GIS-BASED SPATIAL EPIDEMIOLOGY OF THE TOP FIVE INFECTIOUS

DISEASES IN GHANA

By

SYLVESTER DODZI NYADANU

(BSC. MOLECULAR BIOLOGY AND BIOTECHNOLOGY)

A Thesis submitted to the Department of Computer Science

Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

HEALTH INFORMATICS

Department of Computer Sciences

Institute of Distance Learning

November, 2016

DECLARATION

I hereby declare that this submission is the true findings of my own researched work presented towards an award of a second degree in MPhil Health Informatics and that, to the best of my knowledge, it contains no material previously published by another person nor submitted to any other University or institution for the award of degree except where due acknowledgement has been made in text .However, references from the work of others have been clearly stated.

Sylvester Dodzi Nyadanu (PG8954413)   ………………
……………...

Student Name & ID   **Signature**   **Date**

Certified by:

Dr. Harry Tagbor   ……………..   ………………….

Academic Supervisor   **Signature**   **Date**

Certified by:

Dr. J. B. Hayfron-Acquah   ………………
…………………

Head of Department   **Signature**   **Date**

DEDICATION

Dedicated to my wife, Rachel L. Garglo, and children, Rasyl Nunana and Lislyn Nuseli
Nyadanu

ACKNOWLEGEMENTS

assisted and encouraged me in some way. Am so thankful to my wife, parents and siblings for their emotional and economic support to bring me to this height. To Mr and Mrs Kordorwu David, thank you very much for your prayers and encouragement in shaping my life holistically. On the whole, I will eternally be appreciative to my God who never fails me but always makes complexity simplicity and impossibility possibility in my life. Thank you Father.

ABSTRACT

Infectious diseases are general burdens in developing countries but within a country or region, specific areas can be experiencing elevated rates of infection. To address this and support public health policy, Geographic Information System (GIS) and other spatial techniques are used extensively in geo-health informatics.This study explored GIS-based spatial epidemiology to describe spatial patterns, identify hotspots and assessed the determinants responsible for spatiotemporal variations of infections in Ghana. The outpatient department data of Malaria, URTI, Diarrhoeal diseases, Intestinal worms and Typhoid fever from 2010-2014 were obtained from

Ghana Health Service, health determinants processed from 2010 census and geo-coded per the 170 districts to create the GIS database. Incidence rates were estimated, smoothen and spatially analysed. The rates of the infectious diseases persistently showed significant spatial dynamics with different intensities (Moran's I > 0 and Z-score > 1.96, p< 0.05). The disease and yearlyspecific hotspot analyses pinpointed a number of hotspot districts. The divergence in the intensities of clustering and the number of hotspots observed undoubtedly showed location and disease-specific disparities in health intervention programmes, resources allocation and/or spatial variability of determinants across the country. There is a critical need to further strengthen/improve and prioritise disease-specific control strategies in the identified hotspot districts and to also give broader and uniform attention to other leading infectious diseases, especially Intestinal worms and Typhoid fever. From the spatial lag regression model, policy and/or social interventions to increase basic education attainment but decreasing urbanisation, intermigration and sex ratio while considering neighbourhood-effects could reduce the rates of infections in the districts.

Geo-health informatics should be included into health researches and reports in Ghana as data mining tool to enhance public health promotion and health policies.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

Abbreviation	Full Meaning
AICc	corrected Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
CDC	Center for Disease Control
CHIM	Centre for Health Information and Management
DALY	Disability-Adjusted Life Years
DBMS	Database Management System
DR	Dependency ratio
DCU	Disease control units
DSS	Decision Support Systems
EBS	Empirical Bayesian smoothing
EPR	Employment-to-population ratio
ESRI	Environmental System Research Institute
GBD	Global Burden of Diseases
G/SDDS	Geographical/Spatial Decision Support System
GDHS	Ghana Demographic and Health Survey
GHS	Ghana Health Service

LH..outliers of Low-High values

LISA...Local Indicators of Spatial Autocorrelation

LL..clusters of Low-Low values (Coldspots)

MMDAs...Metropolis, Municipal and District
Assemblies

MOH..Ministry of Health

NCDs...Non-communicable Diseases

n.d...No Date

NTDs...Neglected Tropical Diseases

OLS...Ordinary Least Square regression

OPD...Outpatient Departments

OI..Overcrowding Index

PHC...Population and Housing Census

RR..Relative Risks

SPSS..Statistical Package for Social Sciences

TB..Tuberculosis

WHO..World Health Organisation

UNICEF...United Nations Children's Fund
URTI..Upper Respiratory Tract Infection

US...United States

UTM..Universal Transverse Mercator

VIP..Ventilated Improved Pit

WC...Water closet

LIST OF SYMBOLS

Symbol   Meaning

i.j   two areal units (districts)

i.i.d   independent identically distributed errors

W_{ij} spatial weight neighbourhood matrix corresponding to the observation of district

                       pair i, j

X_{it} morbidity counts of the reported infectious disease for i-th district in the year t

P_{it} mid-year population for t

R_{it} incidence rate per year for the i-th district in the year t

n   number of weights (number of districts)

R   smoothed cumulative rates

$\overline{R}$ mean incidence rate of infection

Σ summation

I_M global Moran's Index

I_i local Moran's Index

p   probability value (significance level)

ε vector of random (spatially autocorrelated) error terms/residuals

ρ spatial autoregressive coefficient for spatial lag model

λ spatial autoregressive coefficient for spatial error model

β coefficients estimated parameters in the spatial regression model

Gi*   Getis-Ord General G-statistics for hotspot analysis

z   z-statistics value

LIST OF APPENDICES

1.1 BACKGROUND OF THE STUDY

1.1.1 The basis of Infection

Infectious diseases accounts for 30% of the global burden of diseases and one-quarter of all deaths (Sen and Bonita, 2000). Infectious diseases are caused by transmission of specific pathogenic agents to a susceptible host. The pathogens are categorised as viruses, bacteria, fungi, protozoan and parasitic worms. Examples of infectious diseases are malaria, typhoid, tuberculosis, diarrhoeal diseases, pneumonia, HIV/AIDS, anthracis, Ebola and so on. Transmission of pathogens involve direct contact which occur from host to host, some involve near-direct contact (such as pathogens excreted by one host and inhaled or consumed by another), whereas others rely on an arthropod vectors or intermediate hosts (Ostfeld *et al*, 2005; US CDC, 2012). If direct infection is from an infected humans or animals, it is described as contagious or communicable, otherwise anthroponotic infection (infection spreads through inanimate such as air/waterborne particles or vehicles). Zoonotic infections, accounting for nearly 70% of the emerging infectious diseases, are infections of lower vertebrates that are transmissible between animals or from animals to humans (Saker *et al*, 2004: Potter, n.d). From the epidemiological triangle theory, infection results from the interaction among the agent, host, and environment. The agent refers to the pathogens, physical and chemical factors that cause diseases. The environment includes physical factors such as geology and climate; biologic factors such as vectors and intermediate
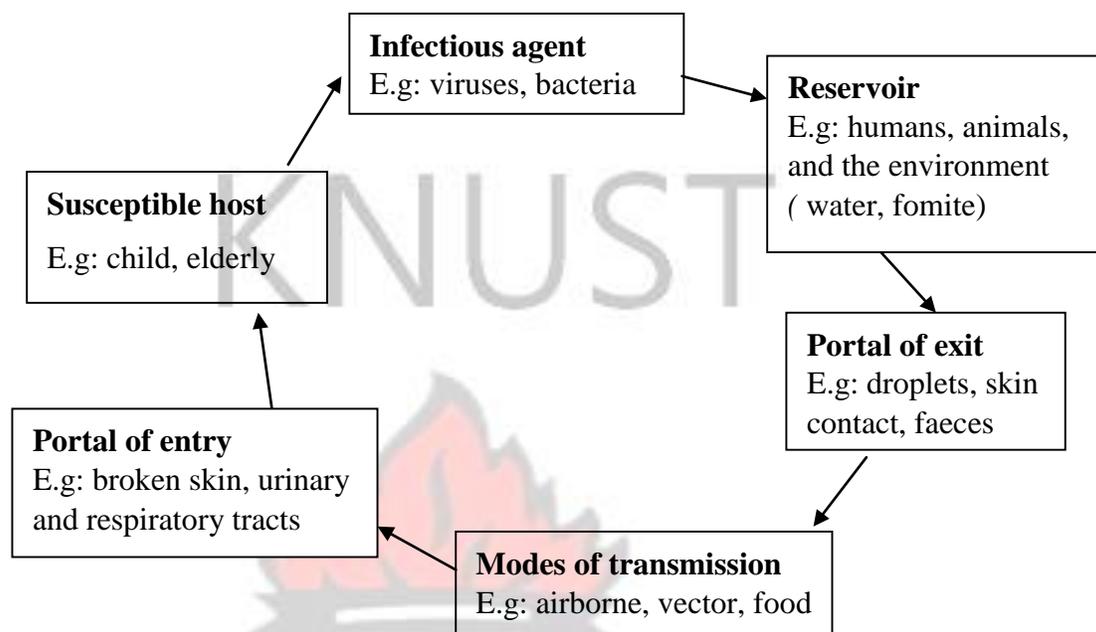
hosts, technology and socioeconomic factors such as overcrowding, sanitation, water quality, air pollution and the availability of health services. Most of these factors are promoted by human activities and have strong influence on the emergence and/or reemergence of infectious diseases (Santana, *et al*, 2011; Bonita *et al*, 2006; Rajiv *et al*, 2003). The cycle or the chain of infection (Fig 1.1) describes the processes leading to the causation of infectious diseases (WHO, 2001).

Fig1.1: Cycle of Infection (US CDC, 2012)

Knowledge of the reservoir, sources, transmission and multiplication processes or the analysis of the causal chain offers many entry points for intervention (Bonita *et al*, 2006). Bhutta *et al*, (2014) reported five public health strategies promoted by WHO to control, eliminate, and eradicate infection. This includes preventive chemotherapy; innovative and intensified disease-management; vector control and pesticide management; provision of safe drinking water and hygiene, and

education; and veterinary and public health services. The usual intervention strategies to infectious diseases (IDs) are to

- Eliminate sites of pathogens, vectors and intermediate hosts
- Reduce human exposure to pathogens or vectors: this includes regulating the natural and built environments, using protective materials like ITNs and chemoprophylaxis behaviour modification and isolation.

- Boost the immune response of humans against pathogens through vaccination

- Treat the infected person (Saker *et al*, 2004).

1.1.2The Burden of Infectious Diseases

The health fact sheet on Ghana (WHO, 2010a) reported that 73% of years of life lost are due to communicable diseases. Even-though non-communicable diseases (NCDs) are also becoming serious burden of public health, IDs still remains the most reported cases of morbidity and mortality, leading to loss of most productive years mostly in children and young adults (Saker *et al*, 2004). Over the years, IDs such as malaria, tuberculosis (TB), HIV/AIDS, neglected tropical diseases (NTDs: a group of 17 bacterial, parasitic protozoal and viral infections) and nutritional disorders are major causes of disability and mortality. With increasing NCDs, developing countries are currently battling with a double burden of disease where there is co-existence of infectious diseases and NCDs. Globally, a death increase of 111,000 were attributed to malaria and NTDs in the last two decades (Bhutta *et al*, 2014). The epidemiological profile of Ghana on NTDs identified lymphatic filariasis, schistosomiasis, soil-transmitted helminthiasis, onchocerciasis, trachoma and loa-loa as endemic in Ghana (WHO, 2010b). The main infectious diseases of

high mortality are   acute respiratory infections, such as pneumonia, killing

approximately 2 million persons per year; diarrhoea (caused by salmonella and

cholera), killing approximately 1.5 million per year; measles (about 1 million); and

malaria (about

1 million), whooping cough, rabies and dengue fever (GBD, 2014). In the 2004
bulletin of health information in Ghana, malaria ranked first, accounting for 44.7%
of the total outpatient visits; followed by upper respiratory tract infections (6.8%),
skin diseases (4.3%) and diarrhoea accounting for 4.2% (Adams *et al*, 2004).
Globally, Africa alone accounts for 9 out of every 10 child death each due to malaria
and AIDS, and half of the global child deaths attributed to diarrhoeal diseases and
pneumonia. The leading cause of death in the developing countries include
pneumonia followed by heart diseases, diarrhoea, HIV/AIDS and stroke but heart
disease, followed by stroke, lung cancer, pneumonia and asthma or bronchitis in
the developed countries (WHO 2009). Global deaths from HIV/AIDS and malaria
have increased to 1.5 million and 1.17 million respectively in 2010 but was 1.1
million for malaria in 2012. The NTDs, however, accounted for 152,000 deaths
while 1.2 million deaths were attributable to TB in that same year. The NTDs, HIV,
TB and malaria are as a group referred to as 'infectious diseases of poverty (IDoP)
which are primarily concentrated in rural populations with geographic overlap,
resulting in high levels of co-infections (Amo-Adjei, 2013). Thus IDoP
disproportionately affect the poorest population in the world, contributing to a cycle
of poverty due to decreasing productivity resulting from prolonged illness,
disability and social stigma (Bhutta *et al*, 2014). As a group, IDoP, including
diarrhoeal diseases and respiratory infections are amongst the top five causes of
premature death and disabilityadjusted life years (DALYs). The consequences of
IDoP are enormous, ranging from impaired physical growth and cognitive
development in infants to huge economic burden on the individual, families, society
and the government(GBD, 2014).

1.1.3 The Geographic Information System (GIS)

A Geographic Information System (GIS) is a computer mapping technological tool

that organises and processes geo-referenced data from varied sources and integrates

them into an interactive map project to aid making intelligent decisions. GIS is

therefore considered as a geographical decision support system (GDSS). The

thematic map with table of content facilitate the addition of layers of information of the real-world locations, often with the aid of Global Positioning System (GPS), describing events occurring at places on the earth's surface (Jebara, 2007). The GIS execution stage usually consists of the relational database creation, including a complete set of geo-referenced data, and the elaboration of user interface for a spatial analysis (Rayed, 2012). The systematic phases of GIS implementation are

Data Capture ──► Data Management ──► Data Manipulation and Analysis ──► Data ►
Presentation

The output of GIS analysis communicate the results of the geovisual data analysis as maps (Otto and Rolf, 2009). All geo-referenced data are stored in the database management system (DBMS) of the GIS as a geo-relational model (or dual architecture) for geovisual analysis of the real world. The operational components of GIS are shown in Fig 1.2. These components, relating in a hierarchical way are implemented distinctively in the GIS framework. In the spatial data analysis, the areas stored in GIS are represented by a closed polygon with attributes stored in a relational DBMS table to form the geo-relational model (Camara *et al*, n.d, Otto and Rolf, 2009). GIS therefore allows a dynamic link among databases to enhance integrating existing data systems with spatial data and maps, enabling updates to be automatically reflected on the maps for the real space-time analysis of situations (Kaushal and Johnson, 2003; Hansen and VanOsdel, n.d). Generally, GIS software has useful space-time data analysis tools for spatial epidemiological studies (Verma and Gupta, 2014).

5
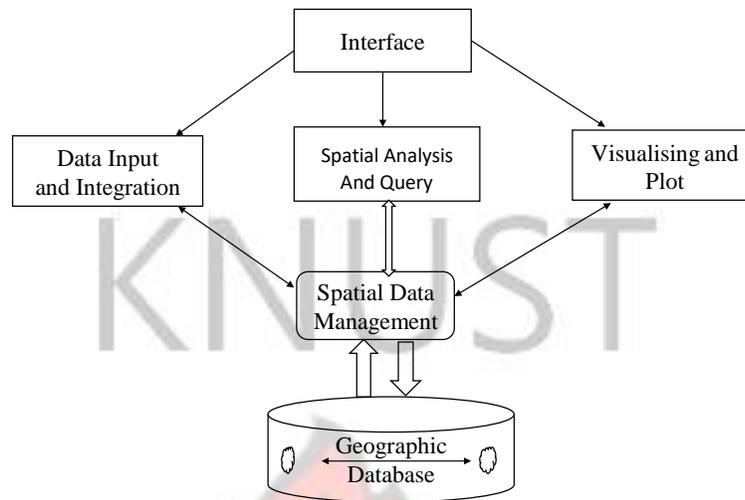
Fig 1. 2 The operational architecture of GIS (Source: Câmara *et al*, n.d)

1.1.4 Spatial Epidemiology: a GIS-aided epidemiology

1.1.4.1 Epidemiology and Spatial Epidemiology

Epidemiology studies the distribution and determinants of health-related states or events in specific populations, and then applies it to the control of health problems. Systematically, epidemiological studies involve counting cases or events, determining frequencies (rates) and then comparing the rates. Among the several epidemiological study designs, observational (descriptive and analytic) studies are mostly used to describe diseases or health-related events and health impacts (US CDC, 2012). While descriptive study describes occurrence and distribution of the outcome in terms of three major variables: people, place and time (Adetokunbo and Herbert, 2003), analytic study describes association between exposure and outcome in order to elucidate the determinants. Contrary to the observational study, interventional (experimental) study is designed to test a hypothesis by modifying an exposure within the study population with random assignment of exposure.

