Kaduna North, Igabi, SabonGari and Zaria.

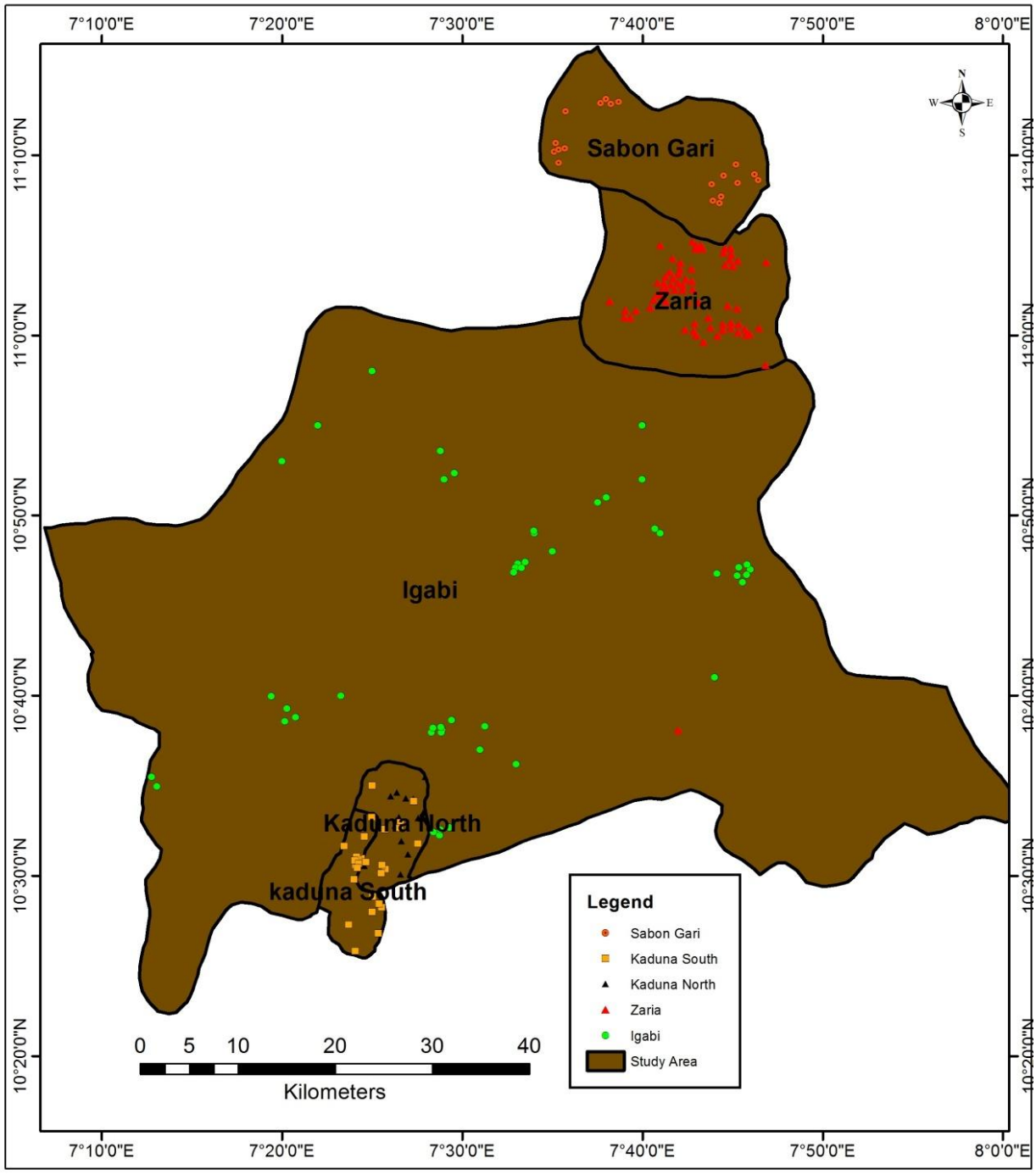

Figure 4.1: Spatial Location of the Cholera Epidemic in the Study Area

Figure 4.1 shows the map of Kaduna State indicating the five Local Governments in study with atotal of 187 locations with cholera incidence between the years 2010 to 2015 in the study area.

Five Local Government Areas were highly affected with Zaria LGA having 71 locations affected by the cholera incidence. Igabi LGA has 53 locations, SabonGari with 19 locations and Kaduna North and South with 16 locations and 28 locations respectively. On the reported cases across the LGAs under study, 378 and 447 cases were reported in Kaduna North and South respectively, while 217 cases were reported in Zaria LGA, 1363 cases in Igabi LGA and 301 in SabonGari LGA. The disease mapping provides a rapid visual summary of complex geographic information as supported by Goovaert and Jacquez (2004). The results show the advantages of advances in technology that allows disease and spatial distribution of disease mapping as pointed out by Kulldorff*et al,* Rosenberg (1999), Ali *et al*.,(2001,2002a, 2002b).

4.2.3 Spatial Patterns of Cholera Epidemic in the Study Area

The results of the analysis arepresented on Figure 4.2 to 4.6.

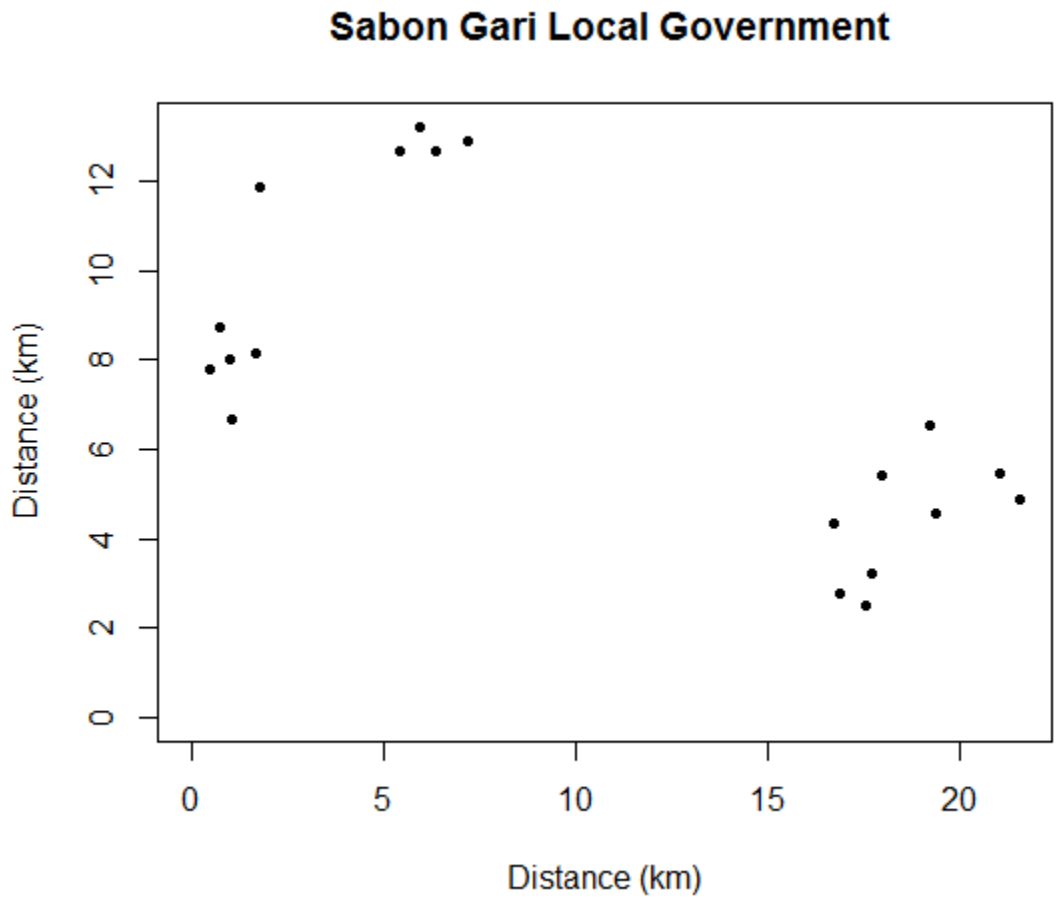

Figure 4.2: Locations of Cholera Epidemic in Sabon-Gari LGA

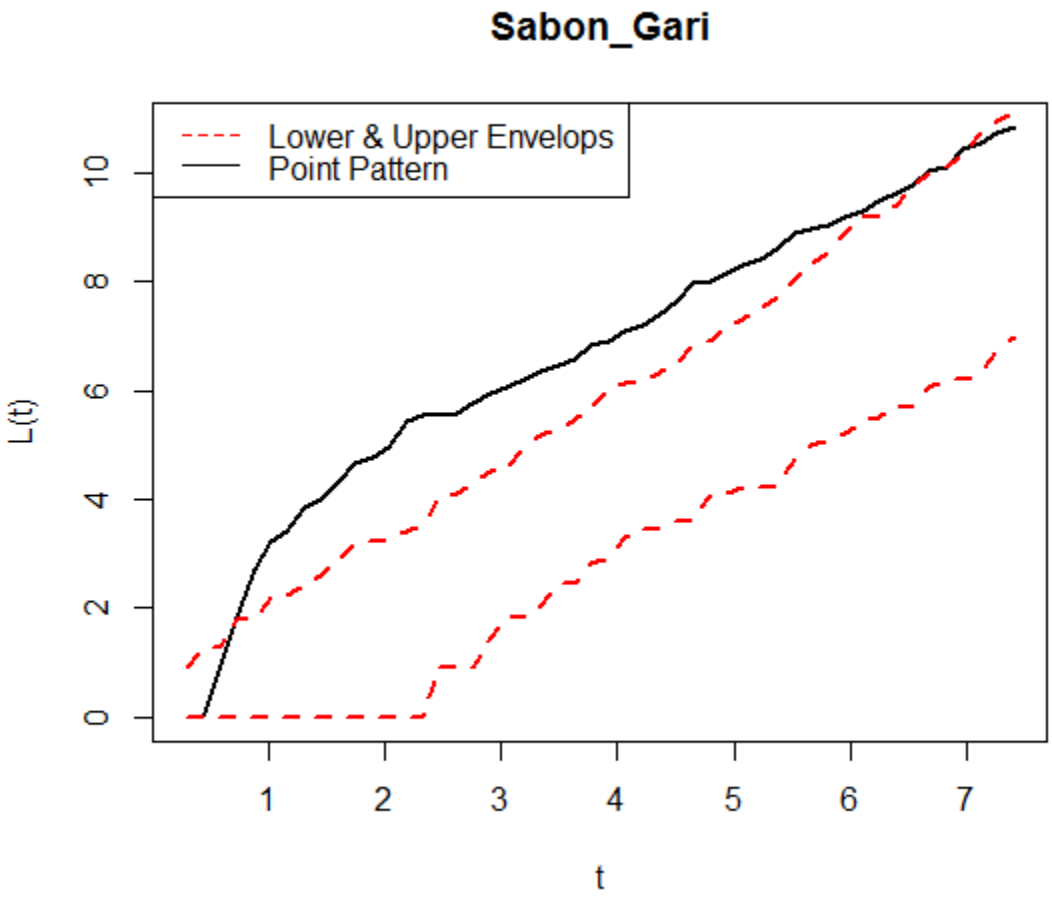

Figure 4.2.1: The L(t) function and the 99% upper and lower envelops of the spatial pattern Fig 4.2.

The black line in Figure 4.2.1 is the observed L-function and is above the Upper Confidence Envelopfor values of h between 0 to 7.5 indicating clustering within the range. The result revealed that the disease was clustered in Sabongari LGA like the research conducted by Diego *et al.*, (2010) which show that spatial clustering of cholera cases was detected at different temporal and spatial scales.The result of Figure 4.3 was generated from the map in Figure 4.1 to show the chats of clustering locations.