6

However, where experimental studies are infeasible due to practical and ethical reasons, observational studies based on existing exposure conditions are mostly used. Despite some challenges, observational studies in the investigation of determinants of diseases provide frameworks for interpreting most spatial analyses of public health data (Waller and Gotway, 2004). Within the two broad categories (observational and interventional studies), epidemiological studies are further organized into other forms depending on the mode, how, time, and other factors of data collection and analysis (US CDC, 2012).

It is very obvious that the location of people can be of great importance in identifying patterns of diseases to highlight the factors governing the spatio-temporal dynamism of diseases (Waller and Gotway, 2004; Ostfeld *et al*, 2005). Spatial Epidemiology is a GIStechnology aided intersection of three fields of study viz Epidemiology, Statistics/Biostatistics and Geography/Geology (Waller and Gotway, 2004; Shaddick, n.d). Spatial Epidemiology emerged as a principal scientific discipline to understand the causes and consequences of spatial heterogeneity in infectious diseases (Ostfeld *et al*, 2005). Spatial epidemiology is the description and analysis of the geographic or spatial variation in diseases with respect to background features such as socio-economic, demographic, environmental and climatic factors and many others. This means that spatial epidemiology studies the space-time distribution of diseases and their relationships to potential risk factors. The ability of GIS to provide spatial relational database, containing data from diverse sources and the spatial statistical functionalities and then generate thematic maps of the areas of interest is the basis of spatial

7

epidemiological studies (Basommi, 2011; Kolivras, 2006). The core purposes of spatial epidemiology are to

- evaluate differences in rates observed from different geographical areas

- smoothen the rates to account for heterogeneity

- identify disease clusters and unusual disease distribution

- assess the significance of potential exposures

- model the spatial relationship between exposures and rates

Fundamentally, the knowledge of space-time patterns of diseases and spatial structures characterisation is critical for better understanding of how populations interact with the environment or background features. Presently, GIS is the most useful technology in mapping diseases because of its capability to link health data of interest to all kinds of geo-referenced background variables such as socio-demographic and economic data, environmental and census data and so forth (Mika, 2004). Spatial epidemiologists were traditionally just using maps when analysing associations between location, environment, and disease (Rai *et al*, n.d). This is because diseases and other health events are related to some spatial exposures which are difficult to be linked and presented in the spreadsheet. But with GIS technology, health reports are presented with the spatial analysis components, and may therefore be perceived to be adequate for either epidemiological or strategic healthcare management purposes (Dhama *et al*,2013). With its ability to integrate data from different sources, mapping and geovisual analytical functionalities, spatial epidemiology fostered by GIS technology is potentially powerful resource for

8

health policy development and public health promotion (Buckeridge *et al*, 2002; US CDC, 2012). The applications of GIS in public health and epidemiology proved to be helpful in researches to generate new knowledge in public health, to design focused interventions (based on locating high prevalence areas and populations at risk), management, planning, monitoring and evaluation of health intervention programmes, dissemination of information, decision making and advocacy for optimal resource allocation (Rajiv *et al* 2003; Mika, 2004). Several researches has justified the use of decision support systems (DSS), including geographical decision support systems (GDSS) as an Information Communication Technology (ICT) tool to simplify and assess complex problems in health planning. GIS is considered as the central spatial component and as an innovative method to implement DSS due to the computational, analytical, problem solving and visualisation capabilities (Gudes *et al*, n.d) for a broader understanding to almost every community member. The most important public health reason for investigating the spatial distribution of infectious diseases and in searching for infectious agents is to help guide disease prevention and control strategies at different locations through practical ecological approach with the aid of GIS/GPS technology

(Mika, 2004; Jenkins, 2003).

1.1.4.2 Spatial Statistical Methods in Spatial Epidemiology

Starting with disease mapping by John Snow in 1885, technological advancement such as GIS has added spatial statistical analyses components to epidemiology. Spatial statistics involves the statistical analysis of observations with associated

locations in space in which there is application of statistical theory and techniques to geospatially-referenced data (Krivoruchko *et al*, 2003; Waller, 2007). Spatial statistical analysis follows the Tobler's first law of geography which states that: "*Everything is related to everything else, but near things are more related than far things*" (Scott and Janikas, 2010; Dhama *et al*, 2013). This implies the correlation of a variable with itself through space, known as spatial interaction or autocorrelation. Thus the basis of statistical analysis of dependency and heterogeneity in spatial patterns is to measure the degree of spatial association

(spatial autocorrelation) among data in relation to locations (Ord and Getis, 1995; Getis, 2004).The spatial analysis involves descriptive stage concerned with description and mapping of health-related states followed by the spatial analytical stages to account for the observed spatial patterns. If background spatial correlation is established, a predictive stage uses a model-based approach to make predictions (Mika, 2004). Since diseases occur at particular locations but have no boundary, spatial statistical analysis study the interaction of human activities in one place with those in another. This reveals patterns to indicate that data are not independent of one another, contrary to the assumption of independency for most traditional statistical tests. The spatial statistics toolbox of ArcGIS includes both statistical functionalities and general purpose utilities for the spatial analyses. The four groups of the statistical functions toolsets and the common spatial statistical approaches employed in spatial epidemiological analyses are Measuring Geographic Distributions, Analysing Patterns, Mapping Clusters and Modeling Spatial Relationships (Scott and Janikas, 2010; Murack, 2013). A concise descriptions of the spatial analytical approaches are given in Table 1.1 below.While the other

statistical tools answer the question where something is happening, spatial regression analyses answer the question why something is happening at that place. Other geostatistical methods used are Mantel's test, spatial correlogram, Knox's test, Kulldoff's spatial scan statistic, Barton's test, Ederer-Meyer-Mantel test, Jacquez's k nearest neighbour, Ripley's K function and many more (Osei, 2010; Waller, n.d).

Table 1.1: Common spatial statistical approaches in spatial epidemiology

Spatial statistical approach	Description and function(s)
Disease mapping	A visual concise description of the geographical distribution of a disease. It exploits spatial dependence in order to smooth rates for better predictions. Gives clues to aetiology by examination of disease frequency map with exposure maps and provides estimates of risk by area
Spatial Smoothing	Computes smoothed rates to replace raw rates for each region by averaging the raw rate of the neighbouring regions. Removes heterogeneity in the rates due to either small rates or if the rates are based on very different population sizes The three most useful smoothing methods are: • locally-weighted mean, • locally-weighted median • empirical Bayesian smoothening (EBS).

Spatial pattern	A global statistics that measures spatial autocorrelation
	Identifies and measures the spatial pattern of the entire study area
	Exploratory step that indicates if feature values are clustered, dispersed, or random
	But does not indicate where specific patterns occur
	The common spatial statistics are the Global Moran's I index, Geary's C and Getis-Ord G spatial statistics
Cluster-outlier detection	A local statistics that measures specific areas of clustering
	Residual spatial clustering provides clues to disease aetiology
	Uses individual features and relationships to nearby features to identify variation across the study area Describes cluster-outlier as:

- Outlier: low value is surrounded primarily by high values (Low-High) or a high value is surrounded by low values (High-Low).

- Hotspots cluster: high values (High-High) cluster together

- Cold spot clusters: low values cluster together (LowLow).

- Outliers and Clusters are measured with Local Indicators of Spatial Autocorrelation (LISA) or an Anselin Local Moran's I, Hot/coldspots are measured by Getis-Ord G^*

    spatial statistics

| Modeling spatial relationships | A spatial regression modeling |
| | Models the association between outcome and exposures Approaches are: |

- logistic regression for point data

- linear and loglinear models for count data (common models are Geographically weighted regression and Ordinary Least Squares Regression as linear regression models

(Sources: Meza, 2003; Odoi *et al,*2003*;* Krivoruchko *et al*, 2003;Yrigoyen, 2007;Scott and Janikas, 2010, Murack, 2013; Verma and Gupta, 2014; Ord and Getis, 1995; Anselin,1995; Pui-Jen Tsai, 2012; Arie *et al*, 2010)

1.2 STATEMENT OF PROBLEM

Presently, in both medical and geo-sciences, exploration of the spatial distribution of diseases and risk factors becomes the relevant focus of health researches (Marek *et al*, 2015). This method helps to better understand the detailed mechanisms that drive heterogeneities in the distribution of diseases and health exposures. There is therefore a growing interest in the use of GIS technology or/and spatial epidemiology to manage and analyse health data for prolific policy making and programmes. In the sub-saharan Africa and for that matter Ghana, these researches are limited or neglected because most researches are rather based exclusively on the biological aspects and the characteristics of the individuals contracting the diseases (Appiah *et al*, 2011). Other researches such as Arku *et al* (2015), Basommi (2011), Osei (2010) and Marek *et al* (2015) also acknowledged this and indicated the usefulness of spatio-temporal epidemiology in control and prevention of infectious diseases. Even-though non-spatial studies are equally important, they omit or rarely unveil the spatial variability in rates and risk factors, and are unable to identify and draw attention to high risks or hot spot areas (Osei, 2008; Basommi, 2011).

13

Essentially, since risk factors and other health events are never distributed randomly in the population, diseases are likely to occur in some members of the population more than others (Dhama *et al*, 2013). However, existing studies tend not to incorporate distributional (equity) effects in assessing the causal impact of the distribution of health outcomes and social health determinants (WHO, 2013).With the aid of spatial

epidemiology, characterising the spatial variations in diseases and identifying the factors that put one population at greater risk than others is achievable. This is an important aspect to be considered critically for public health policy decisions and in planning effective and efficient health intervention for optimal utilisation of limited resources

(Almeida and Werneck, 2014). From the Health Metrics Network report of Ghana (2005), Ministry of Health (MOH) is now drawing attention on transforming health data generated by health agencies into clearly and understandable information for practical actions especially at the local levels. This is the focus of the Centre for Health

Information and Management (CHIM). To facilitate this, spatial epidemiology is the best approach because geovisual analytics are useful data mining and epidemiological tools for translating data into information and knowledge. This helps to identify areas of true underlying geographical distribution of disease incidences (Toprak and Erdoğan, 2008). Thus through spatial epidemiology, identification of high risk areas which is very difficult for health practitioners is thereby made easier (Um *et al*., 2014) and communicated clearly to many. A spatial epidemiology is therefore indispensable for wide scope health needs assessment for a matter-of-fact action

1.3 RESEARCH AIM AND OBJECTIVES

This study aims to employ a GIS-based geovisual analysis with ecological study approach to investigate the top five infectious disease in Ghana. The following specific objectives are thereby formulated:

1. To evaluate the spatial patterns in the incidence of the top five infectious diseases in Ghana

2. To identify areas with high risks of infection (hot-cold spots detection)

3. To assess the significance of determinants and their spatial variability to the incidence of infection

4. To apply spatial regression model to predict & monitor the incidence of infection at a particular place

1.4 RESEARCH QUESTIONS

The following inferential epidemiologic questions were proposed.

1. What is the spatio-temporal patterns in the spread of infectious diseases in Ghana?

2. What specific areas are with unusually high risk of infection in Ghana?
3. What significant determinants are contributing to the prognosis of infection and do they appear to vary in space?

4. Can we apply spatial regression model to predict the incidence of infection?

1.5 SIGNIFICANCE OF THE STUDY

Mapping the public health resources, diseases and other health events in relation to their surrounding environment and the existing health and social infrastructure

provide a powerful tool for monitoring and management of diseases and public health programmes (Johnson and Johnson, 2001). Spatio-temporal analysis links health data with the background exposures in space and time (Bouwmeester, 2010).This provides enough relative data and information on the pace and extent at which cases and determinants occur in various parts of the country for strategic planning and healthcare management (Appiah *et al*, 2011).

Usually, chloropleth maps show clear relationship among neighbouring areas with respect to the spatial distribution of the study phenomenon. Visual analytics is therefore fundamentally oriented on the clearer dissemination of statistical results to decision makers and the general public while providing the clear communication of the interpretations made by analysts (Marek *et al*, 2015). This is largely due to the fact that maps 'speak for themselves', and so audience needs no special knowledge or training to understand visuals or maps.

Geovisual analysis of health data is therefore the quickest way of visualising the entire health scenario and could communicate more than one type of health information at a time in a more practical and understandable format to all. With this, when changes in disease patterns, health events and risk factors are detected, the information is transmitted not only to health practitioners but to community members as well to enhance community participatory health intervention and planning of community health (Boulos, 2004; Saker *et al*, 2004; Jenkins, 2003).

For healthcare decision makers and agencies, this study could assist them to rapidly and practically identify variability in dependency patterns and then put the existing patterns into objective and measurable considerations for effective implementation

of precautionary measures (Munyaneza *et al*, 2014;Camara *et al*, n.d). This is to state that if adequate and quality data are available, a GIS-based spatial epidemiology will give detail meaning to the health data to better define the public burden of diseases and health promotion.This will then highlights that besides treating the individuals, strategies to prevent diseases need to also focus on modifying the environments that facilitate the development and maintenance of the prognostic factors (Congdon, 2015). Additionally, this approach affords each district a unique opportunity to explore their success in attempts to improve health outcomes over time and to compare their health conditions and determinants with other districts. According to the 2010 Global Burden of Disease report, this type of assessment could help districts put their health achievements in context and identify areas for improvement. The statistical and geospatial tools accurately characterise the size and spatial distribution of infection, including stratification of areas by infection risk as an imperative approach for strategic healthcare management system and policy making (Colston and Saboyá, 2013).

1.6 THEORETICAL FRAMEWORK AND SCOPE OF STUDY

Fundamental to epidemiological studies are the frequency measures of morbidity (prevalence and incidence) and comparison of the rates over time between two distinct exposure groups using relative risk/risk ratio to gain insight into a disease prognosis. While prevalence is useful in the study of the burden of chronic diseases and implications for health services, the main measure of acute diseases is the incidence and is more useful for studies of causation (Jeffery *et al*, 2014).This study used the incidence-based approach in assessing the burden of the infectious diseases (IDs) which is same as prevalence-based approaches in stable situations. One major assumption of the incidencebased approach is that the incidence, mortality and disease progression remain unchanged over time (Kwong *et al*, 2010).There is also an underlying supposition that a disease occurs at a single point of time whereby before this time the disease is absent but present subsequently (Jewell, 2009).

Most spatial epidemiological studies generally adopt ecological approaches despite the comparatively high precision, accuracy and reliability of individual-level studies. This is due to high expenses, time consuming, laborious nature, ethical and other practical issues confronting the individual-level studies (Waller and Gotway, 2004). Additionally, due to computationally demanding nature for large populations, individual-multilevel data are suggested to be approximated by ecological analysis with areas (Congdon, 2015). Ecological study uses aggregated data, which are usually routine-based data gathered from the entire population to compare the health outcomes between different groups during the same period of time or in the same population at different time. From ecological inferences, individual disease risk and level of exposure to risk factor from the aggregated data are deduced. At times, this may lead to ecological bias, resulting in ecological fallacy whereby conclusion based on grouped data analysis is different from those based on individual data analysis. But spatial epidemiological study using data from varied geographical points to describe the exact spatial locations of cases or events and exposures is seen as the closest link to the individual-level study (biologic model). This is because average disease risk and exposure levels of individuals reflect in the distribution of diseases or health-related states and determinants in a specified population (Waller and Gotway; 2004, Osei and Dukar, 2008; Osei, 2010).

This study therefore adopted the ecological study approach and used aggregated (routinebased) data collected from the entire population to assess the spatial distribution of infectious diseases in Ghana and the correlated health determinants, including spatial modeling. Attention to link determinants and IDs is focused on

18

selected acute, vectorborne and epidemic infections. However, most socio-demographic, cultural, economic characteristics and other social determinants of health are not disease specific. It is therefore imperative to consider the wide range of infection that are potentially influenced by these determinants (Saker *et al*, 2004). This study therefore combined the five topmost infectious diseases over five-year periods at district levels for a GIS-based spatial epidemiology of infectious diseases in Ghana.