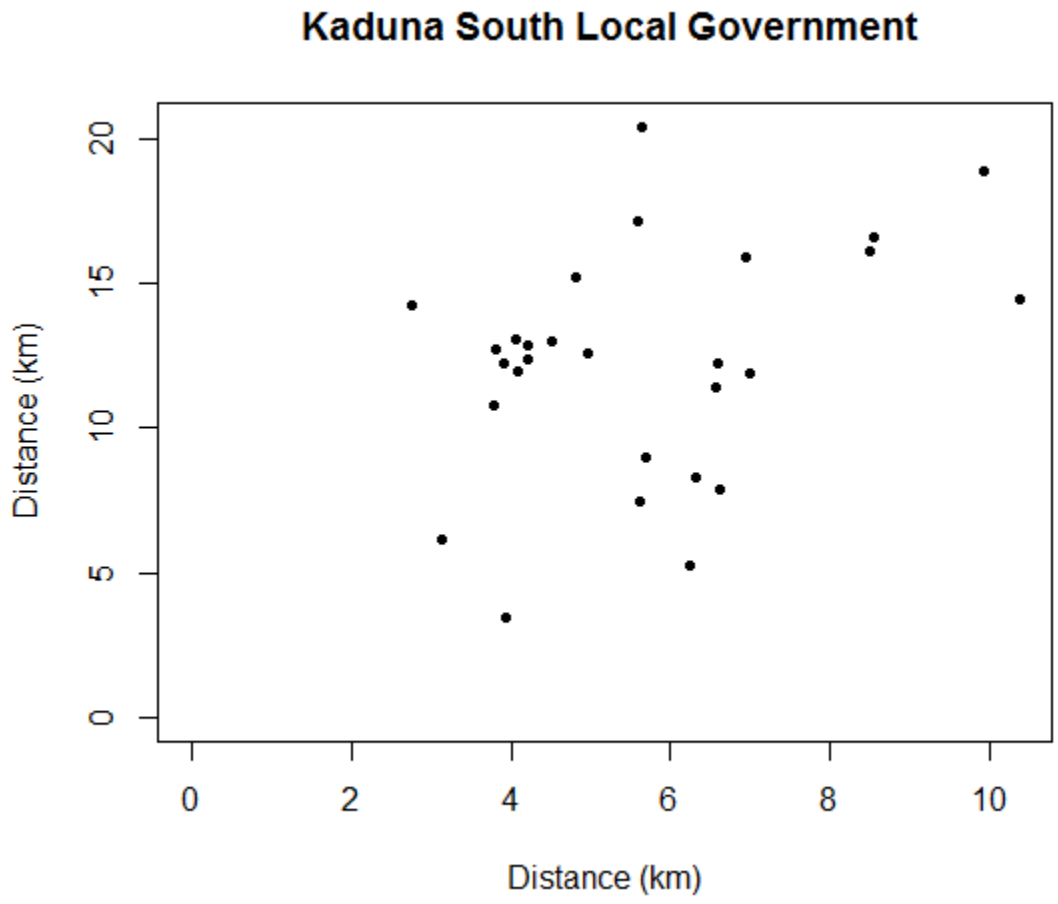

Figure 4.3: Locations of Cholera Epidemic in Kaduna South LGA

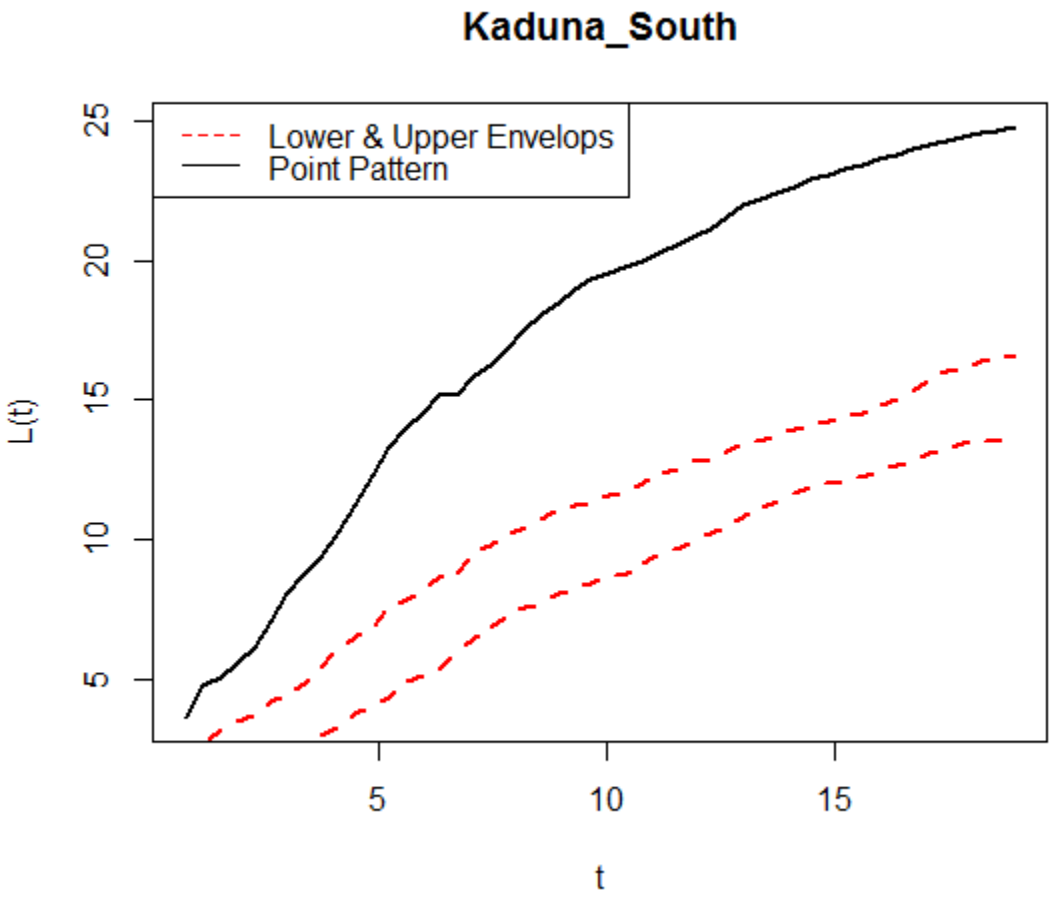

Figure 4.3.1The L(t) function and the 99% upper and lower envelops of the spatial pattern Fig 4.3.

Also from Figure 4.3.1, the black line revealed that the observed L(d) is above the Confidence level for values of h between 3 to 21indicating clustered pattern of spread of Cholera disease.

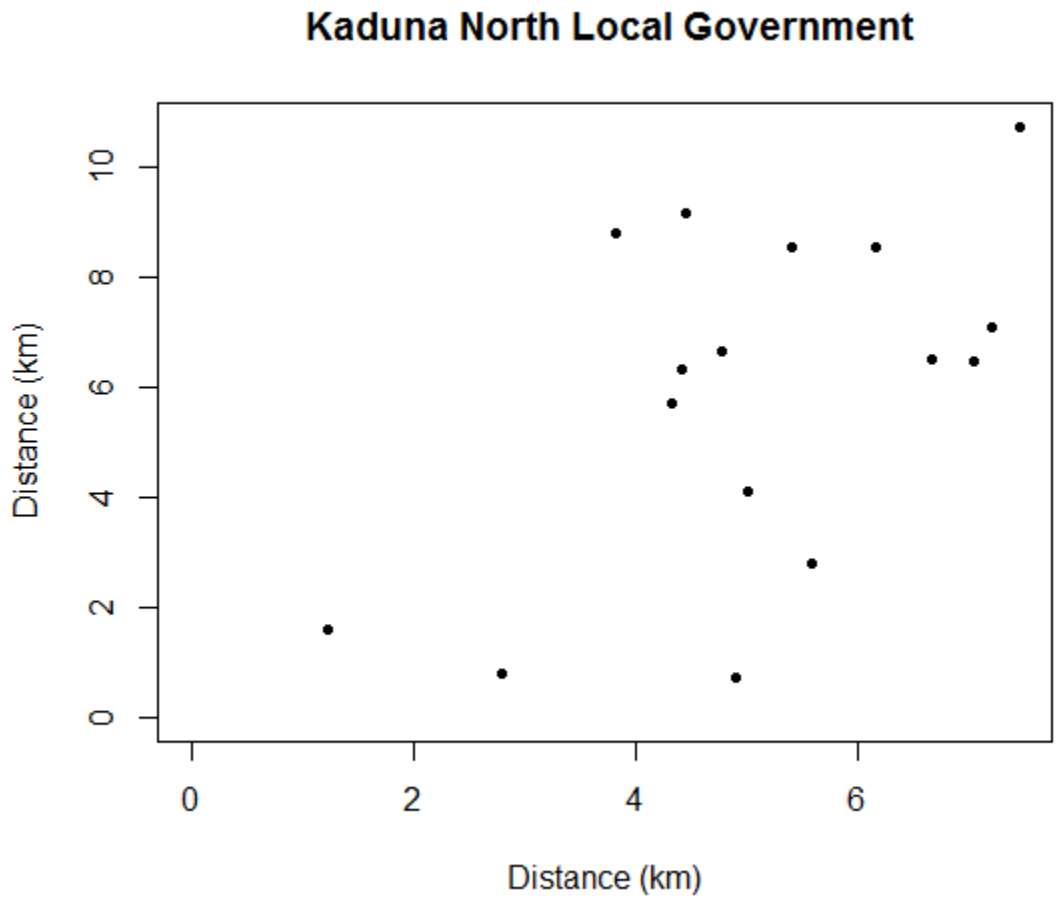

Figure 4.4: Locations of Cholera Epidemic in Kaduna North LGA

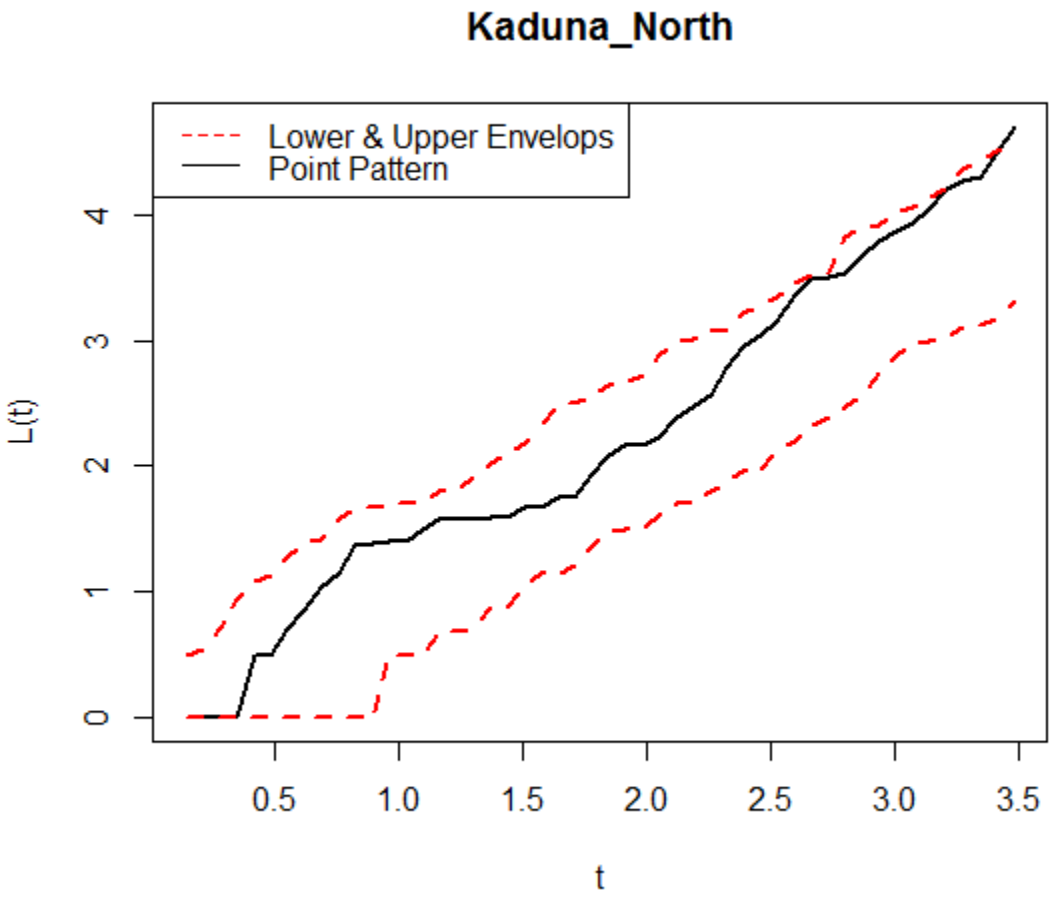

Figure 4.4.1: The L(t) function and the 99% upper and lower envelops of the spatial pattern Fig 4.4

It was found from the result on Figure 4.4.1, that the spatial pattern of cholera disease in Kaduna North is random as shown from the black line that the observed L(t) is roughly within the 99% Confidence Envelops. This findingalso contradicted the research conducted byDiego *et al.,* (2010) which show that spatial clustering of cholera cases was detected at different temporal and spatial scales and cases relative to water sources also exhibit spatial clustering.