A wide range of methods in spatial epidemiology are available but Fig 1.3 gives the framework of the logical and sequential processes involved in conducting a spatial epidemiological studies (Anamzui-Ya, 2012).
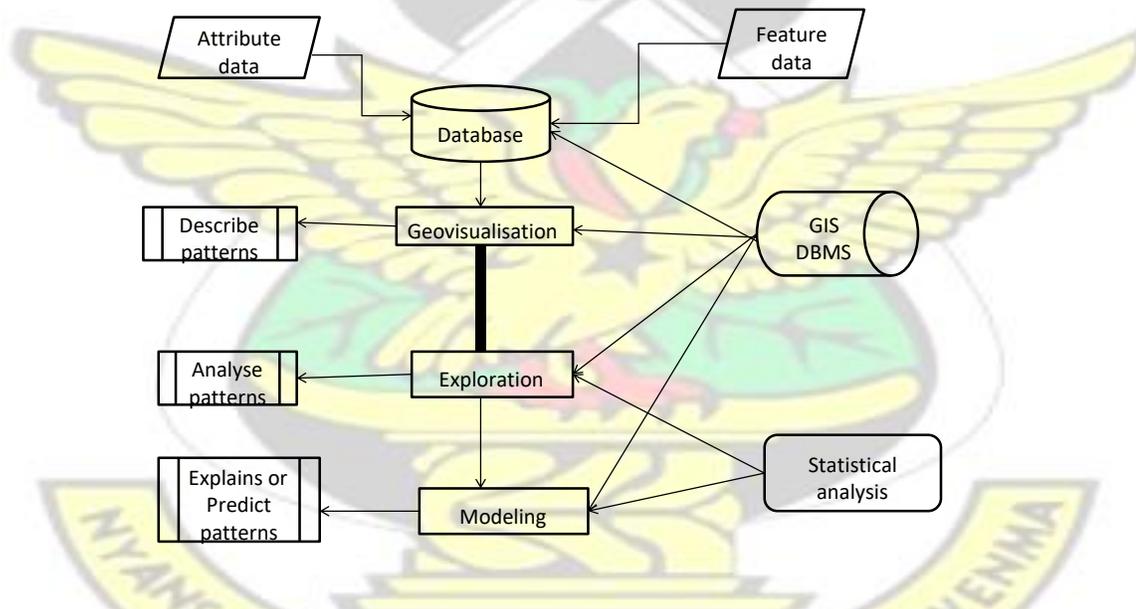
Fig 1.3 The conceptual framework of spatial epidemiological studies (Source: Pfeiffer *et al*, 2008, adapted from Anamzui-Ya,2012)

1.7 LIMITATIONS OF THE STUDY

Although the data is gathered from qualified and statutory bodies, there might be possible data quality and adequacy problems. There are possibilities of diagnosing OPD cases based on clinical symptoms rather than laboratory confirmation which may result in under or over-diagnosis. This is likely to affect the spatial variations in the incidence infection. Because of the potential influence of hydrological, geological and climatic factors on infection, omission of such factors due to unavailability of data could influence the results of the predictive capacities of the spatial regression models. Disease control intervention policy programs which might be in place currently in each district was not included in the modeling. There might be a likelihood of spatial error even-though the spatial statistical techniques tried to account for the different sizes and shapes of the districts. The spatial regression models, however, account for this spatial error.

The study is ecological (aggregated data)study design, hence ecological inference on much smaller groups or individual disease risk and level of exposure should be avoided or made conscientiously. This is to avoid ecological fallacy whereby conclusion based on grouped data analysis is different from those based on individual data analysis.

1.8 ORGANIZATION OF THE STUDY

The study is structured into five (5) chapters. The first chapter provides the introduction of the study and consists of the background of the study, statement of problems, objectives, research questions, significance of the study, theoretical framework and scope of study, research limitations and then organization of the

study. Chapter two involves review of related literatures on the spatial epidemiology of infectious diseases. Chapter three presents the materials and methods, which is consisted of the materials and datasets used, background of the study area, methods of data collection and the appropriate approaches used for the spatial data analysis. Chapter four comprises presentation and discussion of results. Chapter five summarises findings, draw conclusions on the key findings and give appropriate recommendations. The reference list then follows and work ends with the appendices.
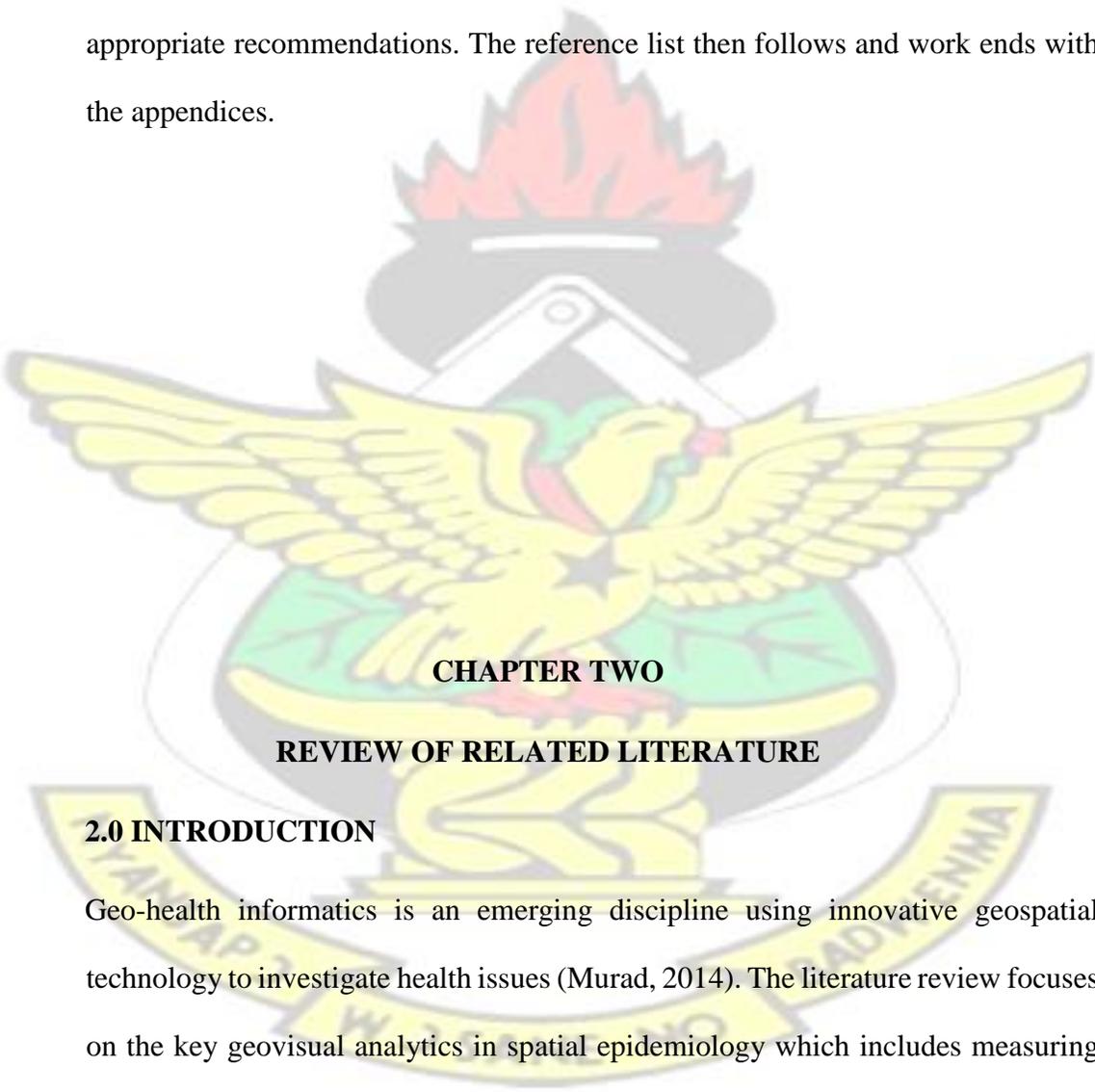
CHAPTER TWO

REVIEW OF RELATED LITERATURE

2.0 INTRODUCTION

Geo-health informatics is an emerging discipline using innovative geospatial technology to investigate health issues (Murad, 2014). The literature review focuses on the key geovisual analytics in spatial epidemiology which includes measuring geospatial patterns, detection of spatial clusters and analysing geospatial relationships of infection.

2.1 GEOSPATIAL PATTERNS OF INFECTIOUS DISEASES

Spatial statistical studies normally distinguish between two different kinds of spatial effects, known as spatial autocorrelation (spatial association or interaction) and the spatial structure (spatial heterogeneity). Studying the interaction of activities in one place with those in another place is very important in spatio-temporal analyses of infectious diseases due to the dynamic nature of emergent infections. The identification of spatial pattern in data is useful in gaining a better understanding of a geographic phenomenon, monitoring conditions on the ground, comparing patterns and tracking changes (Khalid, and Al-Zahrani, 2013; Mitchell, 2015). In GIS-based spatial analysis, mapping of diseases and spatial statistics are used to accurately identify, quantify and describe the distribution of infectious diseases across a study area (Eisen and Eisen, 2011). Fundamental to statistical analysis of dependence and heterogeneity in spatial pattern is finding the degree of spatial association among data representing related locations (Getis, 2004). A global spatial autocorrelation analysis is used to evaluate whether the pattern of infections among related locations is clustered, dispersed, or random. This spatial patterns of infection are measured by Moran's index, Geary's c and the General G statistics (Anselin,1995).Many of these spatial dependences are commonly based on Moran's I index of spatial association (Krivoruchko, 2003). The complexities in climatic, environmental, physical, and social variables results in great differences in space-time infestations across regions. This makes analysis of spatial pattern of infectious diseases very important as an epidemiological tool to monitor diseases, risk factors and distribution of health facilities and other health resources. Accounting for these variations for disease risk assessment and monitoring are

needed to facilitate healthcare policy intervention (Onwuemele, 2014, Norstrøm ,2001; Eisen and Eisen, 2011). In the spatiotemporal investigation of the incidence of Japanese Encephalitis (JE) epidemics in the Gorakhpur, Verma and Gupta (2014) used Moran's I index to provide a better understanding of the influence of geographical changes on the pattern of occurrence of the JE disease. This is based on the high accuracy of combining spatial statistics with disease mapping in identifying patterns in the spread of this infectious disease. In a related study, spatial pattern analysis of epidemiological locations of interest of human

African trypanosomiasis (HAT) have enabled HAT distribution and risk in Democratic Republic of Congo to be known with more accuracy than ever before (Lumbala *et al,*2015*).* GIS-based spatial analysis has been demonstrating its power in redefining heterogeneity in the spread of infectious diseases of diverse nature. Distribution of sexually transmitted infections such as chlamydia, gonorrhea, and HIV were mapped successfully and each found to clearly define spatially heterogeneous areas of infection (Law *et al*, 2004). By mapping the locations of farms and other facilities with animals, veterinary epidemiologist were also able to investigate outbreak of zoonotic diseases for strategic intervention of zoonosis (Sanson *et al,* 1994; Norstrøm*,* 2001). Risk maps are very useful to inform risk-based disease surveillance and control systems. This is because in preparing a control strategy, the exact disease status and case definition are critical. In a spatio-temporal pattern analysis of reported dengue cases, Jeefoo *et al* (2010) found out that the space and time of the cases showed outbreak movement and spread patterns that could relate to entomologic and epidemiologic factors of a particular region. Odoi *et al* (2003) identified spatial and temporal patterns in giardiasis distribution

but the global tests of spatial autocorrelation was non-significant at all spatial scales. It is therefore plausible that disease excess limited to one or two regions may not have a significant impact on the overall measures. Assessing the spatial pattern of health facilities was also found in many researches such as Simarro *et al* (2014), Usman *et al* (2013), Masters *et al* (2013), Verutes *et al* (2012) and many others, to be playing very active roles in addressing spatial inequality in the distribution of health facilities to improve upon healthcare and the spread of infectious diseases. Schwanke *et al* (2013) also explored GIS to demonstrate how resources distribution mapping at the local level can inform health policy decisions on where to stock resources in advance and for reallocation towards pandemic preparedness.

2.2 DETECTION OF SPATIAL CLUSTERS OF INFECTION

Since measuring spatial patterns of interaction cannot indicate where specific infection is occurring, measuring clustering identifies and map areas of unusually high risk (hot spots) and low risks (cold spots) as well as outliers. Comparing clusters to the locations of other features gives a better understanding of why clusters occur. This enables identifying possible contributing factors in order to decide what action to take (Mitchell,

2015). Thus identification of hotspots draw attention of policy makers and healthcare providers for strategic healthcare management and interventions. Moran Local Indicators of Spatial Associations or Anselin Local Moran's I and Getis-Ord Gi* spatial statistic are used to assess the spatial clustering for the hot/cold spots detection. Spatial clustering of landscape, socio-demographic and economic, and climatic associations with disease morbidity are inevitable in an attempt to understand the outbreak dynamism of diseases (DeGroote *et al,* 2014*).* Cluster detection is useful in investigation of disease outbreaks and the possible causes of the outbreaks. This is because by the identification of the background characteristics of the populations at risk, geovisual analysis of disease certainty enhance more meaningful inferences from the spatial health data. It is then easily

then used by health agencies and researches to track disease sources and the movement of the infectious agents for effective and efficient response to outbreaks (Um *et al*, 2014; Johnson and Johnson, 2001). The Getis-Ord Gi*(d) cluster analysis technique used in Japanese Encephalitis (JE) outbreak investigation specifically identified that blocks on the built up areas reported significantly higher incidences and used for the control of the disease (Verma and Gupta, 2014). Toprak and Erdoğan, (2008) used GIS

based spatial analysis to identify statistically significant spatial clusters of typhoid fever among the provinces of Turkey. This investigation also explained the individuals returning from a recent travel to an endemic regions as the cause of the epidemic. Spatial analysis of the rate of malaria infection in Nigeria using principal component analysis clearly indicates high concentration of malaria infections in some few states with seasonal variations (Onwuemele, 2013). Even in a situation of decline in infectious diseases like malaria transmission, it becomes increasingly important to monitor changes in the malaria incidence rather than prevalence. Alegana *et al* (2013), and Osei and Duker (2008) demonstrated using spatial clustering analysis to respectively identify specific areas with high malaria and cholera incidence respectively to guide control and monitoring of infectious diseases. Cluster-outlier detection is helpful in zonoosis investigation as well. Schimmer *et al*., (2010) used GIS to investigate Q-fever outbreak and confirmed that a single dairy goat farm was the source of the human Q-fever outbreak at Netherlands in May 2008. Several other studies and review works, including Norstrøm (2001), Eisen and Eisen, (2011), Dhama *et al* (2013), Ekpo *et al*, (2013*)* and elsewhere had demonstrated harnessing GIS/GPS and DSS technologies to tract infection plan prevention, control and eradication of diseases and for general health system strengthening in most developed country. The visual and statistical evidence

of spatial clustering (Odoi *et al*, 2003) of infection put GIS-spatial epidemiology as useful tool in health researches and policy making.

2.3 DETERMINANTS AND MODELING SPATIAL RELATIONSHIPS

GIS analysis and other spatial tools are being used to establish spatial regression relationships between the risk factors and incidence of infection. This gives more insight into what is happening in a location, predict where something might occur, or begin to examine why things occur where they do (Mitchell, 2015). Regression models are used to explore spatial relationships to better understand the factors behind the observed spatial patterns and then predict spatial outcomes (Cianci *et al*, 2015; Scott and Janikas, 2010). Environmental, socio-demographic and economic, climatic and geological factors, sanitation and other social determinants of health are correlated with rates and then used in the spatial modeling. Density maps of Malaria, Filarial, Kala-azar and Dengue in

Varanasi by Rai *et al*, (n. d) exposed that factors such as poor sanitation and hygienic conditions including improper disposal and management of wastes in the city area accounted for higher cases in the city compared to rural areas. Geospatial studies also has the potential to substantiate or/and unveil aetiology of some diseases of confusing pathogenesis. This is exemplified in a study conducted by Molla *et al* (2014), disclosing high quantities of smectite, mica and quartz minerals within the soil in the pathogenesis of podoconiosis and acute adenolymphangitis. Assessing the risk factors of the population at risk in relation to disease morbidity and mortality is relevant in health interventions. Spatial variation in attributable risks and disease frequency reflects the impact of the determinants on health which is important for health planners to practically prioritise health interventions with the limited resources (Congdon, 2015).The strategies may involve not only treating or changing the behaviour of the individuals but also modifying the prevailing determinants such as changing water source. A GIS supported investigation of a cholera outbreak in Abeokuta of Nigeria identified the municipal waterworks as the source of the outbreak (Shittu *et al*, 2010), possibly from lower quality of drinking water supply in these areas (Odoi *et al*, 2003). Socio-economic indicators, land surface temperature and soil type proved to be best predictors in the environmental visceral leishmaniasis risk model (Tsegaw *et al*, 2013; Almeida and Werneck,

2014). Upon profiling the determinants, particularly identifying the significant determinants, spatial risk models are capable of estimating the occurrence or frequency of measurable features within a particular geographical area and displaying the outputs in thematic maps. With several spatial modeling approaches, disease mapping and modeling have been helpful in suggesting the cluster and spread pattern of a disease (Gesink *et al*, 2006).