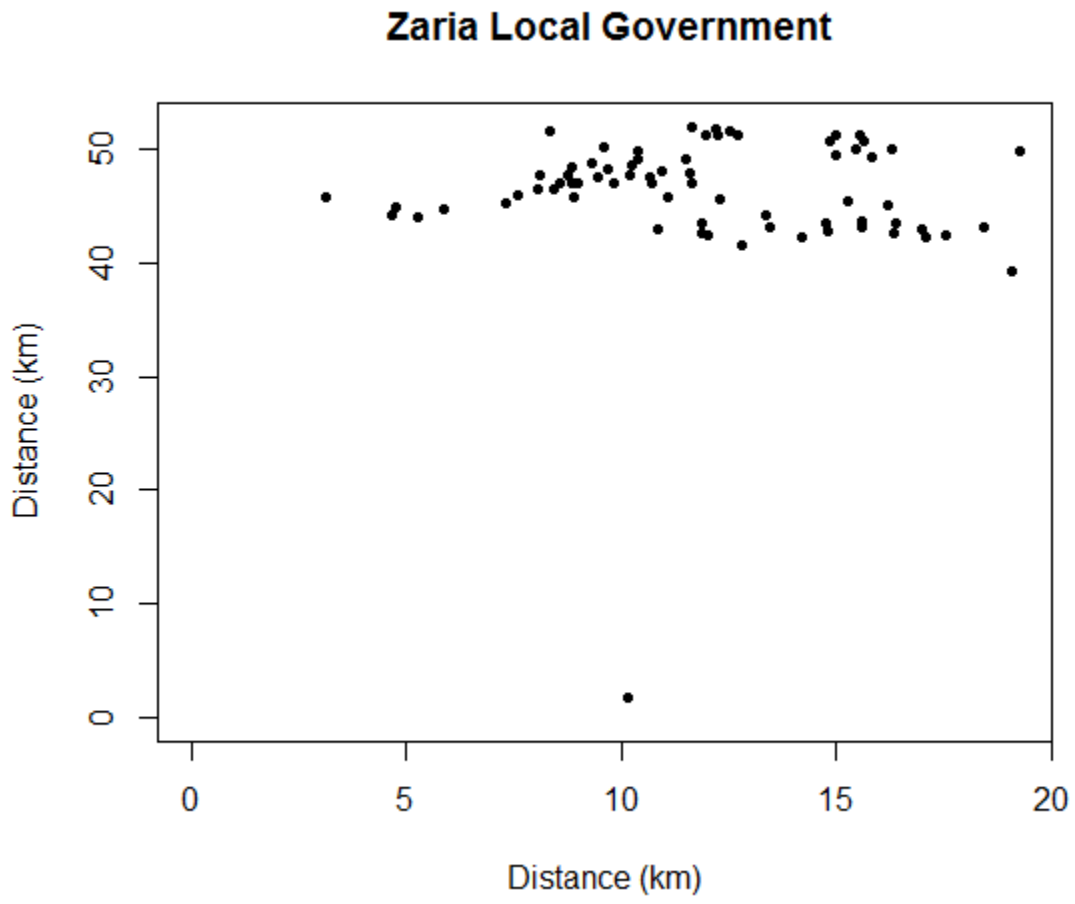

Figure 4.5: Locations of Cholera Epidemic in Zaria LGA

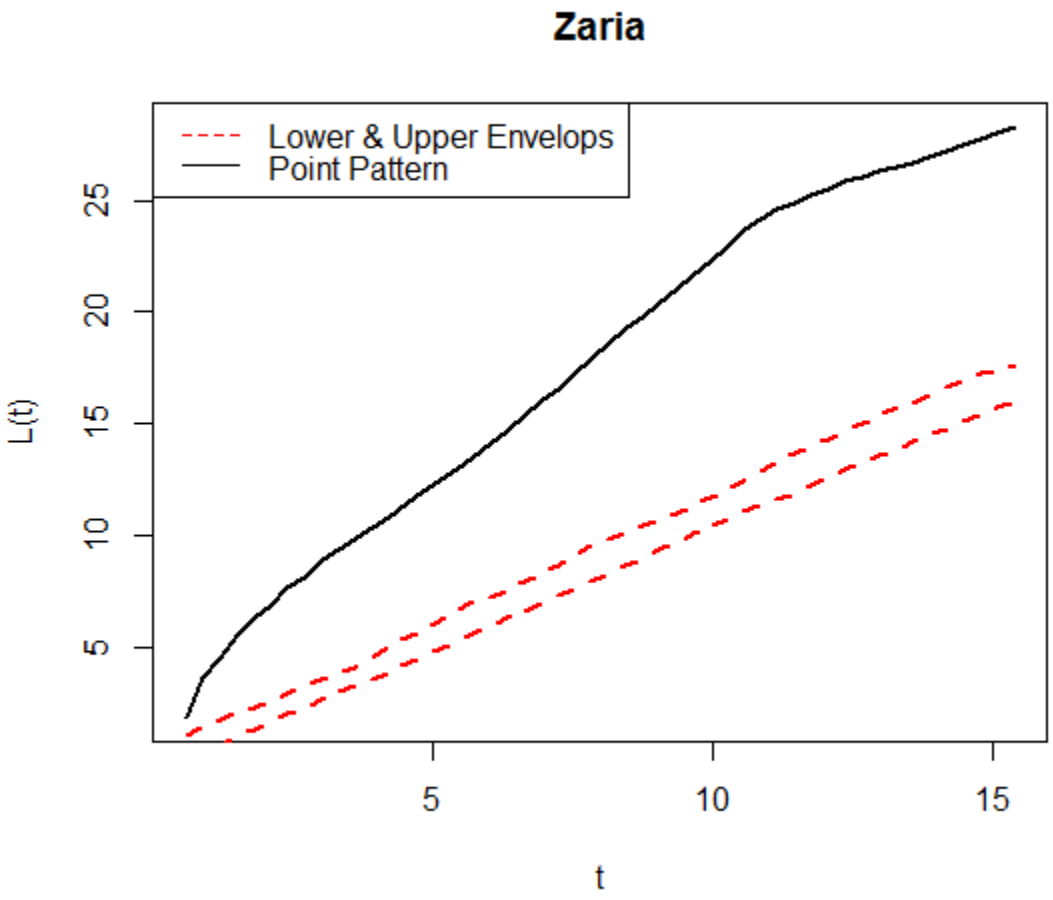

Figure 4.5.1: **The L(t) function and the 99% upper and lower envelops of the spatial pattern Fig 4.5**

The spatial pattern of cholera disease in Zaria is clustered. As indicated in Figure 4.5.1, from the black line that the observed L(t) is constantly above the Upper Confidence Envelopas in the research finding of Ruiz-Moreno et al (2007) which observed that clustering of cholera in Bangladesh occur at different spatial scales. This result confirmed the outbreak reported in Zaria in Kaduna State with 192 cases recorded in two weeks at 2015 (UNICEF, 2015). However, according to WHO (2013) the cholera outbreaks are attributed to the fact that Cholera is most likely to be found and spread in places with inadequate water treatment, poor sanitation, and inadequate hygiene.

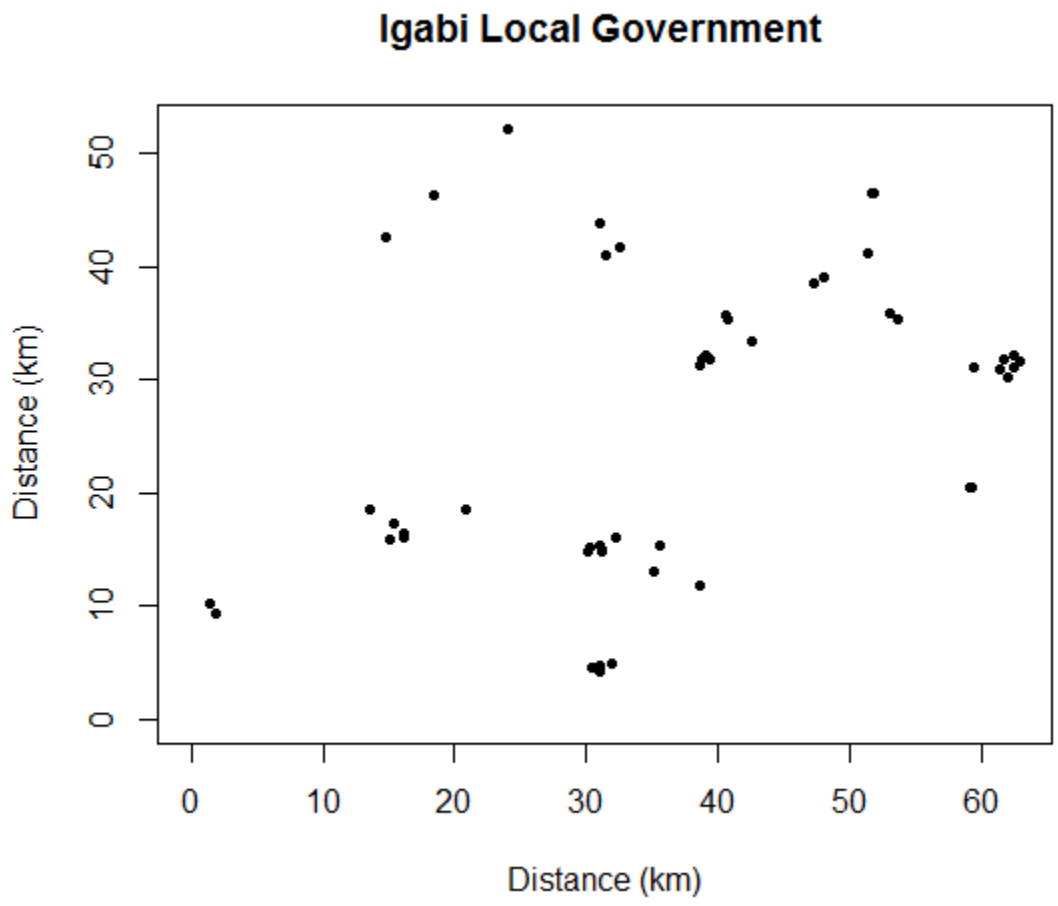

Figure 4.6: Locations of Cholera Epidemic in Igabi LGA

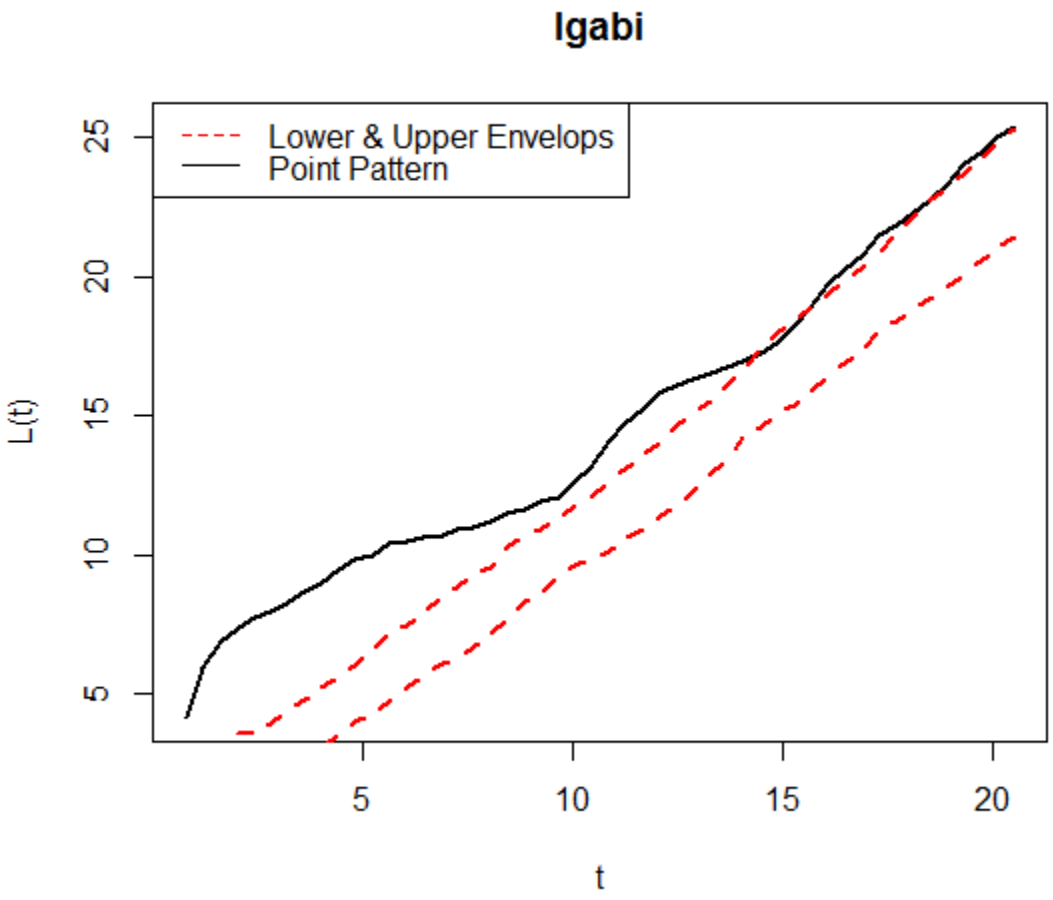

Figure 4.6.1: **The L(t) function and the 99% upper and lower envelops of the spatial pattern Fig 4.6**