The application of remote sensing or GIS-based data in habitat suitability models are being applied intensively to understand and predict factors affecting especially vectorborne diseases (Hönig *et al*, 2011; DeGroote *et al*., 2012). Ehlkes *et al* (2014) applied geographically weighted Poisson regression model to detect the associations between land use/land cover and mosquito-borne diseases. In a similar works, McCann *et al* (2014) and Cianci *et al* (2015) explored landscape and meteorological variables to model malaria vector habitat locations and found random forest model to be most accurate than the non-linear discriminant analysis and logistic regression models in malaria prediction. The logistic regression modeling and mapping was, however applied by MacMillan *et al* (2012) for successful prediction of the human plague case endemicity in West Nile region of Northwestern Uganda. Hay *et al* (2009) employed model-based geo-statistics Bayesian predictive framework on contemporary database of PfPR surveys to construct a continuous, global *Plasmodium falciparum* malaria endemicity map for predicting values at non-sampled locations. Bayesian hierarchical conditional auto-regressive modeling is one of the commonly used modeling method used and cited in Osei (2010), Alegana *et al*, (2013), Ekpo *et al* (2013) and many others for modeling various forms of infectious diseases and to identify the determinants. Other spatial models employed in spatial epidemiology include Boosted Regression Trees (Cheong *et al*, 2014), Maximum-entropy modeling (Atkinson *et al,* 2013; Hönig *et al*, 2011), ecological niche modeling (Monroe *et al*, 2014), Ordinary Least Squares regression, Spatial Lag model and Spatial Error model (Bradshaw, 2013). Useful but extremely underutilised in Ghana, this study therefore aimed to use GIS-based spatial epidemiology to describe the spatial patterns, identify clusters and then explain or predict infection from the top five infectious diseases in Ghana using spatial regression models.

CHAPTER THREE

MATERIALS AND METHODOLOGY

3.0 INTRODUCTION

This chapter outlined the materials and methods used in the study. It began with the

description of the study area, materials used, datasets and data collection

techniques, spatial and ecological analysis and then the processes of spatial regression modeling.

3.1 DESCRIPTION OF STUDY AREA

The Republic of Ghana is located on latitude 8° 00' N and longitude 2° 00' W and has a total land area of 238,589 km^2 (National Population Council, 2011). The country consists of ten administrative regions which are subdivided into presently 216 districts (instead of 170 districts for the 2010 census) to ensure equitable resource allocation and effective administration at the local levels. The population of Ghana as the end of 2014 based on 2.4 percent growth rate of the 2010 census was 27,043,093 with 51 percent of the population living in urban areas (GSS, 2014; GDHS, 2014). There are number of government and private health and education facilities. Each of the ten (10) regions has regional hospitals and there are four medical teaching hospitals. Besides the numerous private health facilities, including the mission hospitals, every district in Ghana has a government or government-mission owned hospital. Data collection on health status and health service performance are largely derived from three types of sources: populationbased data collection including household surveys, disease surveillance, and routine and service-generated statistics. The population-based surveys are carried out by the Ghana Demographic and Health Survey (GDHS) and GSS. All health related activities in the country are coordinated by the Ministry of Health (MOH).

3.2 MATERIALS USED

Boundary maps of the districts obtained from the Survey and Mapping/Planning Division, Accra.

Application software: GIS software and ArcGIS developed by Environmental System Research Institute (ESRI), GeoDa software, spatial software developed by Luc Anselin through the Center for Spatially Integrated Social Science at the University of Illinois,

Microsoft Excel, and Statistical Package for Social Sciences (SPSS).

3.3. STUDY DESIGN

Ecological study design was used in this study since the data used for the analysis were aggregated at the local administrative district levels.

3.4. DATASETS AND SPATIAL DATABASE CREATION

A geo-relational database for the coordinates, population at risk, disease rates and determinants were created to form the attribute table for the spatial analysis.

3.4.1 Spatial data preparation and cartographic display

The spatial database provides information on the spatial location of each metropolis, municipal and district assemblies (MMDAs) in Ghana. The spatial scale of analysis in this study was at the 170 MMDAs levels using the 170 MMDAs administrative boundary maps obtained from the Survey and Mapping Divisions (SMD), Accra and shapefile from Geomatic department of KNUST. The170 districts were used even-though there are currently 216 MMDAs because according to the SMD, the new MMDAs created after the 2010 Population and Housing Census (PHC) was conducted did not have their boundary maps finalised yet due to boundary disputes and other technical issues. Spatial data were obtained by extracting the coordinates of the digitised boundary map and the shapefile into a National Grid, Ghana Metre-

Grid Universal Transverse Mercator (UTM) coordinates system using ArcGIS version 10.1.The cartographic display of the 170 administrative districts and 10 regional polygon layers were then generated.

3.4.2 Population database

Population at risk, the number of people under study is an important factor in estimating the measures of disease and health events in order to give quite reliable impression of the health status of the population (Jewel, 2009). The entire population is at risk of infectious diseases. The population database for each district was obtained from the 2010 PHC conducted by the Ghana Statistical Service (GSS) and the intercensal projections over the periods of 2011 to 2014. The population data were geo-referenced and added to the

attribute table.

3.4.3 Case morbidity database of Infectious Diseases

In Ghana, routine-service-based data are generated through the activities of the Ghana Health Service (GHS) and its affiliates. It is mandatory for all health facilities within a district to present their annual case reports to the disease control units (DCU). The data is then deposited at the centralised reporting system, the Centre for Health Information and

Management (CHIM) of Ghana Health Service (GHS) using the Ghana Health Service Dhmis. In this study, data on annual morbidity reports from health facilities at outpatient departments (OPD) within the districts for the periods 2010 to 2014 were obtained from the CHIM of GHS. The top five morbidity of infectious

diseases were extracted. Morbidity case is defined as a visit at an OPD and clinically diagnosed of an infectious disease (Appiah *et al*, 2011). The extracted morbidity of the infectious diseases were used as the case morbidity attributes of each district.

3.4.4 Determinants database

A determinant is a risk factor, whether event, characteristic, or other definable entity, that influence the occurrence of disease and other health-related events. The risk factor may not necessarily be the causal factor but is associated with an increased probability of an adverse outcome. Because illness does not occur randomly in a population, but occur if and only if the right accumulation of determinants exists, epidemiology is used to search for these risk factors (US CDC, 2012). Many diseases are caused by more than one risk factor (WHO,2009) and since risk factors are usually modifiable, intervening to alter them in a favourable direction can reduce the probability of occurrence of diseases remarkably.

The determinants in this study were processed from the 2010 PHC. The detailed 2010 PHC provided data relating to various aspects of the populations and households, disaggregated to district levels in the regional analytic reports. These data were consulted for information relating to the determinants for each district. This methodology was reported in a number of similar works, including Colston and Saboyá (2013), Bradshaw (2013), Osei (2010), Osei *et al* (2012), Verma and Gupta (2014), and so on. The determinants were extracted and classified into the following risk factors that are known to be risk factors of infection: social and built environments, demographic and economic characteristics, agricultural activities, sanitation and ICT (Colston and Saboyá, 2013, Bradshaw, 2013; US CDC, 2012;

WHO, 2013). The determinant at a location was adjusted towards the overall nationwide average by computing a weighted proportion for each district, giving the relative measure of the determinants. All the determinants were processed in Excel MS office and then transformed into a GIS readable format using Arc GIS version 10.1 software. The determinants are geocoded and implemented as attribute data, mapped and classified using natural breaks (Jenks' optimisation) classification techniques in the ArcGIS. The Jenks' Optimisation has most advantage among the classification methods because it ensures the internal homogeneity within classes while maintaining the heterogeneity among the classes (Osei, 2010).

3.4.4.1 Demographic characteristics

Indicators included were age, sex ratio, education, literacy level and religion.

3.4.4.1.1 Age structure

Age was structured by broad age groups into 0-14 years representing the young dependent population, 15-64 years representing the non-dependent population, and 65 and older representing the old age dependent population per the 2010 PHC. The age structure for a district was computed as the percentage share of the district of that of the nation at large for that particular age structure.

3.4.4.1.2 Sex ratio: male-female ratio

The sex ratio measures the proportion of males to females in the population expressed as percentage. Sex ratio therefore indicates the number of males per 100 females in the population of each district. This indicator is important because of the behavioural differences by gender which can influence disease transmission.

3.4.4.1.3 Educational level

Several researches had established that besides reflecting and sustaining improvements in socio-economic potential of a country, educational level is one of the key indicators for healthy practices. Data on the population of persons 6 years and older who attended or currently in school with reference to the 2010 census period was used. The educational status were classified into five categories viz never attended, basic school (kindergarten, primary, JSS/JHS/middle), secondary/senior high, vocational/technical/commercial and tertiary (higher training institutions, polytechnics, universities and university colleges). To obtain a proportional educational level of each district, a percentage share of each type of education of the respective population of the total was computed for each district.

3.4.4.1.4 Illiteracy level

Literacy is defined as the ability to read and write in any language with understanding, indicating means of sharing societal knowledge. Thus, literacy is important for the development of the individual and society, and the ability to read and understand basic health information is very essential for healthy practices. The illiteracy level of a district was estimated as the proportion of the population aged 15 years and older in the district of the total population of the same age group who cannot read and write any one of the three languages: one Ghanaian languages, English language and French. A large value indicates high illiteracy level while small value indicates high literacy rate.

3.4.4.1.5 Religious affiliation

The district share of the population in each of the four categories of religion (Christian,

Islam, Traditional religions and no and other) were estimated.

3.4.4.2 Social environment

The indicators considered here include urbanisation level, migration and overcrowding living conditions.

3.4.4.2.1 Urbanisation level

Urbanisation is a process by which an increasing proportion of people become permanently concentrated in relatively small areas to form cities. According to GSS criteria, urbanisation is defined as the population of at least 5,000 people or more living in a given locality. Thus a locality within a district with a population of $\geq$ 5,000 people is classified as urban, and less than 5,000 as rural. The percentage share of the total population that are urban populations in each district was computed to obtain a proportional urbanisation level of each district.

3.4.4.2.2 Migration

Migration is a measure of population mobility which involves a geographical movement from a usual place of birth or residence over a defined territory beyond a defined period (Saker *et al*, 2004). Migration either be inter or intra-migration. The former is the movement of people between geographical boundaries within national boundaries while the latter may involve the movement of people within localities in the region or district (GSS, 2010). The information on place of birth and non-Ghanaian population were used to estimate the intra- and inter-migrations

respectively and each type for a district was expressed as the percentage of the district's share of the total.

3.4.4.2.3 Overcrowded living conditions

The indicators used were the population density, population per house, single room occupancy and sleeping room, average household size and households per house. These six variables for the overcrowded living conditions were summed up to form a single index of risk called overcrowding index (OI) for each district. For proportional representation, the OI of each district is then expressed as a percentage share of that all districts (Colston and Saboyá, 2013; Bradshaw, 2013; Osei, 2010)

3.4.4.3 Sanitation

The WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation, JMP, (2010, 2013) standard method of classifying sanitation facilities and drinking-water sources for household facilities as 'improved (safe)" and "unimproved (unsafe) was used in this study. The indicators included were: main source of drinking water, toilet and bathing facilities, and solid and liquid wastes disposal. The JMP methodology was applied to calculate the proportion of households lacking access to improved sanitation.

The sanitation condition of a district was described as the percentage of the district's share of the nation's household who do not have access to improved household facility (Osei, 2010; WHO/UNICEF JMP, 2010). The five indicators processed below were combined to form an insanitary index (II). Larger values of II reflect

poor or bad sanitation condition while smaller values reflect good sanitation condition. The insanitary indices were mapped and classified using the natural breaks classification techniques.

3.4.4.3.1 Lack of access to safe drinking water supply

The different sources of drinking water were classified into improved and unimproved water source according to JMP classification method. An improved drinking-water source is one that by the nature of its construction adequately protects the source from outside contamination, in particular with faecal matter. Improved drinking-water sources included piped water (for the household, public taps or standpipes), tube well/boreholes, protected dug wells, protected springs, sachet water, and rainwater collection. On the other hand, households drinking from unprotected dug wells and springs, vendorprovided water such as cart with small tank or drum and tanker trucks, and surface water

(streams, rivers, ponds, dugouts, lakes, dams, canals) were classified as drinking from unimproved water source. The indicator for drinking water for each district was calculated as the percentage of the district's share of the nation's total households that lack improved drinking-water sources. Larger values reflected poor or unimproved source of drinking water, while smaller values reflected relatively improved drinking water (WHO 2009; Osei, 2010; Colston and Saboyá, 2013, WHO/UNICEF JMP, 2010).

3.4.4.3.2 Lack of access to improved toilet facility

The different types of toilet facilities used by households were classified as Improved and

Unimproved according to JMP method (WHO/UNICEF, 2013 update). By JMP definition, an improved toilet facilities hygienically separates human excreta from human, animal and insect contact and are not shared with other households,

36

otherwise unimproved. Hence Improved facilities include sewers or septic tanks, pour-flush latrines or Water closet (WC), pit latrine with slab and Ventilated Improved Pit (VIP) latrines, provided the facilities are not shared or public. Unimproved toilet facilities are pit latrine without slab or open pit, hanging latrine, bucket/pan and no facility or open defecation (in bush, beach, water bodies, other open spaces or disposed together with solid wastes) and any type of public facilities. Using the JMP method, the sanitation condition of a district with regard to access to toilet facility was estimated as the percentage of the district's share of the nation's households that lack improved toilet facilities. Hence, larger values reflected poor or bad sanitation condition while smaller values reflected good sanitation condition as far as toilet facility was concerned.

3.4.4.3.3 Lack of access to improved bathing facility

Adopting the JMP classification method, own bathroom for exclusive use, share a separate bathroom in the same house and use of private open cubicle were classified as improved bathing facilities for the households. Others such as shared bathrooms in another house, public bath house, open space around house, in water bodies (lake, pond, rivers, dam and others) were unimproved facilities. The indicator for access to bathing facilities for each district was estimated as the percentage of the district's share of the nation's households that lack improved bathing facilities (that is used any other type of bathing facilities besides owned bathroom, share a separate bathroom in the same house and use of private open cubicle). Thus a larger value reflected poor or bad sanitation condition with regard to access to bathing facility while smaller values reflected good sanitation condition.

3.4.4.3. 4 Lack of improved solid waste disposal usage

Among the varied means of solid wastes disposal by households; collected public dumps (container), burned and buried by household were classified as improved wastes disposal otherwise unimproved disposal. The solid wastes disposal indicator for each district was estimated as the proportion of the district's share of the nation's households that used unimproved solid waste disposal facilities (that is open space public dump, indiscriminate dumping, dumping into water bodies and others besides the improved methods). Thus a larger value reflected poor or bad sanitary practices for solid wastes disposal but good sanitary practice for smaller values.

3.4.4.3.5 Lack of improved liquid wastes disposal usage

The liquid waste disposal by households through the sewage system and through drainage system into pits (soak away) and drainage system into gutters were categorised as improved liquid waste disposal method. However, disposal methods including, thrown into open gutters, thrown onto street/outside, thrown onto the compound and others were put into unimproved disposal methods. The insanitary level of a district with regard to liquid disposal was described as the proportion of the district's share of the nation's households that used unimproved disposal methods. Larger values reflected poor or bad sanitation condition for liquid wastes disposal while smaller values reflected good sanitation condition.

3.4.4.4 Household smoke exposure level

The household smoke exposure from cooking fuel and source of lighting were combined to reflect the household smoke exposure index.

3.4.4.4.1 Smoke exposure from cooking fuel

The cooking fuel influences the quality of indoor air which can affect the health of residents. According to WHO report (2009), the main cooking fuel for more than half of the world's population are solid (wood, dung, coal or agricultural residues) on simple stoves or open fires. Under conditions of limited ventilation, solid-fuel use leads to high exposures to indoor smoke and are largely associated with health risks and deaths from lower respiratory infection, chronic obstructive pulmonary and lung cancer, particularly for women and children. The global health risks study (WHO, 2009) approach for measuring indoor smoke exposure level was used to estimate the proportion of a district's share of the nation's household populations using solid fuels (that is using wood, charcoal, crop residues, saw dust and other agricultural residues as main cooking fuel). Larger values reflected high level of solid-fuel indoor smoke exposure while smaller values reflected low level of exposure.

3.4.4.4.2 Indoor smoke exposure level from source of lighting

The source of lighting in housing units gives important information about their quality and a contributor to indoor smoke exposure. This was estimated as the proportion of the district's share of the nation's households that do not use electricity (but use other sources such as candle, kerosene lamp, fire wood, crop residue etc) as their main source of indoor lighting. A larger value reflects high level of indoor smoke exposure while smaller values reflect low level of indoor smoke exposure.

3.4.4.5 Economic characteristics

Socio-economic development of a region or individual is a key determinant for provision and affordability of healthcare and healthy practices. In this study, two indicators; Employment-to-population ratio (EPR) and dependency ratio (DR) were used to indicate economic strength of a district. These give a fair idea about the poverty level of the district. Smaller EPR and higher DR means weak economic strength, giving impression of the high poverty level and vice versa.

3.4.4.5.1 Employment-to-population ratio (EPR)

This ratio is the proportion of a district's working age population (15 years and above) that is economically employed. The indicator was computed by dividing the total number of employed person aged 15 years and above by the population of the same age group (Osei-Assibey and Grey, 2013). The EPR was estimated for each district. A larger value indicated that a greater proportion of the population that could be working are working (possibly low poverty level) while a low ratio was an indication that a large part of the working population is not gainfully engaged directly (possibly high poverty level). Work or job in this context is a work for wage or salary either in cash or kind (GSS, 2013)

3.4.4.5.2 Dependency ratio (DR)

The dependency ratio measured the number of dependents (child and old age) for every 100 people working (GSS, 2013). As such this indicated the economic pressure on the individual and was one of the key indicators of socio-economic development.

3.4.4.6 Agricultural activities

Agriculture as a predominant contributor to the socio-economic development of the nation, contributing to about 30 percent of gross domestic product (GDP), is also a key determinant. This exposes farmers to varied sources of infection such as soil-transmitted helminthiasis, water bodies, animals (zoonosis) and many more. The main agricultural sub-sectors include crops, livestock, fisheries, agroforestry and non-traditional commodities. Agricultural household is defined as where, at least, one person in the household is engaged in any type of farming activity, namely crop farming, tree growing, livestock rearing and fish farming (GSS, 2012). By this definition, the agricultural activity level of a district was described in this study as the proportion of the district's share of the nation's households engaged in agriculture.To further explore the level of exposure through the agricultural activity, the proportion of the district's share of the nation's households engaged in each of the four main types of farming activities: crop farming, tree growing, livestock rearing and fishing were also computed.

3.4.4.7 Built environment (Housing conditions)

The 2010 PHC report also included information on the principal materials used in the construction of the housing units. Even-though there is still no commonly agreed definition of healthy housing, a number of body of researches on housing environment has acknowledged associations between housing environment and health, either directly or indirectly. Living and housing conditions have been found as the basis of many factors influencing residential health (Bonnefoy, 2007).The material for the outer wall, the floor and the roofing and the type of ownership of

dwelling were used as indicators for the built environmental conditions (Colston and Saboyá, 2013; Bradshaw, 2013). For the sake of simplicity, the type of construction material of housing units were regrouped into two according to the main material used: Traditional and Modern as reported in similar work by Abelti *et al* (2001). Adapting the JMP classification style, modern housing materials were considered as "improved" and local or traditional as "unimproved" to serve as an indicator of the quality of the housing unit. The housing unit indicators processed below were combined to form traditional housing index which was mapped and classified using Jenk's classification techniques.

3.4.4.7.1 Lack of improved Outer wall material of the house

The modernized or improved outer walls were made up of metal sheet/slate, stone, burnt bricks and cement block/concrete. Households living in houses of outer walls made of mud brick/earth, wood, bamboo, thatch/palm leaf, sandcrete/landcrete were classified to be unimproved. A district's traditional housing condition was estimated as the proportion of the district's share of the nation's households that were living in housing units made up of unimproved outer wall. Larger values reflected poor (more traditional) housing condition while smaller values reflected good (more modernised) housing condition.

3.4.4.7.2 Lack of improved Floor material usage

From the information on floor material; terrazzo, burnt brick, cement/concrete, vinyl/tiles and ceramic/marble tiles were classified as modernized or "improved" floor materials, otherwise traditional or "unimproved". Hence the district's floor housing condition was estimated as the proportion of households in the district

using the traditional floor materials; that is used mud brick/earth, stone and wood for the floor. For this, a larger value reflected poor (more local) floor housing condition while smaller values reflected good (more modernised) floor housing condition.

3.4.4.7.3 Lack of improved Roofing material

The modernized roofing materials included metal sheet, slate/asbestos, cement/concrete and roofing tiles, otherwise traditional roofing. The housing condition with regard to type of roofing material was measured by estimating the proportion of households in the district of the nation's households that roofed with wood, thatch/ palm leaf, mud/earth, bamboo and other local materials. Larger values reflect poor (more local) roofing housing condition while smaller values reflect good (more modernised) roofing housing condition

3.4.4.8 Built environment (Ownership of dwelling)

Housing ownership was also included as one of the indicators because the type of ownership or tenancy arrangement gives the impression of the care and maintenance of the house and the environment including the housing facilities (Bradshaw, 2013). The non-ownership status of the housing unit was expressed as the percentage of non-owner occupied households for each district as a proportion of the nation's households .

3.4.4.9 Information and Communications Technology (ICT) usage

The use of ICT can greatly influence disease and threat of infection in several ways. This influence is through improved medical technologies for disease diagnosis and

treatment, quick transfer of health information, continuing medical education, telemedicine, accessible information on current best evidence-based clinical practices, public health education and so on (Saker *et al*, 2004). Thus as a determinant, ICT can reduce the threat of infection and morbidity. In the study, ICT level of a district was measured by the percentage share of the district of the nationwide distribution of population 12 years and older using mobile phones and internet facilities, and proportion of households owning desktop/laptop. The three proportions were combined to reflect the proportional level of

ICT usage (Technology index) for each district.

3.5 SPATIAL EPIDEMIOLOGICAL ANALYSES

This section described the approach taken to estimate the incidence of infectious diseases which was then smoothen using Empirical Bayesian Smoothing (EBS) technique. The smoothed rates were used for disease mapping, evaluating the spatial distribution of infection, correlation analysis and spatial modeling. The space-time data analysis tools of the ArcGIS was useful for the spatial epidemiological study (Verma and Gupta, 2014).

3.5.1 Incidence rate estimation

The probability of a particular event occurring (disease occurrence) or risk is estimated by the ratio of the number of people who actually experienced the event during a given period to the population at risk to whom the event might have occurred (Anselin *et al*, 2006). For rate, person-time is used as the population at risk where the base population is multiplied by the number of years to yield person-years compatible with the event counts. Hence the disease incidence rate was

44

computed from the number of reported cases of the top five infectious diseases at each district per unit population at risk of 10,000 people for each district. That is if x_{it} denotes the reported infectious disease morbidity counts at district i $(i= 1, 2, 3,.....,$n where n =170; number of districts observed$)$for the year t $(t = 2010, ...,2014)$, and

P_t denotes the mid-year population for t
Then the incidence rate per year, R for the i-th district in the year t is given by

$$R_{it} = \frac{X_{it}}{P_{it}} \times 10,000$$

The cumulative incidence rates of infection was calculated for each district by summing all the R_{it} over the periods of 2010 to 2014

3.5.2 Conceptualisation of Spatial Weights and Neighbourhood

Since spatial analysis is in consistent with Tobler's First law of Geography, weighted neighbourhood matrix, a necessary step in using areal data, was created for all spatial analyses. Neighbourhood is those areas (districts) surrounding a given district (as an areal data point) that were considered to influence the observation at the data point (AnamzuiYa, 2012). These neighbouring areas are spatial units, interacting in a meaningful way such that the spatial analysis tool compares the target incidence rate to the neighbouring rates. Neighbourhood is usually defined based on distance: inverse distance, distance band and zone of indifference; and adjacency: K nearest neighbours, polygon contiguities and spatial weights (Murack, 2013; Tsai, 2012). Due to the highly irregular shapes and sizes of the administrative

45

districts, the most appropriate neighbourhood method used at all stages of the spatial analysis in this study was the first-order queen polygon contiguity (contiguity-edges-corners) as recommended in previous studies, including (Tsai, 2012) and Tsai *et al* (2009). The queen polygon contiguity incorporates both the Rook contiguity (that is neighbours sharing common edges) and Bishop contiguity (that is neighbours sharing common corners) relationships into a single measure. Based on this approach, the spatial weight or connectivity matrices were constructed according to districts which share common edges or/and corners.

3.5.3 Rate Smoothing and Standardisation

Rates from small cases and small size of the population at risk are usually unstable, producing dramatic changes in disease. The variance instabilities makes it difficult to compare the rates unless the population is constant or very large everywhere (Anselin *et al*, 2006). Hence in ensuring that mapping the morbidity rates account for the spatial heterogeneity in the population at risk and the random variables in rates, the crude rates were smoothen (Odoi *et al*, 2003) to avoid unbiased visual interpretation. Prior to smoothing the rates, the queen polygon contiguity was used to establish the spatial neighbourhood relationship. Empirical Bayesian smoothing (EBS) technique recommended in literatures among the several available standardization approaches was used to correct the variance instability (Sankoh *et al*, 2002; Waller & Gotway, 2004; Osei and Duker, 2008).The crude incidence rates were therefore smoothed using global EBS technique with GeoDa, an open-source software designed for geospatial analysis (Anselin, 2005). Unlike the local smoothing EBS which borrowed information from nearest areas, the global

smoothing approach used the values in all the neighbouring districts to improve the properties of the rate estimate by adjusting the uncertain values

(Toprak and Erdoğan, 2008). The final smoothed rates have a variance stabilising effect by considerably adjusting rates of districts with relatively small population and/or size (to the rate it should have been in a large population or size) and almost no change in districts with relatively large population or size (Sankoh *et al* 2002; Osei and Duker, 2008). The annual and the cumulative EBS rates were then mapped and classified based on the natural breaks classification technique in ArcGis version 10.1.

3.5.4 Estimation of relative risk

To put rates into meaning perspective, it is often more useful to compare it to a benchmark than considering it in isolation. This yields a rate ratio or relative risk (RR) defined as a ratio of the risk or health event (disease, injury, risk factor, or death) in one group (that is exposed group) to the risk among another group called the reference group or baseline (unexposed group). In aggregated analysis, the group with the lowest rate is taken as the reference group to estimate the relative risks for comparison. Thus the

RR was estimated from the smoothed cumulative rates as

$$RR = \frac{\text{Risk of disease in group of primary interest}}{\text{Risk of disease in reference group}}$$

The interpretations of the RR are as follows:

If RR = 1.0, it indicates identical risk among the two groups (districts)

If RR > 1.0, it indicates an increased risk for the group of primary interest than warranted by the benchmark risk.

If RR < 1.0, it indicates a decreased risk for the primary district of interest, indicating that perhaps available exposures are actually protecting against disease occurrence.

(US CDC, 2012; Bonita *et al*, 2006). A map of the relative risks, sometimes referred to as an excess risk map was then produced (Anselin *et al*, 2006) in ArcGIS.

3.5.5 Spatial patterns evaluation with spatial autocorrelation

Fundamental to the statistical analysis of dependence and heterogeneity in spatial patterns is the determination of whether or not identifiable spatial patterns exist. This is because spatial statistics as proposed in Tobler's law is generally based on the precept that nearby geo-referenced units are associated in some way (Getis, 1995). Finding the degree of spatial association (spatial autocorrelation) measures the correlation of a variable across geographical locations. A typical assumption in traditional statistical theory that observations on a single variable are independent (implying no spatial autocorrelation) may not be necessarily true. Tests for spatial autocorrelation were carried out in this study as an adequately suitable benchmark to identify statistically significant dependent or independent phenomena (Verma and Gupta, 2014). Among the number of spatial autocorrelation measures, stated earlier in section 1.1.4.2, the most useful among them is the Moran's index, I. Hence

in exploring the spatial dependence to show how the incidence rates autocorrelated among neighbouring districts in the country, a Global

Moran's I was estimated (Tsai, 2012; Osei, 2008; Toprak and Erdoğan, 2008). The Global Moran's I statistic, similar to the Pearson correlation coefficient but incorporated spatial component is simplified and estimated as:

$$I_M \square \overline{nwij} \square (R_i (\square RR_i)(\square \overline{RR_j}) _2 \square R)$$
$$\square$$

i and j = two areal units (districts), n = number of weights = number of districts wij= the element in the spatial weight matrix corresponding to the observation pair i, j to define the spatial interaction among the districts in the study area. The weights were rowstandardized such that each row sums to 1 ($\Sigma wij = 1$) if districts i and j share a common boundary; otherwise $\Sigma wij = 0$, for non-neighbouring districts.

R_i and R_j = smoothed cumulative incidence rates for the neighbouring districts i and j

$\overline{R}$ = mean incidence rate of infection

The spatial weights matrix containing information on the neighborhood structure of each district was created based on the queen polygon contiguity edges-corners

conceptualisation of spatial relationships described above.
The ArcGIS spatial statistic tool was then used to compute the Global Moran's *I,* z-score value and p-value to evaluate the significance of the index in describing the degree of spatial concentration or dispersion for the disease incidence.

Interpretation of Moran's I

The spatial test statistic was computed for a significance test against the null hypothesis of no spatial autocorrelation. Usually, if spatial autocorrelation exists, it will be exhibited in similarities and dissimilarities between contiguous districts as indicated by numbers between -1 to 1.

If I= 0, implies random distribution; null hypothesis of no spatial dependency is accepted. On the contrary, fail to accept null hypothesis if

I > 0, then rates are Clustered

I < 0, rates are Dispersed

Also, reject null hypothesis if z-test statistic > 1.96 (or < -1.96) at 95% confident level (less than a 5% chance that, in the population) that spatial auto correlation exits (Anselin,

1995; Verma and Gupta, 2014).

3.5.6 Cluster-outlier detection and Hotspot Analysis

The global spatial autocorrelation analysis shows clustering as to whether the pattern of infectious diseases among the districts is clustered, dispersed, or random but cannot indicate where specific patterns of clusters and outliers occur. To identify concentrations of high and low values, and spatial outliers, various local versions of Moran's I, Geary's c, and the General G statistic for each areal unit are often used (Anselin,1995). This study used the Anselin Local Moran's I_i in the ArcGIS for the cluster-outlier analysis to detect areas of unexpectedly high rates of infection across the country. For this, local Moran's I value, a z-score, p-value, and a code representing the cluster type for each feature were calculated. Being a

50

relative measure, the local Moran's I index was interpreted within the context of the computed z-score or p-value at 95% significance level (P<0.05). The cluster-outlier (CO) type therefore enabled distinction to be made among a statistically significant similarly clusters of high values (High-High, HH, called hotspot) or low values (Low-Low, LL, called coldspot) and dissimilarly outliers whereby high value is surrounded primarily by low values (High-Low, HL) or low value is surrounded primarily by high values (Low-High, LH).The cluster-outlier detections were computed on a yearly basis and then the overall period. Mapping highlighted the clustering patterns of the disease by displaying on the spatial statistical maps from 2010 to 2014, specific locations with significant value of hotspots and coldspots (Verma and Gupta, 2014; Tsai, 2012). These measures identified clusters of points with values higher in magnitude than that expected to be found by random chance at 95% significance level (Bradshaw, 2013). A statistically significant hot/cold spots district is surrounded by districts with similar values, but with significantly higher/lower values than its neighbours (Toprak and

Erdoğan, 2008;Verma and Gupta, 2014).

Interpretation

Positive I_i value: $I_i > 0$ and additionally for a z-score value larger than +1.96 Then the feature (the district) is defined as surrounded by districts with similar values, either high or low (HH or LL), hence a cluster.

If $p < 0.05$, then the clusters of high values (HH) or low values (LL) are statistically significant at 95% significance level.

Negative I_i value: $I_i < 0$ and additionally with a z-score value less than -1.96, district is surrounded by districts with dissimilar values; either HL or LH, thus an outlier. If $p < 0.05$, then the areal units are statistically significant outliers (HL or LH).

For statistically significant positive z-scores, the larger the z-score, the more intense the clustering of high values (hotspot). But for statistically significant negative z-scores, the smaller the z-score, the more intense the clustering of low values (coldspot). With colour codes and abbreviations, a spatial statistics maps of the Local Moran's index identified the outliers as either HL or LH and the clusters as coldspot (LL) or hotspot (HH). The tests for hot spots served as a focused test that emerged from the pack if its local structure is sufficiently unusual. This approach provided some protection against the biases that may arise when only selected areas were tested (Ord and Getis, 1995). Hotspot analysis is an important measure in spatial epidemiology because even clusters of cases occurring randomly can have an influence on the spread of infection. Identification of hotspots (HH) draw attention of policy makers and healthcare providers for strategic healthcare management and interventions (Verma and Gupta, 2014).

3.5.7 Modeling spatial relationships: spatial regression modeling

3.5.7.1 Correlation and Selection of Predictors

All the determinants processed above were used as independent variables (potential predictors) and the smoothed cumulative incidence rates as dependent variable (response) for correlation analysis in SPSS. A Pearson's correlation coefficient was

then computed. Pair-wise correlation of the predictors was also carried out to check for multicollinearity

(where a predictor correlated significantly with other predictors besides the response). Multicollinearity poses a lot of problems (such as regression coefficient being very high with opposite sign and not significant as against known theory and most previous works etcetera) in multivariate relationship and makes the model dubious. Hence a stepwise multiple regression model building approach was used to address the multicollinearity by removing redundant predictors. As a result, the maximum number of statistically significant independent variables were selected as candidate predictors for the spatial modeling. This improved the performance and credibility of the spatial model as well as less data collection requirement for future application of the spatial predictive models.

3.5.7.2 Spatial Model building and Model Selection

All the risk factors showing a statistically significant association with infection and passing the stepwise multiple regression selection were considered and included in the spatial regression modeling as candidate predictors. The spatial linear regression models within the ArcGIS framework are Ordinary Least Square regression(OLS) and Geographically Weighted Regression (GWR). The OLS, a starting point for all spatial regression analyses, is a global model that performs global linear regression while GWR is a local model to model spatially varying relationships among datasets over the space.