Moreover, the observed black line in Figure 4.6.1 is above the Upper Confidence Envelop for values of h between 3 to 20indicating clustering pattern of cholera in Igabi LGA. The situation in Igabi LGA was as a result of socioeconomic and demographic factors because most of the people living in that areaare farmers and fishermen and it has been reported by (Ali *et al* 2002a, 2002b; Borroto and Martinez-Piedra 2000; Sasaki *et al* 2008) that socioeconomic and demographic factors significantly enhance the vulnerability of a population to infection and contribute to epidemic spread of cholera.

4.2.3 Spatial patterns of cholera epidemics in the study area

The result for the entire spatial pattern model from the study area is presented in Figure 4.7.

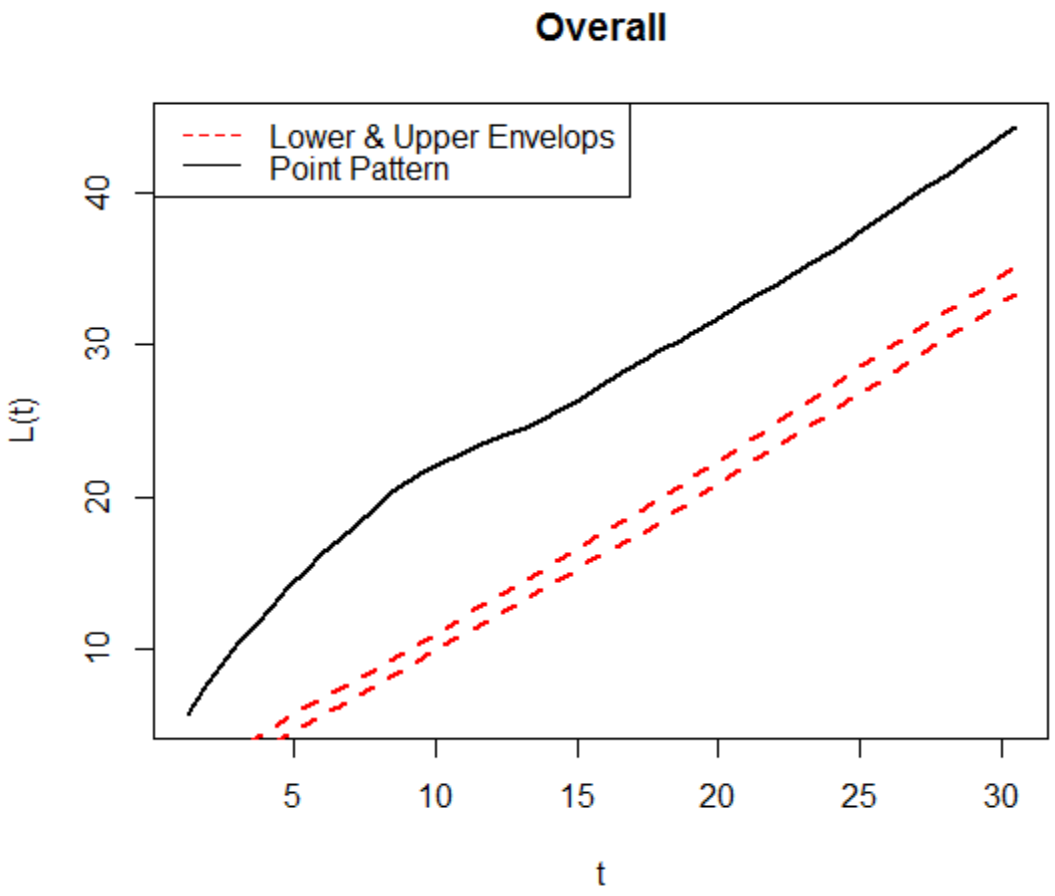

Figure 4.7: Combine Patterns of Cholera Epidemics in Five LGAs

It was found from the combined cholera records that the spatial pattern of cholera disease in the whole study area is clustered. As indicated in Figure 4.8, the observed red line is above the Upper Confidence Envelop for values of t between 2 to 30 showing clustering pattern of Cholera epidemics in the area. The result is not surprising because cholera cases have been recorded as endemic in Nigeria since 1991 (Utsalo*et al* 1991; Eko et al 1994; Falade and Lawoyin 1999; Hutin 2003) as reported from various parts of the country. This is also due to the fact that

Kaduna State was recorded amongst 19 with suspected cholera cases from January 2014 to October 2014.

4.3 Degree of Spatial Clustering of Points and Disease Risks

The result for the Degree of spatial cluster of Cholera locations and disease risks were presented in figure4.8 to 4.11. The point density was used to calculate a magnitude per unit area from point features that fall within a neighborhood around each cell in the study area. In evaluating the degree of spatial cluster, only the points that fall within the neighborhood were considered in calculating the density.

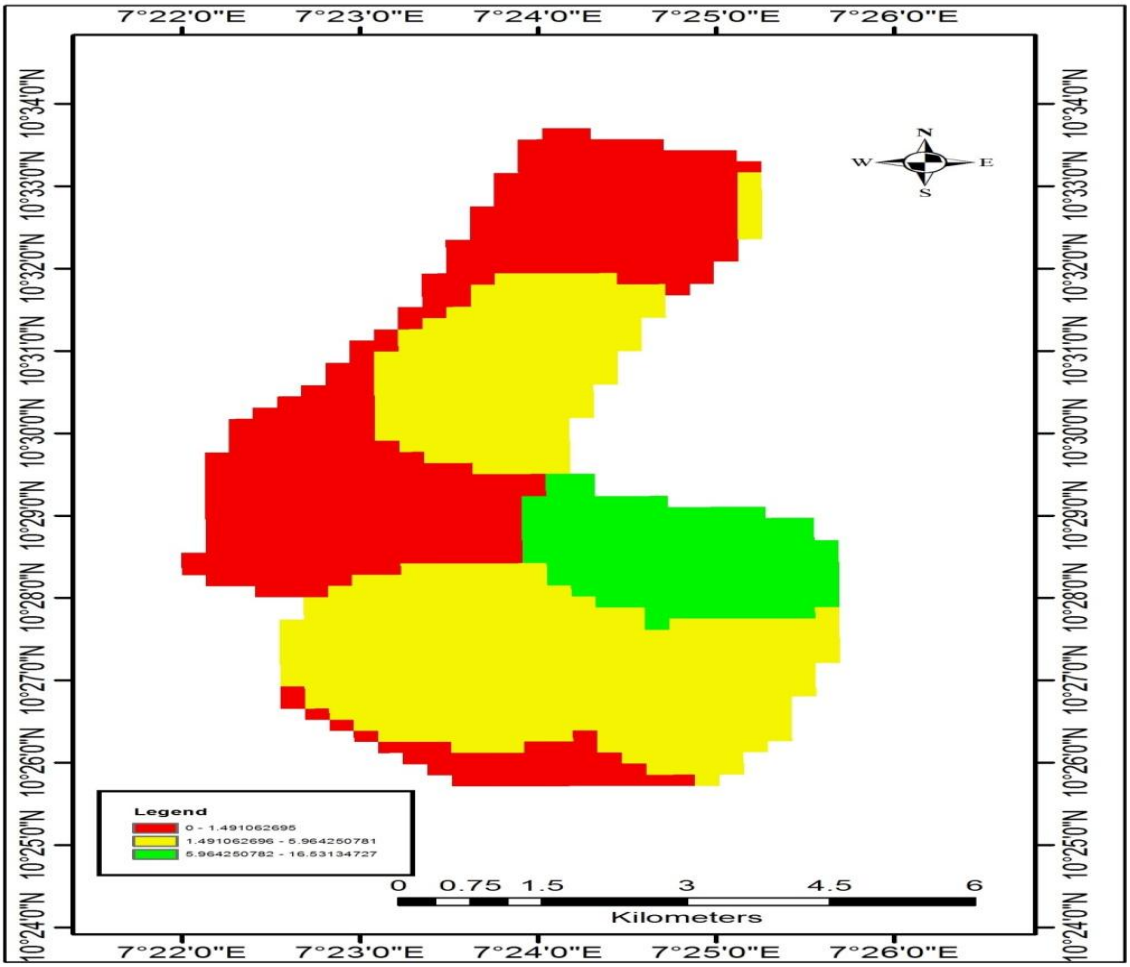

Figure 4.8: **Spatial Clusters of Cholera Disease in Kaduna South LGA**

From figure 4.8, the result shows that the yellow color constitutes higher cluster which ranges from 1.4 to 5.9, red moderately cluster (0.1 to 1.4) and green (5 t0 16) with lowest spatial cluster respectively. This shows that Spatial Clusters in Kaduna South LGA ranges from highly to moderately clustered.

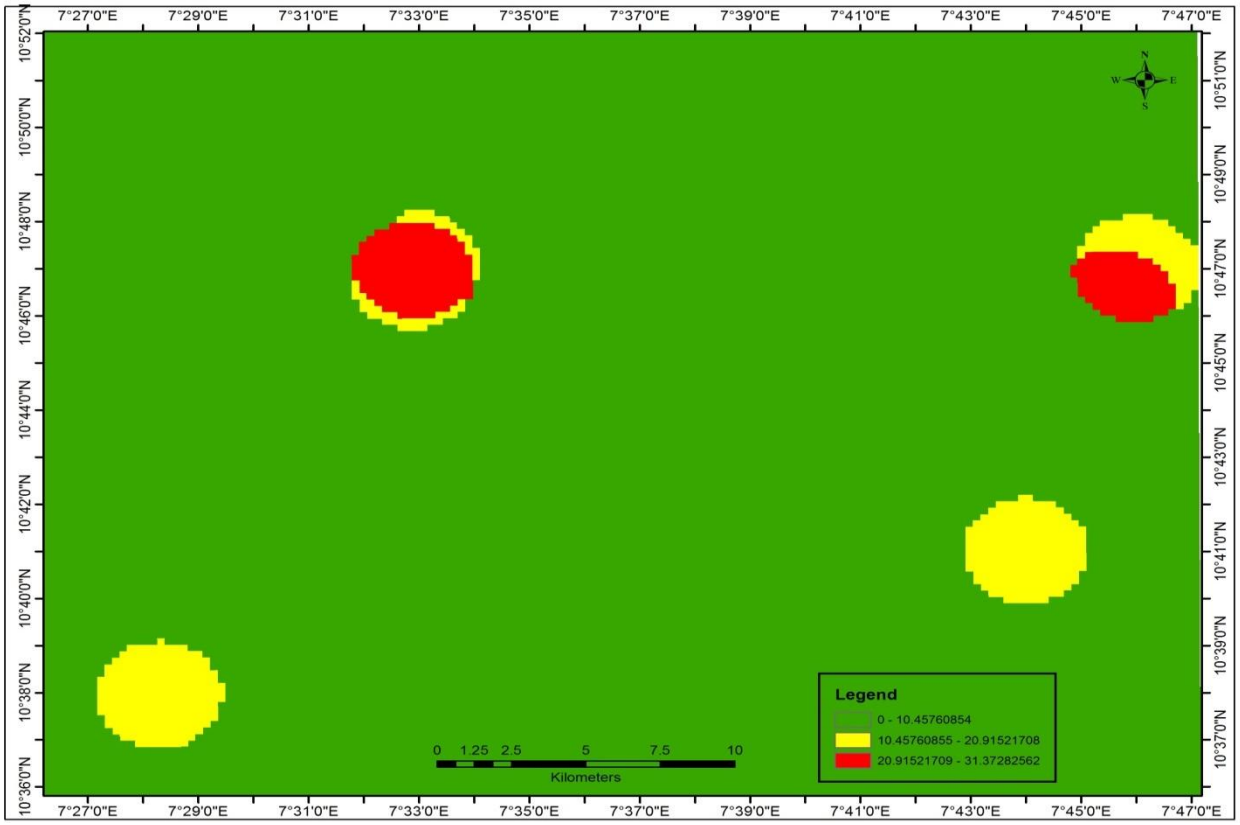

Figure 4.9: **Spatial Clusters of Cholera Disease in Igabi LGA**

From figure 4.9, the result shows that the green color constitute higher cluster which ranges from 0 to10, followed by yellow (10 t0 20) and red moderately cluster (20 to 31) with lowest spatial cluster respectively. This shows that Spatial Clusters inIgabiLGA was low clustered.