GWR model incorporates regional variation into the regression model (Murack, 2013). Before modeling, the queen's polygon contiguity edges-corners conceptualisation of weighted spatial relationships was created. The two linear spatial regression models were performed with the smoothed cumulative incidence rates as the dependent variable (response) and the significant selected determinants as the independent

variables(candidate predictors). Because the predictors could be spatially autocorrelated, other spatial regression models, Spatial Lag Model and a Spatial Error Model were tested in GeoDa to check for the improved model. The Spatial Lag Model incorporates a spatially lagged dependent variable to assess the existence and strength of spatial interaction among neighbours. The Spatial Error Model assumes that the errors of a model are spatially correlated, so that the model's deficiencies are similar among the neighbours throughout the study area (Anselin, 1999 cited in Bradshaw, 2013). The parameter estimates and model diagnostics were used for the performance of the candidate predictors and model assessment respectively. The residuals (observed minus predicted dependent values) which is the unexplained portion of the dependent variable was mapped and the spatial autocorrelation run on the residuals to check for the completeness of the model in finding out if there is misspecification (omission of key determinants) or not. A significant residual clustering means there could be misspecification (that is a key variable is missing from the model). Since the adjusted- R^2 values for Spatial Lag and Spatial Error models are pseudo values, model comparison and selection was based on corrected Akaike Information Criterion (AICc) which measures the best relative goodness of fit of a model. The model with lowest AICc value has the

highest goodness-of-fit and was selected as the best predictive model for the dataset (Murack, n.d; Osei, 2010). The Geographically weighted regression (GWR) was used as secondary check on the non-stationarity in the datasets (a condition where a single global model cannot explain the relationship in the datasets due to local variations). This looked for and indicated where non-stationarity was significantly taking place on the map. GWR estimated an OLS-like regression for each feature to indicate that different relationships exist at each point in space for which reason the fitted coefficient values of a global model may not adequately represent the detailed local variation (Bivand, 2014).

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 DATA ANALYSIS AND RESULTS

4.1.1 Geo-visual spatial exploratory analysis of infection

The smoothed incidence rates and determinants were geocoded with district identification codes (Appendix I) and implemented as attribute data for the spatial epidemiological analysis in ArcGIS environment.

The ranking order of incidence among the investigated top five infectious diseases was Malaria, Upper Respiratory Tract Infection (URTI), Diarrhoeal diseases, Intestinal worms and Typhoid fever. The geovisual exploratory data analysis of the incidence rates presented as smooth cumulative incidence maps of the five leading infectious diseases reflected a non-random spatial pattern of possible clustering or

dispersion of infections. The graduated colour quantities and legend gives the visual impression and the numerical range of the rates respectively. Fig 4.1 showed a fairly high malaria rates in eight districts

(Districts 4, 33,103, 116, 152, 155,156 and 159), most of which are in the Upper and Brong-Ahafo regions. Eight districts (districts 2,4,6,8,24,111,118 and168) again reported high rates of URTI (Fig 4.2). Even-though the smooth incidence map of Diarrhoeal diseases (Fig 4.3) showed quite uniform possible clustering and dispersion across the country, some districts (districts 2,4,4,8,103,111 and 168) recorded relatively high rates. The spatial incidence of intestinal worms in Fig 4.4 showed comparatively low risk in the northern part but higher incidence in the middle and southern part of the country, most of which were in the Brong-Ahafo region.
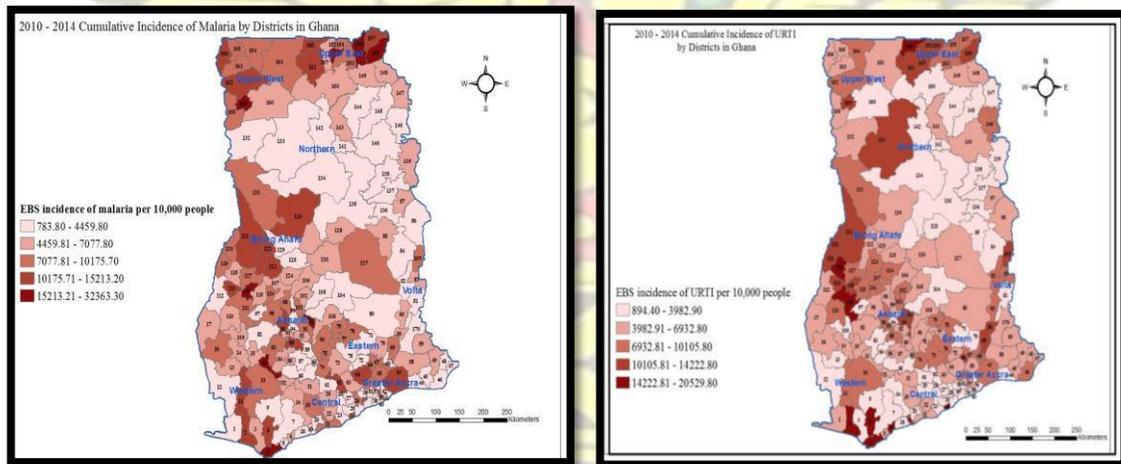

Fig 4.1 Cumulative incidence map of Malaria Fig 4.2.Cumulative incidence map of URTI
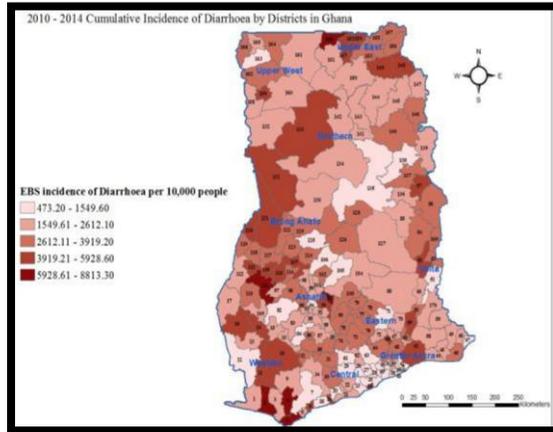
Fig 4.3 Cumulative incidence map of
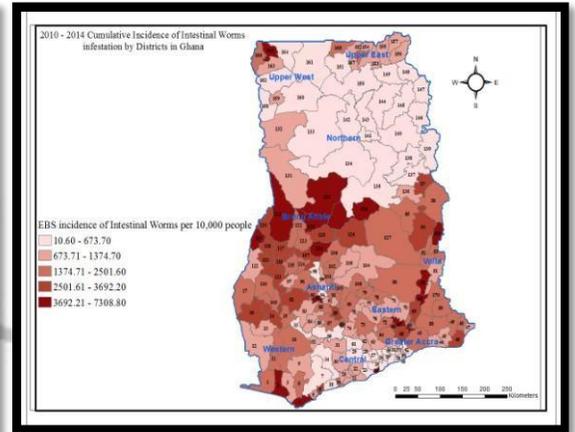Diarrhoeal diseases

Fig 4. 4. Cumulative incidence map of
Intestinal worms infestation
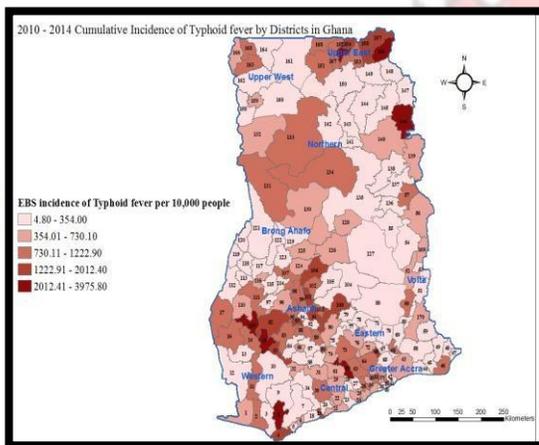
Fig 4.5. A cumulative incidence map
of Typhoid fever
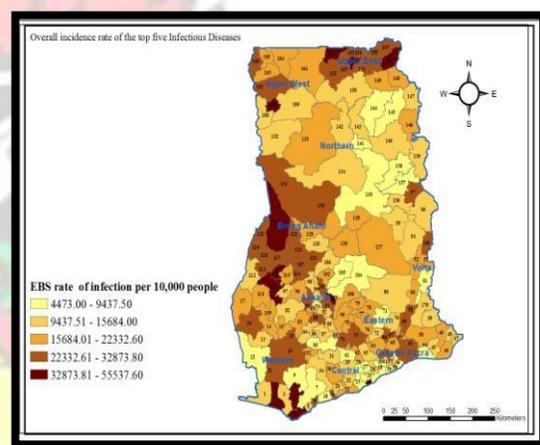
Fig 4.6. A cumulative incidence map of the
top five infectious diseases

The visual observation of Fig 4.5 indicated clear spatial variability in the incidence of Typhoid fever in most districts across the country but reasonably high in seven districts (districts 8, 15, 62,109,146,152 and156). Within the five-year period of the investigation, the result showed that the overall incidence of infection from the five infectious diseases depicted great variation in the country with respect to absolute location (Fig 4.6).
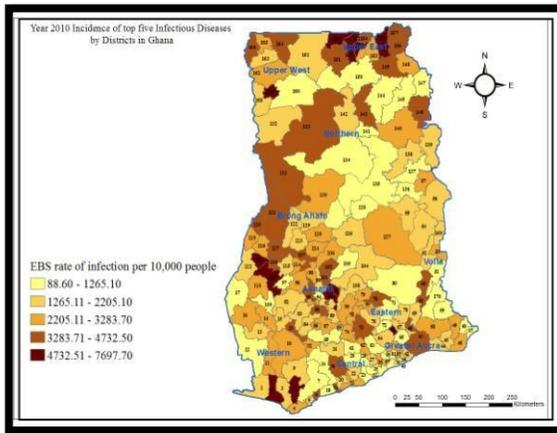
Fig4.7A 2010 incidence map of infection   Fig 4.8 A 2011 incidence map of infection
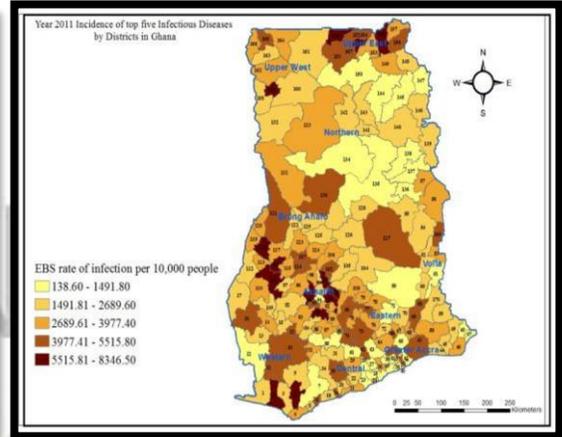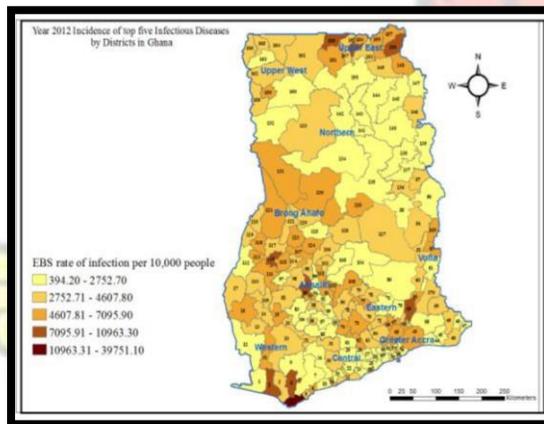
Fig 4.9 A 2012 incidence map of infection   Fig 4.10 A 2013 incidence map of infection
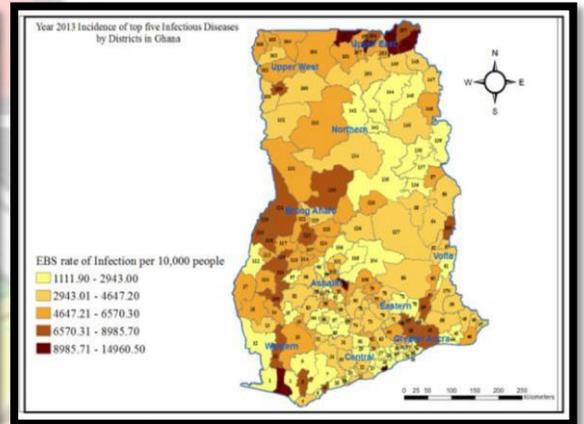
Generally, the results of the geovisual exploration of the annual incidence of infection showed yearly increment in the incidence and spatial variability over the periods 2010 to

2014 (Fig 4.7 to Fig 4.11).

With reference to the average minimum infection rate of 4,473 cases per 10,000 people in district 135 (East Gonja), the rate ratio or relative risk map showed that seven districts (2, 4, 8, 103, 116, 152 and 168) had 8.2 to 12.4 times higher infection rate as compared to the reference district.
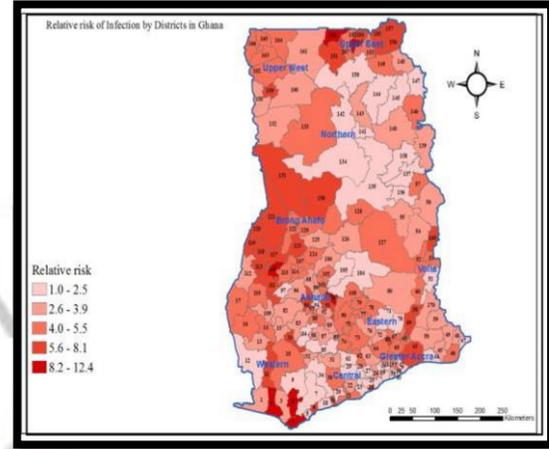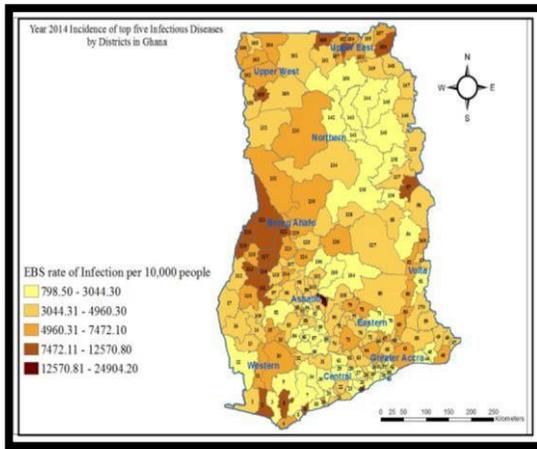
Fig 4.11A 2014 incidence map of infection   Fig 4.12 A relative risk map of infection

4.1.2 Spatio-temporal patterns of Infection

The results of the spatial autocorrelation analysis for evaluating the observed spatiotemporal patterns is given in Table 4.1. The result showed a positive and statistically significant Moran's I, reflecting significant accumulation of high rates (clustering) for four of the five infectious diseases and the overall rates of infection (Moran's I > 0 and Zscore > 1.96 at p< 0.05).The URTI, however, with negative Moran's I was an indication of dispersion which is not significant (Moran's I < 0 and Z-score < -1.96 at p > 0.05). A further assessment of the level of geospatial patterns in the incidence of infection is the magnitude of the z-score, indicating great differences in the intensity of spatial clustering in the incidence of the diseases. The temporal spatial patterns (Table 4.2) indicated statistically significant clustering for the years 2012 and 2014 but non-significant clustering in 2011.

Table 4.1 Spatial autocorrelation on the cumulative incidence of the infectious diseases

Disease	Moran's I	P-value	Z-score	Spatial pattern evaluation
Malaria	0.1642	0.0003	3.6004	Significant Clustering

URTI	-0.0504	0.3495	-0.9355	Non-significant Dispersion
Diarrhoeal diseases	0.0953	0.0355	2.1025	Significant Clustering
Intestinal worms	0.3260	0.0000	6.9148	Significant Clustering
Typhoid fever	0.1371	0.0024	3.0303	Significant Clustering
Overall Infection	0.1720	0.0003	3.7100	Significant Clustering

Table 4.2. Spatial autocorrelation analysis of infections for the periods 2010 to 2014

Year	Moran's I	P-value	Z-score	Spatial pattern evaluation
2010	-0.0093	0.9444	-0.0698	Non-significant Dispersion
2011	0.0575	0.1877	1.3175	Non-significant Clustering
2012	0.1131	0.0028	2.9867	Significant Clustering
2013	-0.0644	0.2173	-0.2339	Non-significant Dispersion
2014	0.1569	0.0041	3.5329	Significant Clustering

The years 2010 and 2013 were, however, non-significant dispersion at the 5% significant level. In and broader perspective over the periods, infection has moved from dispersion in 2010 to the greatest clustering in 2014 (highest z-value = 3.5329). Generally, the spatiotemporal analysis confirmed the non-random yearly increasing incidence and clustering of the infections as observed geo-visually in the smooth incidence maps.

4.1.3 Identifying areas with high risks of infection: hotspots detection

With different colour codes and abbreviations, the results of the cluster-outlier analyses were shown in Fig 4.13 to Fig 4.23 for the investigated infectious diseases and the annual rates of infection. The spatial statistical maps identified specific locations on the map with statistically significant clusters of high rates (hotspots), or low values (coldspots) and the outliers with respect to specific diseases and the annual rates of infection from the top five infectious diseases.
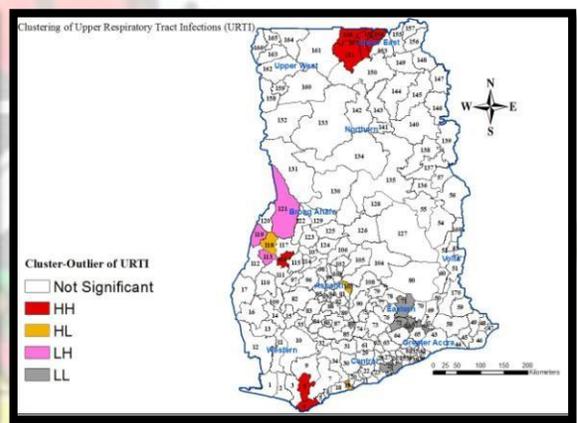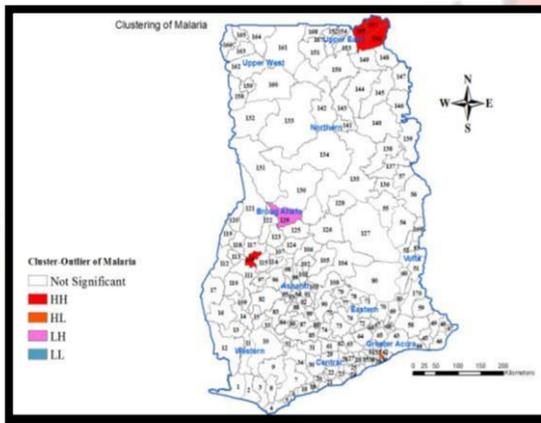
Fig 4.13 Cluster-outlier detection of Malaria Fig 4.14 Cluster-outlier detection of URTI
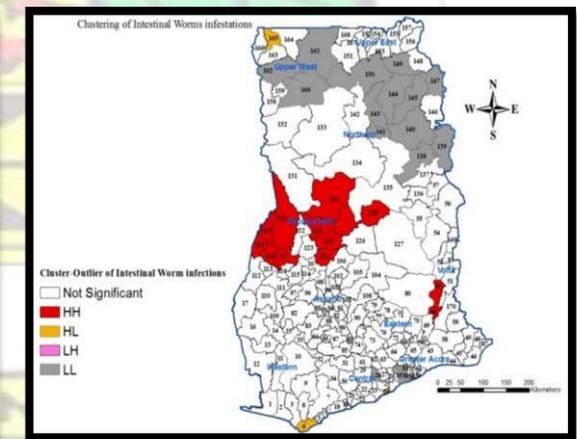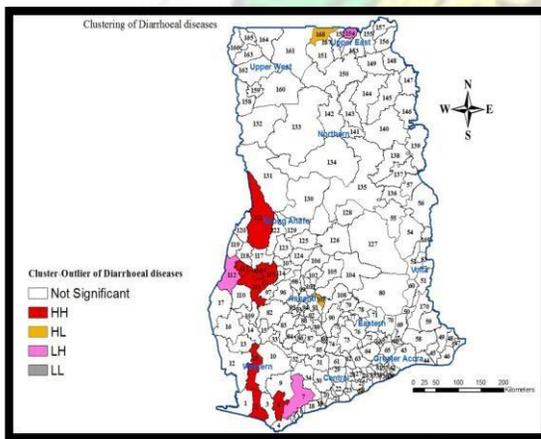
Fig 4.15 Cluster-outlier detection of   Fig 4.16 Cluster-outlier detection of
         Diarrhoeal diseases   Intestinal worms
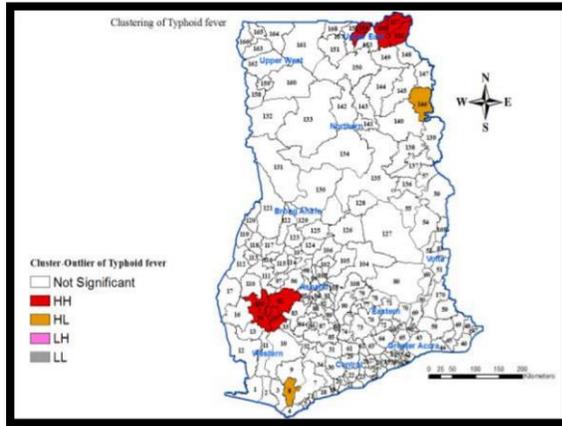
Fig 4.17 Cluster-outlier detection of Typhoid fever
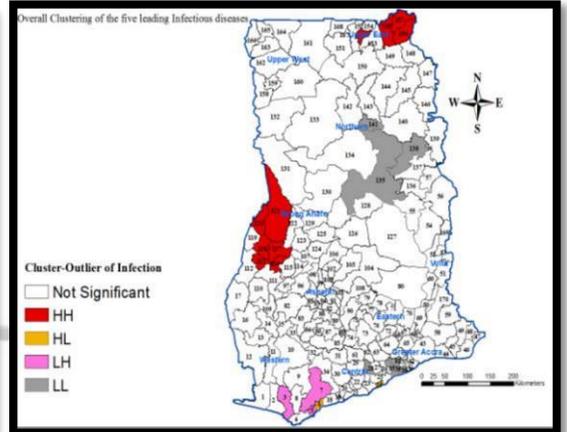
Fig 4.18 Cluster-outlier detection of Overall Infection
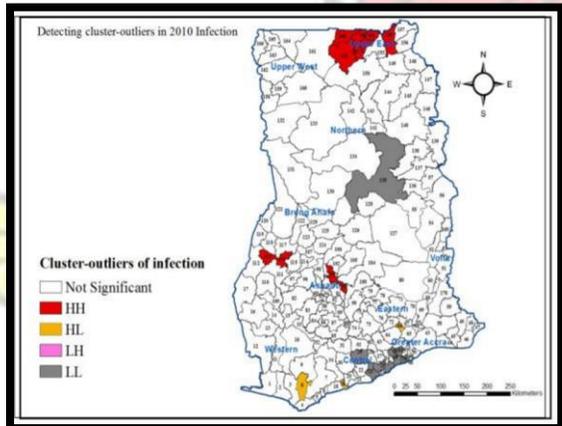
Fig 4.19 A 2010 cluster -outlier detection of Infection
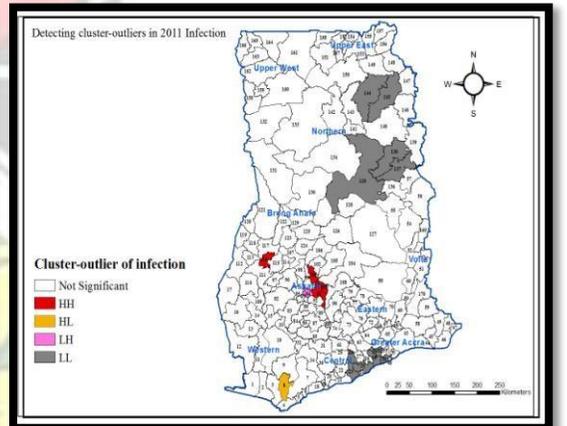
Fig 4.20 A 2011 cluster -outlier detection of Infection
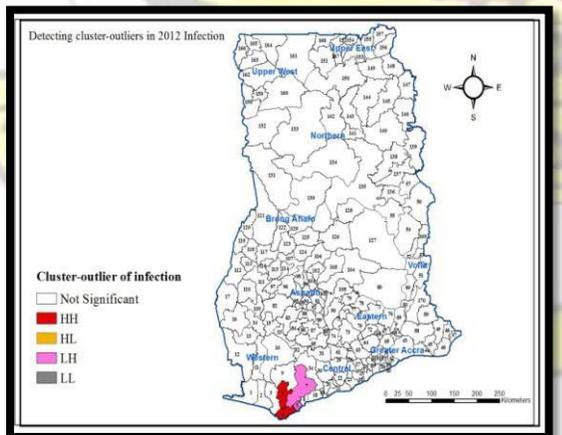
Fig 4.21 A 2012 cluster-outlier detection of Infection
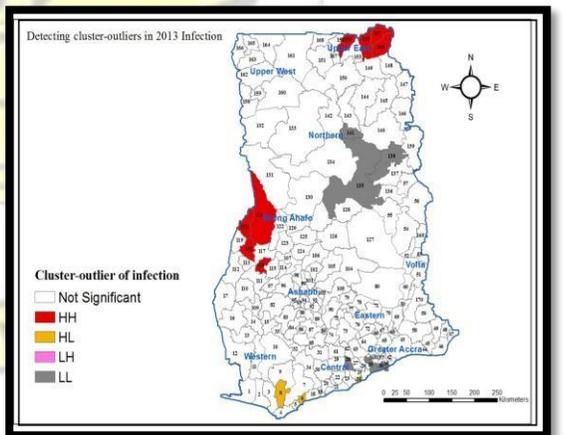
Fig 4.22 A 2013 cluster -outlier detection of Infection
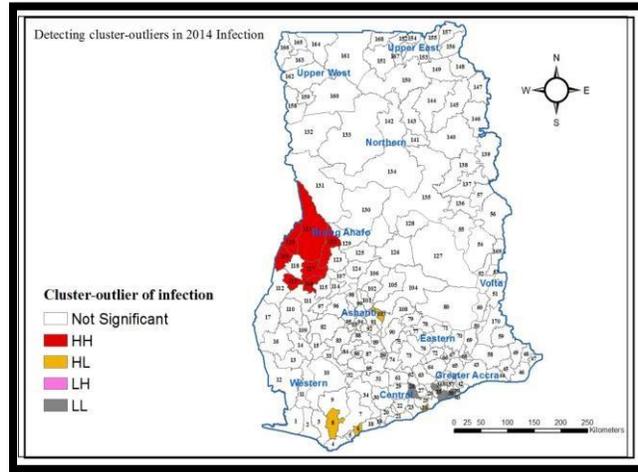
Fig 4.23 A 2014 cluster-outlier detection of Infection

From the spatial statistics maps of the cluster-outliers analysis, the disease-specific

and year-specific hotspot locations were described and summarised in Table 4.3

and Table

4.4 respectively below.

Table 4.3 Summary of hotspot locations of the top five infectious diseases in Ghana

Disease	Hotspot district codes	Hotspot Regions
Malaria	155,156,157,116	3 in Upper East;1 in Brong-Ahafo
URTI	151,152,154,167,168, 116, 4, 8	5 in Upper East and 3 in Western
Diarrhoea	121, 113, 115, 111, 116, 11, 8, 2	5 in Brong-Ahafo and 3 in Western
Intestinal worms	117,118,119,120,121,124,125,128,129,130, 50, 60	10 in Brong-Ahafo and 2 in Volta
Typhoid fever	152,154,155,156,157,82,14,15,109	5 in Upper East, 2 in Western, 1 in Ashanti and 1 in Brong-Ahafo

63

Overall Infection	113,116,117,118,120,121,152,155,156,157	6 in Upper East and 4 in Brong-Ahafo

Table 4.4 The annual hotspot locations of infection from the top five infectious diseases

Year	Hotspot district codes	Hotspot Regions
2010	151,152,154,155,167,168,113,116,101,103	6 in Upper East, 2 in Brong-Ahafo, 2 in Ashanti
2011	91,101,103,116	3 in Ashanti and 1in Brong-Ahafo
2012	4,8	Both in Western
2013	152,154,155,156,157,116,118,120,121	5 in Upper East, 4 in Brong-Ahafo
2014	113,116,117,119,120,121,122	All in Brong-Ahafo

4.1.4 Significant Determinants and their Spatial variability

The Pearson correlation analysis disclosed that out of the total of 33 risk indicators

assessed, 20 showed to be contributing significantly to the incidence of infection

(Table 4.5).

Table 4.5. Significant determinants of infection

Significant risk indicators of infection

Sex ratio*[c]

Proportion of population aged 0 - 14 years**

64

Proportion of population aged 15 - 64 years**

Muslim**

No and other religion**

Not literate*

Basic education**c

Secondary education**

Inter-migration*c

Urbanisation level**c

Overcrowding Index (OI)*

Unsafe drinking water**

Unimproved Toilet facility**

Unimproved Bathing facility*

Unimproved Solid waste disposal*

Unimproved Liquid waste disposal**

Household Smoke Exposure Index**

Traditional Housing Index*

Non-owner occupied housing units**

Employment-to-population ratio (EPR)**

* statistically significant at $p < 0.05$;   ** statistically significant at $p < 0.01$ c: candidate predictors

The results of the geovisual analyses and mapping of the significant determinants were given in Fig 4.24 to Fig 4.39. The great differences in the intensity of the