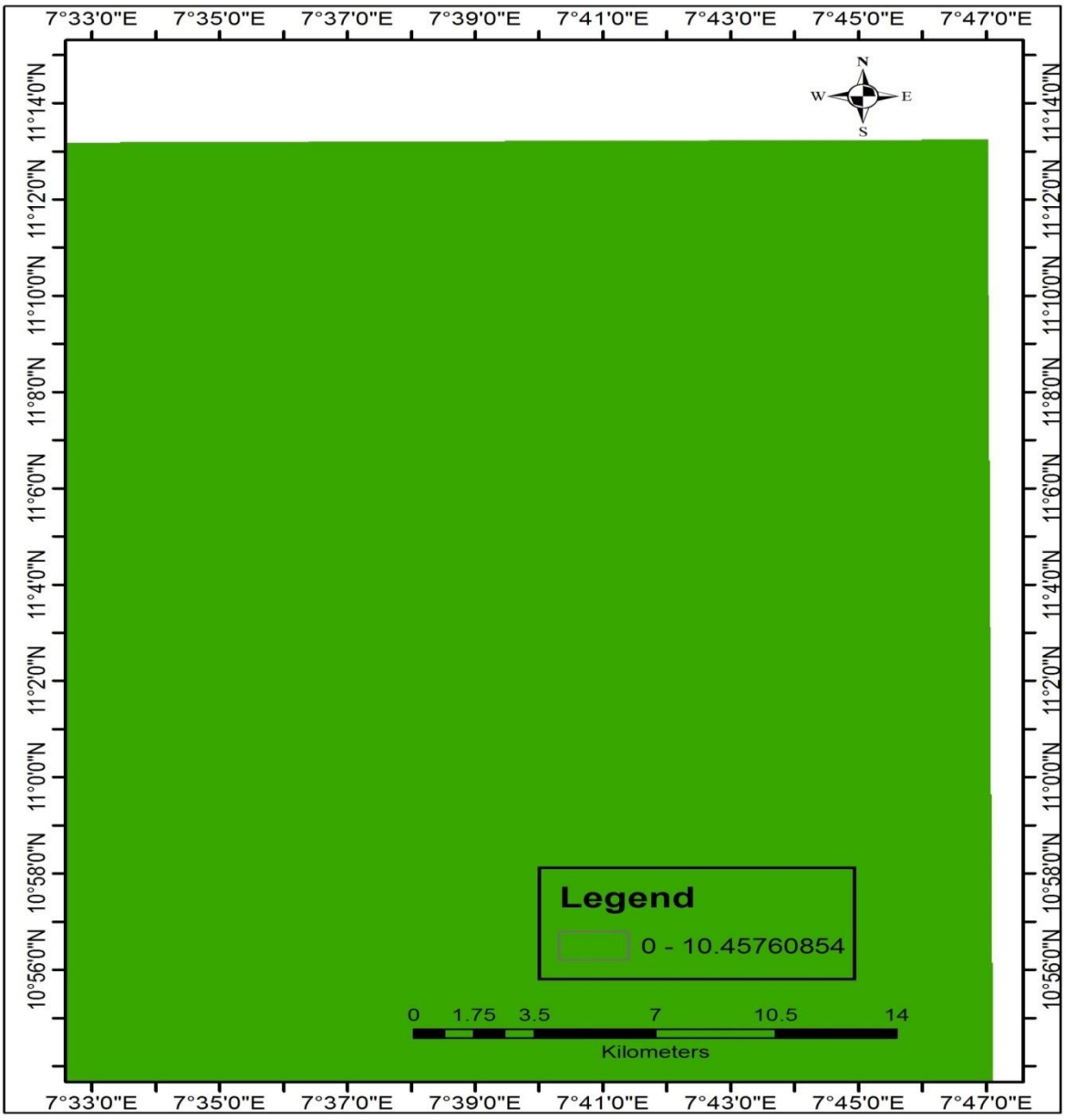

Figure 4.10: **Spatial Clusters of Cholera Disease in Sabongari and Zaria LGAs**

From figure 4.10, the result shows that the green color constitutes higher cluster which ranges from 0 to 10 of lowest spatial cluster cholera diseases. This shows that Spatial Clusters inSabongari and Zaria LGAs are low clustered.

4.3.1 Cholera disease risk

Figure 4.12 to Figure 4.16 respectively shows the spatial pattern of cholera risk map in the area

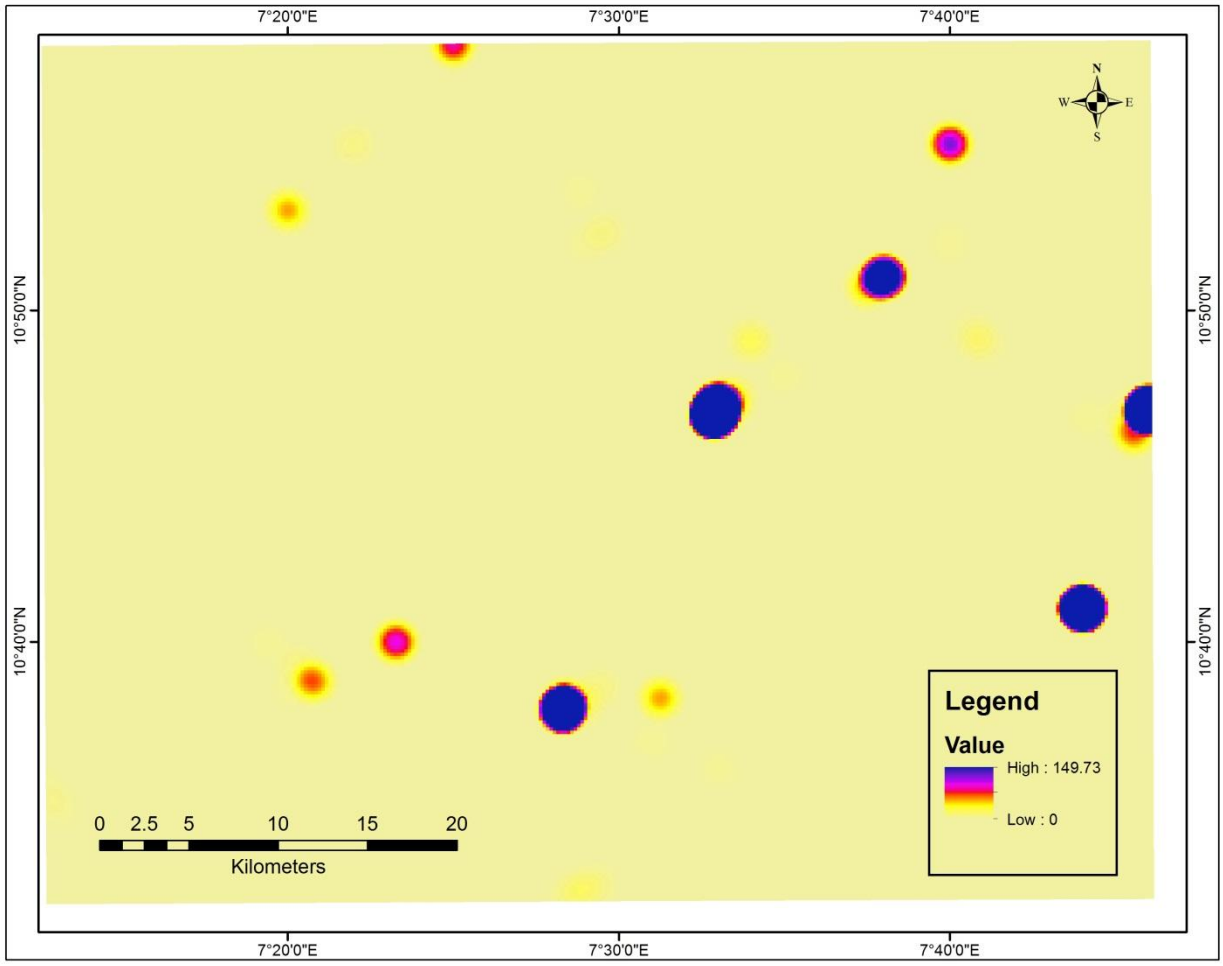

Figure 4.12: **Cholera Disease Risk in Igabi LGA**

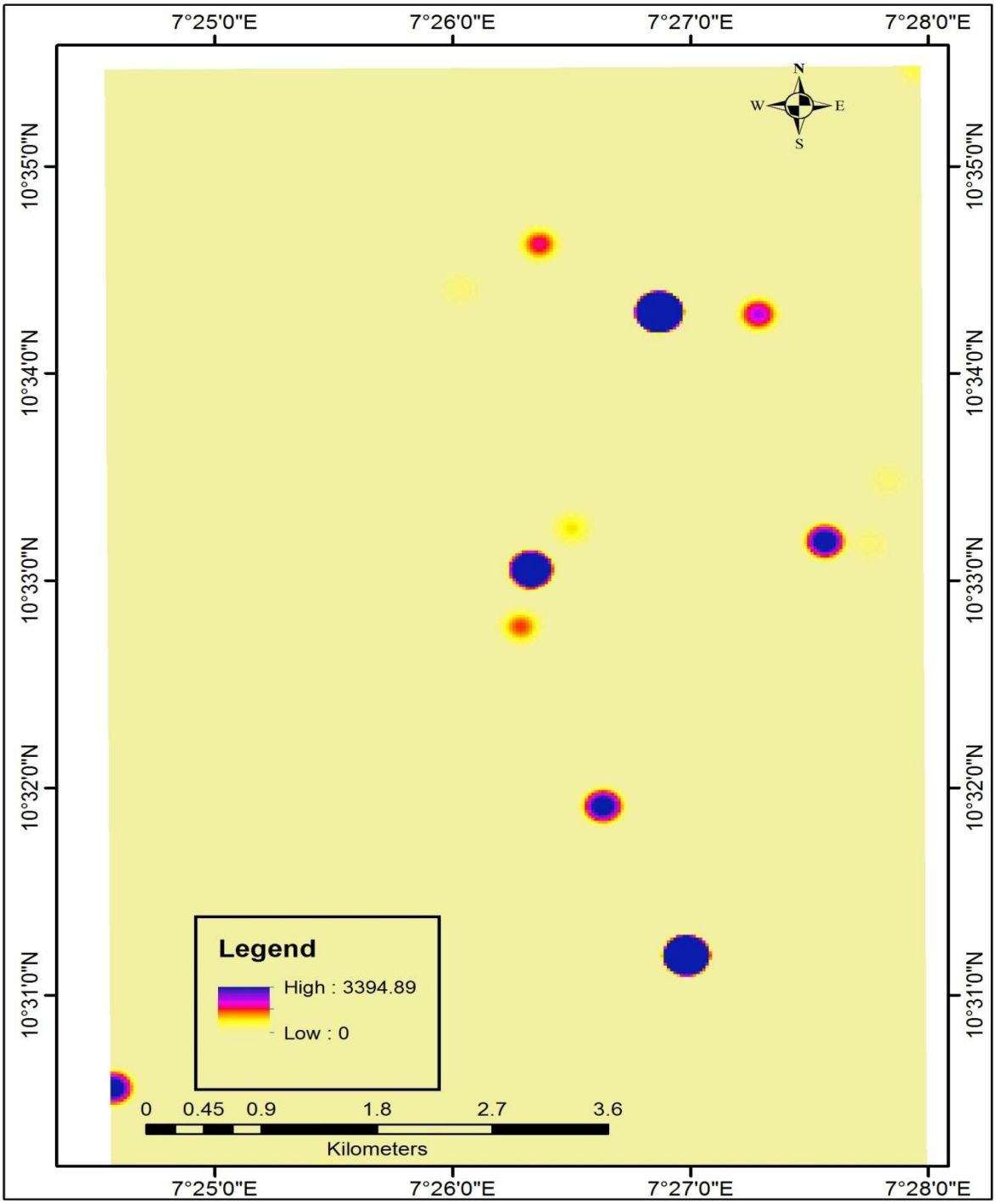

Figure 4.13: Cholera Disease Risk in Kaduna North LGA

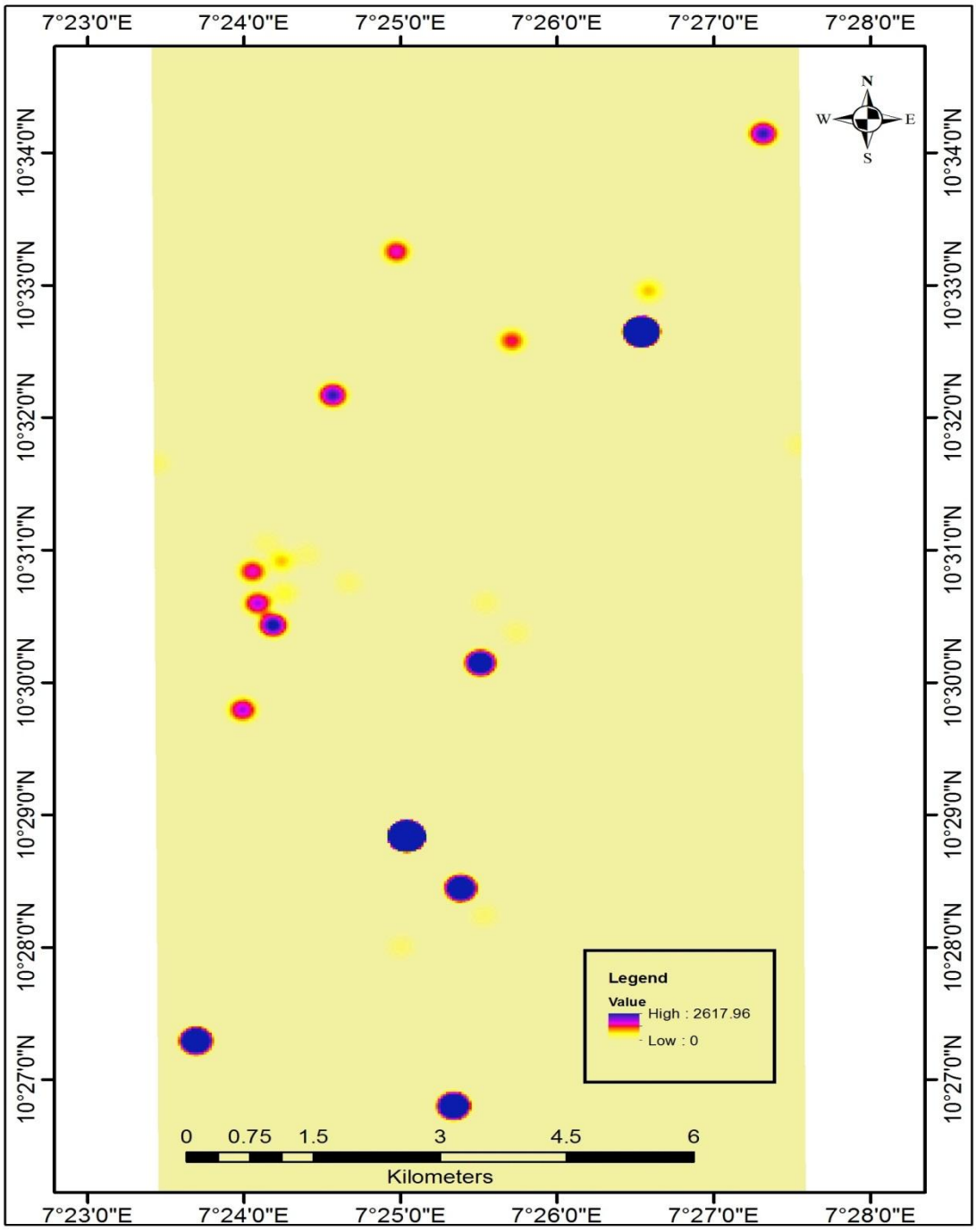

Figure 4.14: Cholera Disease Risk in Kaduna South LGA

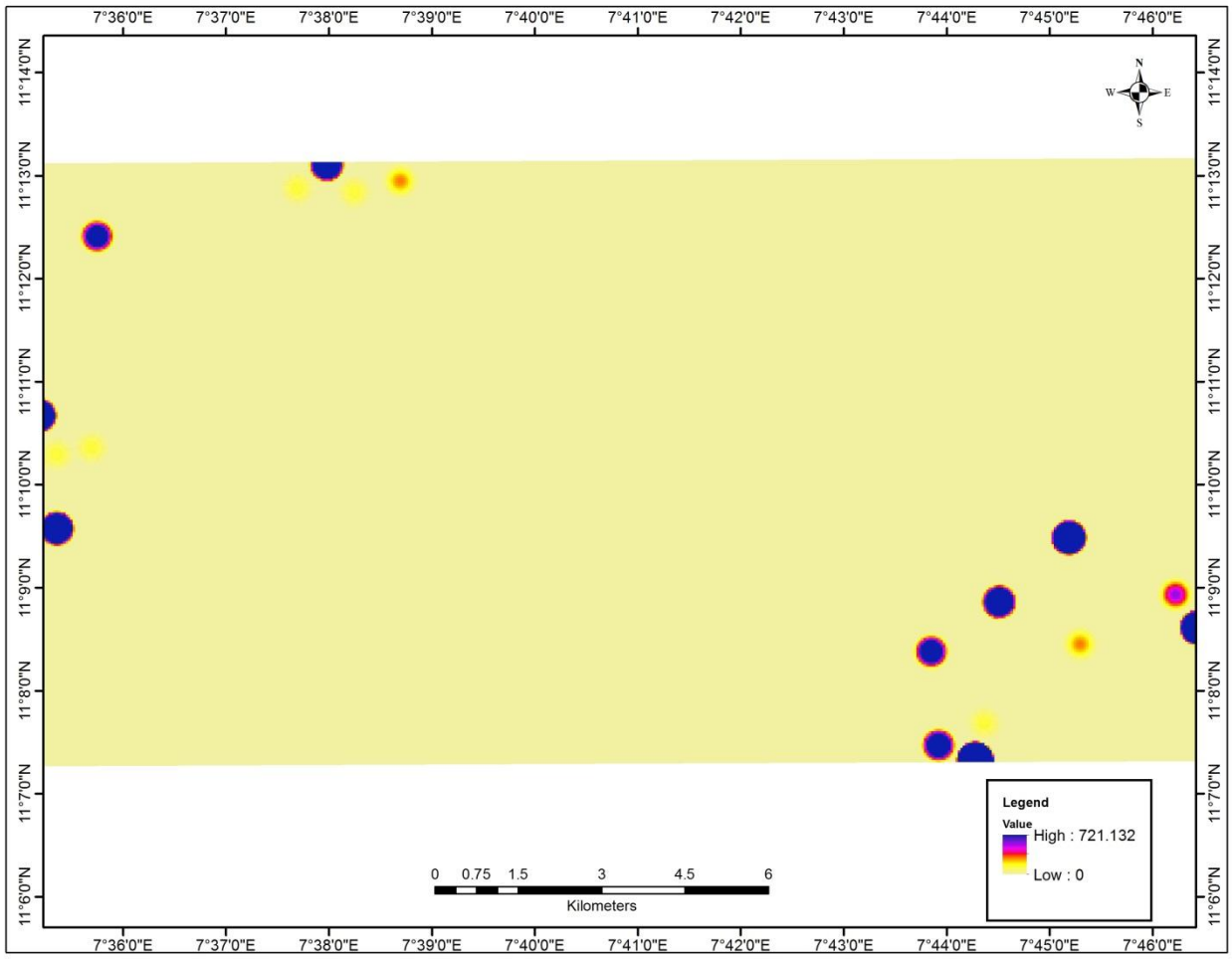

Figure 4.15: Cholera Disease Risk in Sabongari LGA

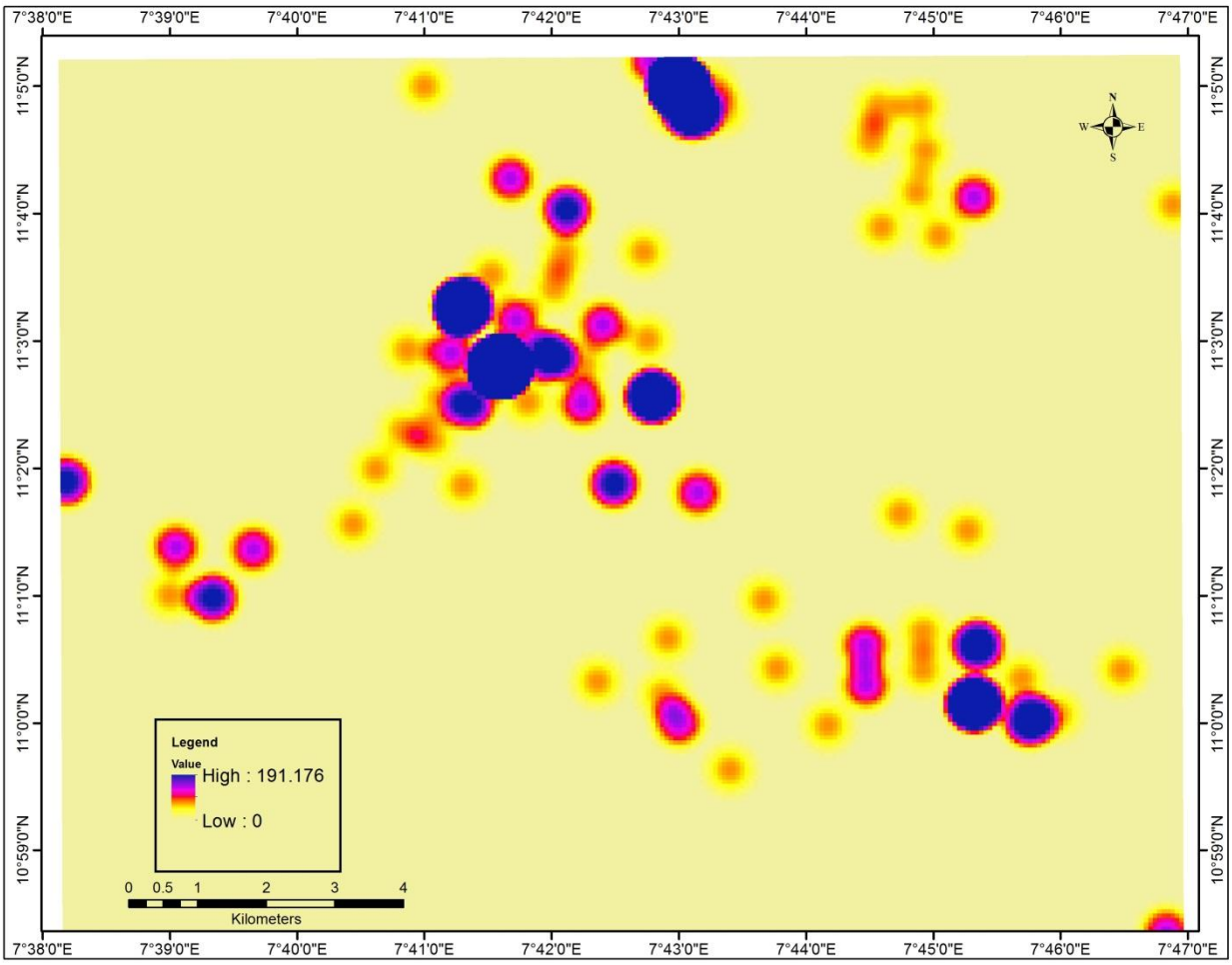

Figure 4.16: Cholera Disease Risk in Zaria LGA

As shown in Figure 4.12 to Figure 4.16, in order to evaluate the spatial pattern of cholera diseases risk, there is the need to calculate a magnitude per unit area from point features using a kernel function to fit a smoothly surface to each point. The larger values of the search radius parameter produce a smoother, more generalized density raster whereas the smaller values produce a raster that shows more detail.

From the results obtained in Figure 4.12 to Figure 4.16, it was found that all Local Governments in the study area areat risk of the Cholera disease outbreaks. The high concentration in cholera epidemic in the study area could be as a result of the rivers found in the community because the

research conducted by WHO (2008) pointed out that African Great Lakes have been suspected to play a role as a reservoir of the bacteria Vibrio cholerae (v.cholerea), while human infection and movement are probably involved in the propagation of the disease inland. Also, the discharge of fertilizers and chemicals as the result of agricultural activities into the tributaries (rivers) in the study area is one of the major determinant factors that lead to the cholera outbreak. This situation is similar to what happens during the 19th century spread of cholera from its original reservoir or source in the Ganges Delta in India to the rest of the world, before receding to South Asia (WHO, 2008).

4.3 Inventive Cholera Control Plan

Figures 4.8 to 4.11, have revealed that all the Local Government Areas in study should be given priority during the wet season for cholera control measures especially in Kaduna South. The areas that are vulnerable to the disease risk are shown in figures 4.12 to 4.16 and should be used as a decision making guide for cholera control plans in the study area especially. Finally, the GIS interface is an inventive cholera control decision making tool because the integration of GIS and epidemiological approaches are helpful tools to control the disease spatially and temporally as pointed out by Chang (2008). As such the outcome of this research serve as an innovative cholera control plan in five Local Government Areas of Kaduna State as supported by Chin-Lai (2009). These intervention includes;

1. Case Management,
2. Surveillance of the outbreak including laboratory analysis and data management,
3. The WASH activities such as hygiene promotion, safe burial, household disinfection, water treatment, etc.
4. Community mobilization,

5. Possible oral- cholera vaccine(OCV) campaigns.

Cholera intervention includes the concept of preparedness and long-term prevention. Preparedness aims at improving the readiness of agencies, government, civil societies and communities during outbreaks. Cholera is predictable, preventable and can ultimately be eliminated.However, the usefulness of public health activities including vaccinations and WASH interventions are the main context in reducing the risk of these disease outbreaks.Government and international establishments need to advocate that;

- Cholera hotspots in each area should be identified and become the primary target for interventions.
- Epidemiological and environmental findings should be mainstreamed into national policy.
- Comprehensive strategies for cholera elimination are established with national commitments and high level support from the health sector.
- Resources for cholera prevention measures program are mobilized in hotspots as a long-term investment for the benefit of public health and economic and social welfare.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

The purpose of this study is to carry outspatial analysis on cholera epidemic in five LGAs of Kaduna State with respect to the traditional method of Cholera data handling and management using archive in the State Ministry of Health Kaduna. To analyze the spatial pattern and cluster cholera epidemic in the study area, inventories provided by the Ministry of Health Kaduna state, coordinate of the locations of towns and localities were used for the analysis.

The methods employed include scanning the administrative map of the study area, digitized and update using the Google earth image 2016 in order to produce the up-to-date digital map in ArcGIS 10.3 environment and the R Statistical package respectively. It was found that there wereabout I87 locations of patients with Cholera disease cases in the study from the year 2010 to 2015. Zaria Local Government Area, has the highest number with about 56 locations which constitutes about 37.4% and Kaduna North with the lowest locations of about 16 patients (8.4%). There were about 1,363 cases of people with Cholera disease in Igabi LGA, 447 cases were recorded in Kaduna South, 378 in Kaduna North, 301 in SabonGari and 217 in Zaria which has the lowest cases of Cholera epidemics.

The results of spatial statistical analysis by multi-distance spatial cluster analysis (Ripleys K Function) revealed that the cholera diseases were clustered in Sabongari, Zaria, Igabi, and Kaduna South LGA's, random only in Kaduna North. This finally revealed that the Combine spatial pattern of cholera epidemic in the study area is Clustered.

It was also found that,Kaduna SouthLGA has the highest degree of spatial cluster, while Igabi Zaria and SabonGari LGAfalls within the low areas. It was also found that all Local

Governments in the study area are at risk from the Cholera disease outbreaks.The outcome of this research serve as an innovative cholera control plan in five Local Government area of Kaduna State as supported by Chin-Lai (2009).

Furthermore, the result reveals that GIS could provide a mechanism for data integration, management and output generation in its spatial environment. In this study, spatial distribution of cholera, patterns, cluster and their statistical significance were derived using geospatial technology. Various maps and layer designed for the study area provide valuable records (information) for the assessment, management, monitoring, development and decision on the map. A logical analysis has been translated into a series of what lies here and/or what lies there. The results received descriptive and analytical answers which re meant for health planners, managers and decision-making experts because the types of inference, deduction and conclusion made is left entirely with the health worker authorities.

5.2 Conclusion

The study reveals that GIS is a useful tool to epidemiologists to visualize, manage, and analyze large volumes of data. It can help to better define populations exposures with perhaps better specificity. The results shows the advantages of advances in technology that allows not only disease mapping but also the application of spatial statistical methods such as cluster analysis. The study determined the presence and extent of clustering in the study area. The spatial pattern of cholera mirror the spatial pattern of the population at risk. The association between population density and environmental exposure were also established.

5.3 Recommendation

It is recommended that surveillance should be intensified in all the four LGAs that have high incidence of cholera epidemic most especially during wet seasons. A further study can be carried out on the effects of socioeconomic and environmental factors on the pattern and risks of cholera.

5.3.1 Recommendation for further research

i. This research analyses the spatial pattern of cholera and there is the need for mapping spatio-temporal GIS visualization, and exploratory spatial data analysis for cholera epidemics in the study area and the whole of Kaduna State at large.

ii. This research did not look at the effects of socioeconomic factors on the pattern of cholera. There is the need for another research to look at the effects of socioeconomic factors on the spatial pattern of Cholera epidemics in the study area.

iii. This research did not study the effects of environmental factors on the spatial pattern of Cholera in the study area. There is the need also for a research on the effects of environmental factors on the spatial pattern of Cholera epidemics in the study area because research has shown that environmental factors hinders and triggers the spread of Cholera Epidemics especially in the developing countries.

iv. GIS base can be a powerful tool to direct ones attention to areas of interest and also help visualize diseases hotspots.

v. Geospatial techniques can be used to further identify and evaluate places where there are unusual concentration of diseases.

5.4 Contribution to Knowledge

i. This is the first research that employed the use of the Ripley's k-function method on Cholera epidemic in Kaduna State.

ii. Cholera epidemic data was used in GIS to highlight the previously unidentified disease pattern and clusters in the five Local Governments in the study area.

References

Ajoke, O. A., Solayide, A. O. N., Mary-Theresa, N. and Akitoye, O. C. (2012).Cholera Epidemiology in Nigeria: an overview. *Pan African Medical Journal*.

Akyala, I. A., Bright, E. S., Olufemi, A., Adebola, O. and Patrick, N. (2014). Investigation of Cholera Outbreak in an Urban North Central Nigerian Community-*The Akwanga Experience*. *Public Health Research*, 4(1): 7-12 .

Ali, M., Emch, M. and Sack, R. (2002); identifying environmental risk factor for epidemic cholera; a raster GIS- apporoach. *Health place* ,8: 201-210.

Ali M., Emch, M., Donnay J. P, Yunus M. and Sack R. B. (2002): Identifying environmental risk factors for endemic cholera: a raster GIS approach.*Health Place*, **8:**201–210.

Barreto, M. L. Crescimento e tendência da produçãoscientíficaemepidemiologia no Brasil. Rev SaúdePública. 2006; 40(N Esp), 79-85.

Bernstein, K.T., Curriero, F. C., Jennings, J. M., Olthoff, G., Erbelding, E. J. and Zenilman, J. (2004). Defining core gonorrhea transmission utilizing spatial data. Am. J. Epidemiol. 2004; *[PubMed*: 15229117], 160:51–58.

Borroto and Martinez, P. (1996); Geographical pattern of cholera in mexico; *internationalJournal for epidemiology*.29(4)764-2000.

Carpenter, C. (1971): Principle and practice of cholera control, *Ann. Intern. Med*, 74(6)1021-1050.

Chang, K. T. (2008). *Introduction to Geographic Information Systems* (New York:TheMaGraw

Hill Companies, Inc). 1-418.

Chin-Lai P, Mun-So, F. and Wing, C. (2009). *Spatial Epidemiological Approaches in Disease*

Mapping and Analysis (USA:CRC Press) ,35-49

Cholera vaccines, (2010): WHO position paper. ReleveEpidemiologiqueHebdomadaire/Section d'hygiene du Secretariat de la Societe des Nations = *Weekly Epidemiological Record/Health Section of the Secretariat of the League of Nations. WHO Press*, Geneva: World Health Organisation, 117-128.

Claude-Lyne, *et al* (2013). Predictors of Disease Severity in Patients Admitted to a Cholera Treatment Center in Urban Haiti. *American Journal of Tropical Medicine and Hygiene.* 89(4), 2013. 625–632.

Codeico, C. (2001); Epidemic and epidemic dynamics of cholera; the role of aquatic reservoir *BMC Infecticious disease*.

Colwell, R. R. and Huq, A. (1994): Environmental reservoir of Vibrio cholerae: Thecausative agent of cholera.*Ann N AcadSci*, 740**:**44–54.

Collins, A. E., Islam, M. S. and Williams, L. E. (2006); Socio economics and environmental origin of cholera epidemic in Mozambique; *guidelines for tackling uncertainty in infectious disease.*

Cressie, N. (1993); Statistics for spatial data 2nd edition.

Dickson, S. N., Julius, A., Peter, N. F., Elvis, A. T., Charlotte, W. A. and Odette, D. K.(2015). Assessing the risk factors of cholera epidemic in the Buea Health District of Cameroon. *BMC Public Health*, 15:1128.

Diego, R., Mercedes, P., Michael, E. and Mohammad, Y. (2010). Spatial clustering in the spatio-temporal dynamics of endemic cholera. BioMed Central Infectious Diseases, *International Journal of Health Geographics*, 10:51.

Diggle, P. J., Chetwynd, A. G. and Haggkvist, R. (1995). Morris SE. Second-order analysis of space-time clustering. *Stat. Methods Med. Res*; [PubMed: 7582201]; 4:124–136.

Dirk, P. (2008). Spatial analysis in epidemiology. *Oxford University Press*.

Doguwa, S.I. (1988): Statistical Analysis of Mapped Point Patterns, Ph.D Thesis, University of Essex, U.K.

Elliott, P. and Wartenbrg, (2004); spatial epidemiology current approaches and future challenges. *Environmental health perspective*, 112(9)998.

Emch, M. (1999); diarrheal disease risk in matlab, bagladish, *social science and medicine* 49(4)519-530

Frank, B. O. and Alfred, A. D. (2008). Spatial and demographic patterns of Cholera in Ashanti region –Ghana. BioMedCentral. *International Journal of Health Geographics*, 7-44.

Frank, B. O., Alfred, A. D. and Alfred, S. (2012).Bayesian structured additive regression modeling of epidemic data: application to cholera. *Medical Research Methodology* , 12-118.

Frank, B. and Osei, (2010). Spatial Statistics of epidemic data; the case of cholera epidemiology in Ghana.

Gesler, W. (1986). The uses of spatial analysis in medical geography: a review. SocSci Med 23-963.

Hinnman*et al (2006).*Spatial and temporal of typhoid outbreaks in Washington D.C, *International journal of Health Statistics.*

Hormick, R. B., Music, S. I. and Wensel R. *et al* (1971): The broad street pump revisited; response of volunteers to ingested cholera vibros. *Bull. NY Acad. Med.* 47(10)1181-1191.

Houben, M. P., Coebergh, J. W., Birch, J. M., Tijssen, C. C., Van Duijn, C.M. and McNally, R. J. (2006). Space-time clustering of glioma cannot be attributed to specific histological subgroups. *Eur. J. Epidemiol*. [PubMed: 16547834] 21:197–201.

Huang, L. and Pickle, L. W. (2008). Das B. Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. *Stat. Med.* 27:5111–5142.

Huq, A.,*et al* (2005): Criticalfactors influencing the occurrence of Vibrio cholerae in theenvironment of Bangladesh. *ApplEnvMicrobiol*, 48:420–424.

International Federation of Red Cross (IFRC, 2016). *International Federation of Red Cross and Red Crescent Societies*, Geneva.

Jason K. B.,*et al* (2014).Household-Level Spatiotemporal Patterns of Incidence of Cholera, Haiti. Emerging Infectious Diseases . *www.cdc.gov/eid*, 20.

Keams RA (1993). Joseph AE. Space in its place: developing the link in medical geography. *SocSci Med*, 37:17-711.

Knox, G. (1989). Detection of clusters. Methodology of Enquires Into Disease Clustering. London: Small Area Health Stat. 17-22.

Kuldorff, M. and Nagarwatta, N. (1995); Spatial disease clusters: detection and inference. *Stat. Med.* 14(8):799-810.

Kulldorff, M (2006). Tests of spatial randomness adjusted for an inhomogeneity: a general framework. *J. Am. Stat*. Assoc; 101:1289–1305.

Lauren, M. S. and Mark, V. (2010). Spatial statistics in ArcGIS: Handbook of Applied Spatial Analysis; Software Tools, Methods and Applications,*Springer-Verlag Berlin Heidelberg*.

Lawoyin, T. O., Ogunbodede, N.A., Olumide, E. A. A. and Onadeko, M. O. (1999). Outbreak of cholera in Ibadan, Nigeria. *Eur J Epidemiol.*15(4)367-70.

Lawson, A. B. (2001a). statistical method in spatial epidemiology. John Wiley and sons chick Chester.

Lipp EK, Huq A, Colwell RR: Effects of global climate on infectious disease: The cholera model.*ClinMicrobiol Rev*, 15**:**757–770.

Luque, F. M. A., Schomaker, M., Mason, P.R. (2012). Elevation and cholera: an epidemiological spatial analysis of the cholera epidemic in Harare Zimbabwe. *BMC Health*, 12-442.

Mahalanabis, D. and Sack, D. (2005). Clinical management of cholera. *In D. Barea and W Greenough, III (Eds) cholera plenum medical Book Company*, New york; 253-283.

Matthews, S. A. (1990). Epidemiology using a GIS: the need for caution. Comput Environ Urban Syst;14:213-21.

McNally, R. J., Rankin, J., Shirley, M. D., Rushton, S. P. and Pless-Mulloli, T. (2008). Space-time analysis of Down syndrome: results consistent with transient pre-disposing contagious agent. *Int. J. Epidemiol,*37:1169–1179.

Mukandavire, S. Z., Liao, J. W., Gaff, H., Smith, D. L. and Morris, J.G. (2011) .*Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe*, Proc. Natl. Acad. Sci. 108, 8767–8772.

National Population Commission. (2006). *Population and Housing Census*: Class Size of Household. Abuja, Nigeria.

Nelson, E. J., Harris, J. B., Morris, J. G., Calderwood, S. B. and Camilli, A. (2009). *Cholera transmission: The host, pathogen and bacteriophage dynamics*, Nat. Rev. Microbiol. 7, 693–702.

Nene, S.A., Nayar, S.K. (1996). A Simple Algorithm for Nearest Neighbor Search in High Dimensions. *In IEEE Trans. on Pattern Analysis and Machine Learning*, 18, 8 989

Ohser. J, Stoyan, D. and Ripley, B.D. (1985). On the efficiency of edge correction for spatial point processes. *Biometrical Journal* 23, 522-533.

Osei, F. B. and Duker, A. (2010). Spatial dependency of cholera prevalence on potential reservoir in an urban area, Kumasi Ghana; *international journal of applied earth observation and geo information.*

Paul, F. H., *et al* (2011). Spatio-temporal epidemiology of the cholera outbreak in Papua New Guinea, 2009–2011. *BMC Infectious Diseases*, 14:449.

Public Health Agency of Canada (2008). *GIS for Public Health Practice*.

Ripley B.D. (1976). The second order analysis of stationary point process. *Journal of applied probability, 13, 255-266.*

Ripley B.D. (1977). Modeling spatial patterns (with discussion). *Journal of the Royal statistical Society B 29, 172-212.*

Ripley, B.D. and Kelly, F.P. (1977). Markov Point processes. *Journal of the London Mathematical Society* 15, 188-192

Rita, C. (1996). Global climate and infecticious disease, the cholera paradigm. Science 274(5295:2025-2031). prevention and control; *international journal for environmental studies* 63(5);537-549.

Ruiz-Moreno D., Pascual M., Emch, M. and Yunus, M. (2010): Spatial clustering in the spatio-temporal dynamics of endemic cholera.*BMC Infect Dis*,10**:**51.

Saddique, A. K., Zaman, K., Sack, R, B. and Islam, (1996): cholera epidemic in Bangladesh.

Saulat, J. (2016).Cholera Epidemiology, Prevention and Control. Prevention and Control of Food Related Diseases.

Schmiedel, S., Blettner, M., Kaatsch, P. and Schuz, J. (2010). Spatial clustering and space-time clusters of leukemia among children in Germany, 1987–2007. *Eur. J. Epidemiol*, 25:627–633.

Shapiro, R. L., Musa, R. O. and Adock, P. M. *et al* (1999): Transmission of epidemic vibrio cholera. *Am. J. Trop. Med. Hyg*. 60(2):271-276.

Snow, J. (1855); On the mode of communication cholera. *2nd edition. John churchhill*.

Steven, C. L. C., Kim, D. R., Ali, N. H., Douglas, O. F. and John, C. (2011). Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics*, 10:35.

Tamplin, M. L., Gauzens, A. L., Huq, A., Sack, D. A. and Colwell, R. R. (1990): Attachment of Vibrio choleraeserogroup O1 to zooplankton and phytoplankton of Bangladesh waters.*ApplEnvMicrobiol*, 56.

UNICEF (2015). Cholera outbreak in the West and Central Africa: Regional Update, 2015 - WEEK 37. http://reliefweb.int/report/nigeria/cholera-outbreak-west-and-central-africa-regional-update-2015-week-37-enfr.

WHO, (1993): Guidelines for cholera control. World Health Organisation, Geneva. *http:/www./helid.desastres.net/en/d/jwho.90e*.

WHO, (2001): Cholera, 2000. Weekly epidemiological record 76(31):233-240. *http://www.who.int/wer.*

WHO, (2001):Weeklyepidemiologicalrecord,76(31)233-240 *http://www.who.int/wer/2010/wer853.pdf*

WHO, (2010): Outbreak news. Cholera, Haiti - update**.** *Weekly Epidemiol Rec*,85:489–490.

World Health Organization, (WHO, 2013). http://www.who.int/csr/don/archive/ disease/cholera/en/.

APPENDIX

```
Gph<- function(D,w){
  if(is.matrix(D)==TRUE &is.matrix(w)==TRUE){

    a<-sqrt(sum((D[2,]-D[1,])^2))
    b<- sqrt(sum((D[3,]-D[2,])^2))
    A<- a*b
    d<- sqrt(sum((D[3,]-D[1,])^2))
    t<- seq(from=d/100,to=d/4, length=50)
    Pi<- acos(-1)
    x0<- D[1,1]; y0<- D[1,2]; x1<- D[3,1]; y1<- D[3,2]

Kk<- function(w,t,D){
      n<- 1:nrow(w)
wij<-0
      x0<- D[1,1]; y0<- D[1,2]
      x1<- D[3,1]; y1<- D[3,2]

      for(i in n){
        for(j in n[-i]){

          x<- w[i,1]
          y<- w[i,2]
rij<-sqrt(sum((w[i,]-w[j,])^2))
          if(rij<=t){
            # At the boundary of the window
            # And at the corner points of the window
            if(x==x0){
              if(y==y0 | y==y1) theta<- 3/2*Pi
              if((y1-y)>=t & (y-y0)>=t) theta<- Pi
            }
            if(x==x1){
              if(y==y0 | y==y1) theta<- 3/2*Pi
              if((y1-y)>=t & (y-y0)>=t) theta<- Pi
            }
            if((x-x0)>=t & (x1-x)>=t){
              if(y==y0 | y==y1) theta<- Pi
            }

            # For points inside the window
            if(x>x0 & (x-x0)<t){
              if(y>y0 & (y-y0)<t) theta<- Pi/2 + acos((x-x0)/t) + acos((y-y0)/t)
              if((y-y0)>=t & y<=(y1-t)) theta<- 2*acos((x-x0)/t)
              if(y>(y1-t) & y<y1){
                theta<- Pi/2 + acos((y1-y)/t) + acos((x-x0)/t)}
            }
```

```
          if((x-x0)>=t & x<=(x1-t)){
            if(y>y0 & (y-y0)<t) theta<- 2*acos((y-y0)/t)
            if((y-y0)>=t & y<=(y1-t)) theta<- 0
            if(y>(y1-t) & y<y1) theta<- 2*acos((y1-y)/t)
          }
          if(x>(x1-t) & x<x1){
            if(y>y0 & (y-y0)<t) theta<- Pi/2 + acos((y-y0)/t) + acos((x1-x)/t)
            if((y-y0)>=t & y<=(y1-t)) theta<- 2*acos((x1-x)/t)
            if(y>(y1-t) & y<y1) theta<- Pi/2 + acos((x1-x)/t) + acos((y1-y)/t)
          }
wij<- wij + 2*Pi/(2*Pi-theta)

        }
        else wij<- wij
      }
    }
wij
  }
ysam<- array(data = NA, dim = length(t))
  for(i in 1:length(t)){
    K<- A*Kk(w,t[i],D)/(nrow(w)*(nrow(w)-1))
ysam[i]<- sqrt(K/Pi)
  }
  #library(splines)
  #ty1<- smooth.spline(t,ysam)

  forest<- matrix(data = NA, nrow = length(t),ncol = 101)
sim.w<- matrix(data = NA, nrow = nrow(w),ncol = 2)
sim.w[,1]<-runif(n=nrow(w),min = x0, max = x1)
sim.w[,2]<-runif(n=nrow(w),min = y0, max = y1)
  for(i in 1:length(t)){
    K<- A*Kk(sim.w,t[i],D)/(nrow(w)*(nrow(w)-1))
    forest[i,1]<- sqrt(K/Pi)
  }

  for(j in 2:99){
sim.w[,1]<-runif(n=nrow(w),min = x0, max = x1)
sim.w[,2]<-runif(n=nrow(w),min = y0, max = y1)
    for(i in 1:length(t)){

      K<- A*Kk(sim.w,t[i],D)/(nrow(w)*(nrow(w)-1))
      forest[i,j]<- sqrt(K/Pi)
    }
  }

  for(i in 1:length(t)){
    forest[i,100]<- min(forest[i,1:99])
```

```
    forest[i,101]<- max(forest[i,1:99])
   }


   plot(t,ysam,type='l',xlab = 't',ylab = 'L(t)',col=1,lwd=2,
        main = deparse(substitute(w)))
   #ylim=range(0,max(ysam)))

   #for(i in 1:99){
   #ty<- smooth.spline(t,forest[,i])
   # lines(t,forest[,i])
   #}

   lines(t,lty=2,forest[,100],lwd=2,col=2)
   lines(t,lty=2,forest[,101],lwd=2,col=2)
   legend.txt<-c("Lower & Upper Envelops","Point Pattern")
   legend("topleft",legend = legend.txt,col = 2:1,lty=2:1)
  }

  else stop("D must be a matrix i.e. a set of coordinates\n
w must also be a matrix i.e. a set of coordinates for locations")
}
```