graduated colour as shown in the determinant maps indicated the degree and spatial variations of a particular determinant among the districts across the country. The results obviously showed on the risk maps that for each risk factor, some districts had relatively high measures and with great spatial variations. The insanitary condition risk map (Fig 4.34) was a proportional sum of the unsafe or unhygienic housing facilities; main source of drinking water, toilet and bathing facilities, solid and liquid wastes disposals.
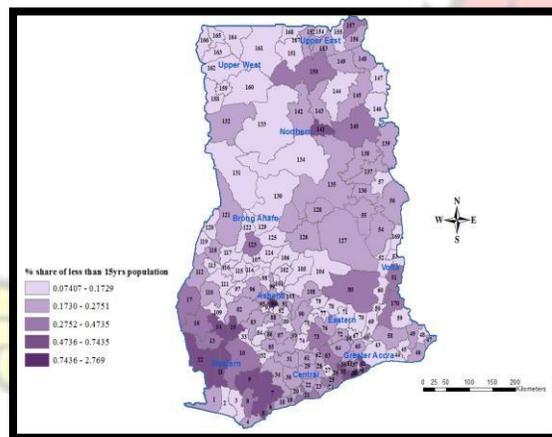
Fig 4.24 The proportion of the total population   Fig 4.25The proportion of the total
less than 15 years by districts   population aged 15 - 64 years
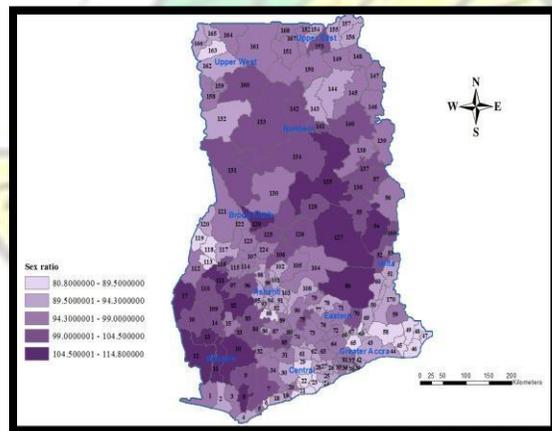
Fig 4.26 The number of males per 100 females Fig 4.27 The proportion of the population
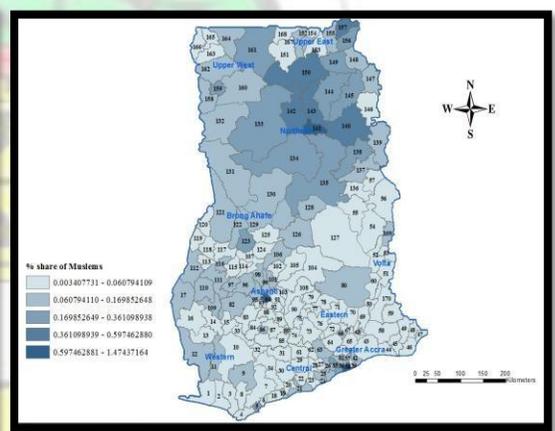
practicing Islamic religion
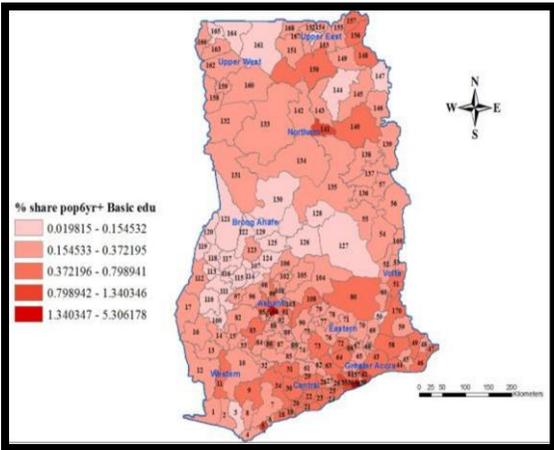
Fig 4.28 The proportion of the population aged 6 years and older with Basic education
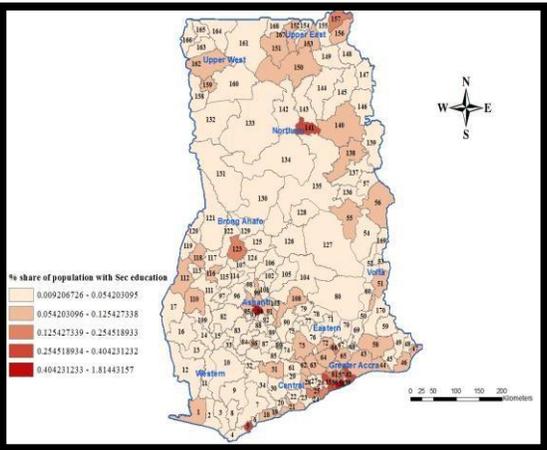
Fig 4.29 The proportion of the population with Secondary education
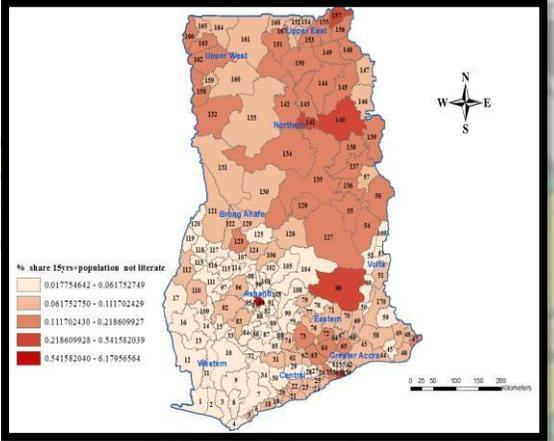
Fig 4.30 The proportion of the population aged Fig 4.31 T   he proportion of the total
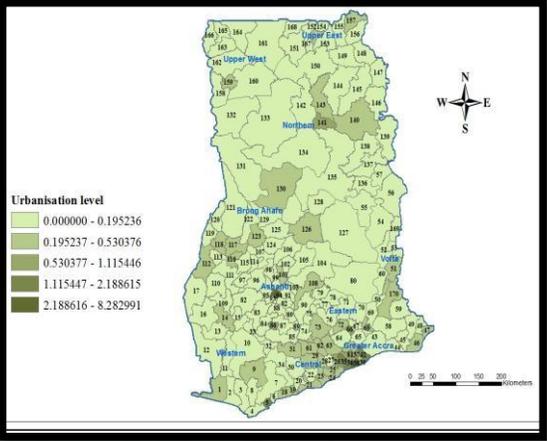
population 15 years and older that are illiterate   that are urbanised
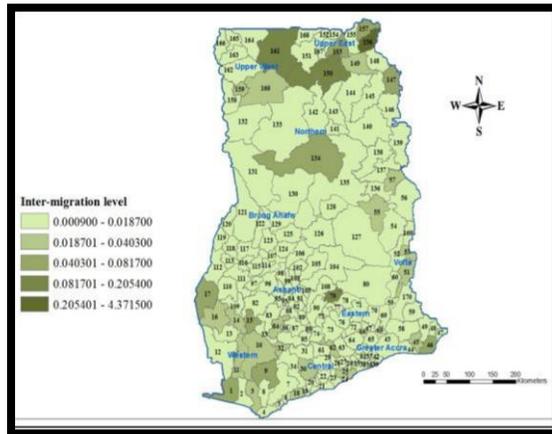
Fig 4.32 The inter-migration level
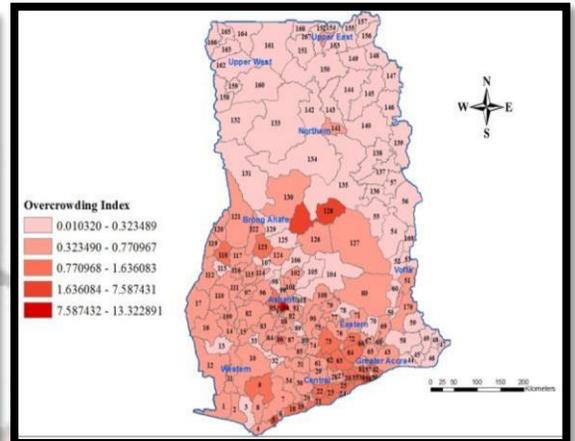
Fig 4.33 The overcrowding level
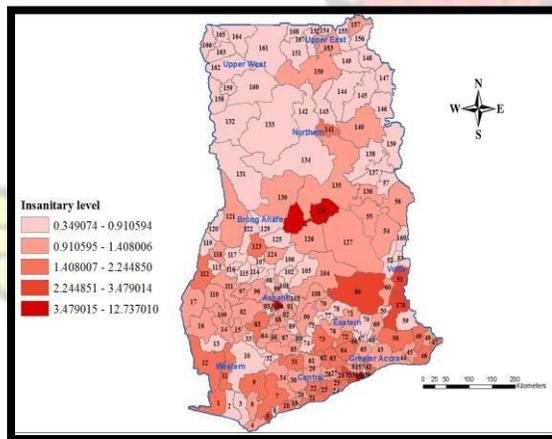
Fig 4.34 The level of insanitary conditions
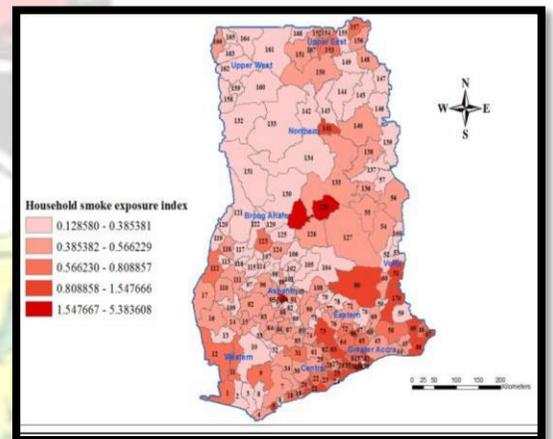
Fig 4.35 Household smoke exposure level
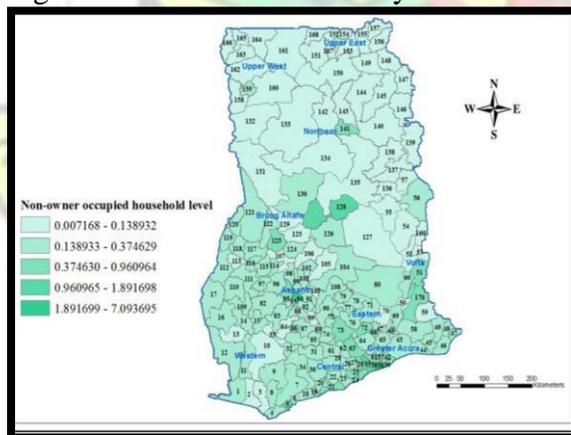
Fig 4.36 The proportion of non-owner
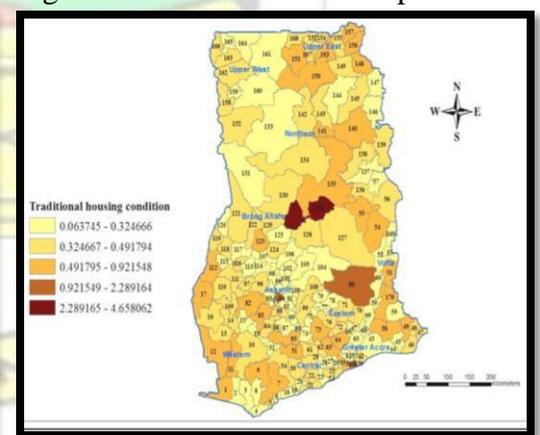
Fig 4.37 The proportion of the traditional

occupied housing conditions   housing conditions
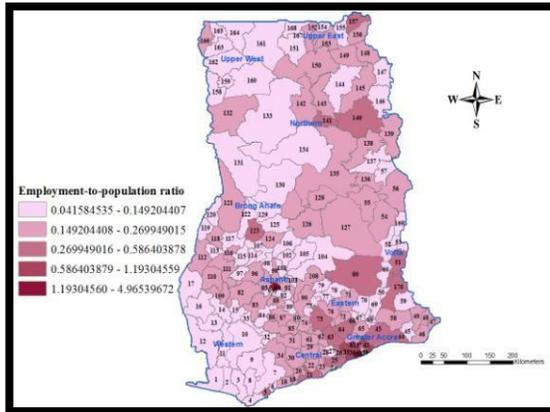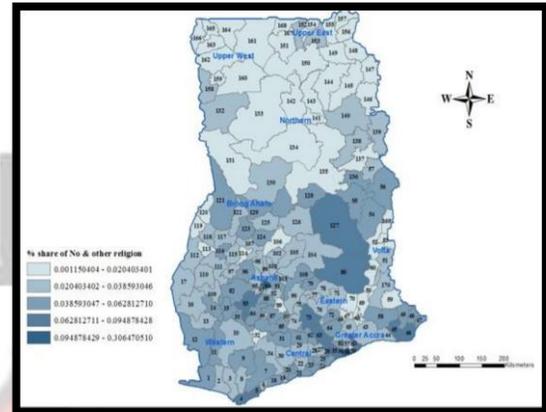
Fig 4.38 Employment-to-population ratio   Fig 4.39 No or other religious affiliations

4.1.5 Modeling Spatial Relationships

High multicollinearity was observed among the significant risk factors and most of the risk factors positively correlated with urbanisation level, overcrowding and sanitation conditions, non-owner occupancy and employment-to-population ratio significantly. However, the result of the stepwise multiple regression model building approach selected a maximum of four significant explanatory risk factors; Basic Education, Urbanisation level, Intermigration and Sex ratio as candidate predictors for effective predictive modeling of infection. Spatial regression modeling and model selection isolated Spatial Lag model as the model with the best goodness-of-fit measure (model with the lowest

AICc, Table 4.6)

Table 4.6 Model selection criterion

Model	AICc
Ordinary Least Squares(OLS)	3564.25
Spatial Lag model	3352.05

Spatial Error model   3561.28

Diagnostics assessment of the selected model (Table 4.7) indicated significant spatial dependency, normal distribution of errors (implying that the model is not biased) and possible stationarity of predictors over space (since Breusch-Pagan tests of nonstationarity were not significant). The Moran's I on residuals was a significant dispersion, indicating non- clustering of missing key predictors.

Table 4.7 Assessment of the Spatial Lag model

Model diagnostic measure	Value	P-value	Model assessment
Corrected Akaike info criterion, AICc	3352.05		Performance
Likelihood Ratio Test (df = 1)	83.481	0.00000*	Spatial dependence
Breusch-Pagan test (df = 4)	8.6989	0.06908	Non-stationarity of predictors
Moran's I on residuals	-0.1009	0.0377*	Misspecification and model biasness

The output of a further model diagnostic tool using Geographically Weighted Regression (GWR) modeling indicated possibility of missing variables and non-stationarity of predictors in very few districts where the models is over-predicting (red colour) and under-predicting (blue colour), Fig 4.40

70
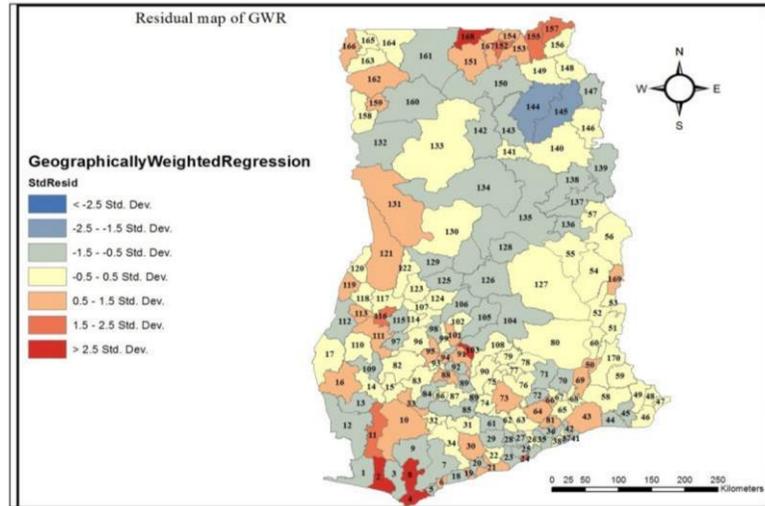
Fig 4.40 Residual map of Geographically Weighted Regression modeling

The parameter estimates of the selected model of all the determinants (Table 4.8), except Basic education, related positively to the incidence of infection but none of them showed to be statistically significant (p > 0.05).

Hence from the result of the Spatial Lag model parameter estimates (Table 4.8) and the model's equation (Appendix IId), the incidence rate of infection (*R*) from the top five infectious diseases at particular district in Ghana could be predicted as

$$R = 0.713833W_{ijR} + 3385.134 + 22.9913(\text{Sex ratio}) - 2391.13$$

$$384 \quad .0142(\text{Urban level}) + 1407.215(\text{Intermigration}) + 4157.8$$
(Basic edu) +

where W_R = spatially lagged dependent variable for weights matrix W_{ij}

W_{ij}= the element in the spatial weight matrix corresponding to the observation pair of districts i, j to define the spatial interaction among the districts in the study area.

71

Table 4.8 Spatial Lag model-maximum likelihood parameter estimations

[n= 170, df =164, Mean Rate =18209.2, S.D = 5989.4, S.E of regression= 4157.8]

Variable	Coefficient	Std. Error	z-value	P-value
Lag coeff. (Rho)	0.713833	0.0573	12.4540	0.00000*
Constant	3385.134	5536.44	0.6114	0.54092
Sex ratio	22.9913	56.4371	0.4074	0.68373
Basic education	-2391.13	2367.834	-1.0099	0.31257
Urbanisation level	384.0142	963.1658	0.3987	0.69011
Inter-migration	1407.215	1477.828	0.9522	0.34099

4.2 DISCUSSION

A number of studies have shown that the appreciation of the space-time effect on the non-random distribution of diseases warrants the GIS-based spatial epidemiological studies of infections.

The smooth maps of the disease rates provided apparent geovisual indication of the incidence of the five leading infections in Ghana in the order Malaria, URTI, Diarrhoeal diseases, Intestinal worms and Typhoid fever. The visual evaluative inspection of the smooth incidence maps is suggesting high incidence rates and possible high spatial variability of the infectious diseases across the country. This

reaffirm these infectious diseases as the leading causes of morbidity in Ghana

(Adams *et al*, 2004) and reflects the

Global Burden Diseases study's ranking of malaria, respiratory infections and diarrhoeal diseases among the topmost causes of mortality globally(GBD, 2014). In most of the instances, districts of the Upper West and East and Brong-Ahafo regions were highly infected. Assessment of the spatial patterns in the incidence of the diseases with global Moran's I showed a high extent to which neighbouring rates of infection correlated. This suggested statistically significant spatial clustering for four of the five topmost infectious diseases with increasing intensity of clustering as measured by the magnitude of the zscores. The higher the z-score, the more intense the clustering of high values as hotspots while smaller z-score values indicate more intense clustering of low values as coldspots. It therefore implies that from coldspot towards hotspot, the order among the top five infectious diseases were URTI, Diarrhoeal diseases, Typhoid fever, Malaria and Intestinal worms (Table 4.1). This is to state that the accumulation of high rates was greatest in Intestinal worm and least in URTI. It therefore implies that Intestinal worm requires additional intervention, taking into account the spatial influence. Also the increasing incidence of infection over the years with highest clustering of high rates in 2014 (Table 4.2) could possibly mean that factors promoting infections in the country are likely to be increasing as the population increases with time. It could also be deduced from the space-time analysis of the infections that the occurrence of a disease in one area is significantly influencing and is influenced by the rates in the connected neighbourhoods as well as the rates in the previous years. This finding lines up with the reviewed studies with regard to the spatial dependency of infectious diseases, implying neighbourhood-effects in the transmission of infections. Therefore by principle, any health intervention at a place should additionally consider what is happening in the neighbouring locations to some extent due to social interactions, especially among the connected or close neighbours. The high incidence rates, high relative risk of 12.4 times in some areas and significant clustering implies that despite several health intervention programmes, some areas (districts) in the country are still experiencing an unusually high risks of infection (especially Intestinal worms) as compared to their neighbours. It hereby means that to aid policy making and optimal allocation of limited resources, synergistic link between efficiency and equity is needed. This link as proposed in the recent economic thinking is essential to justify intervention programmes on social health determinants by governments and stakeholders (WHO, 2013). This justifies the importance of identifying the disease-specific areas with the elevated risks of infection (that is hotspot areas) as compared to the surrounding neighbours. From the clusteroutliers and hotspots analyses maps and Table 4.3, as many as eight districts each were specifically identified as having an unusually high incidence of URTI and Diarrhoeal diseases while four, nine and twelve were hotspot districts respectively for Malaria, Typhoid fever and Intestinal worms. This means that as compared to the neighbouring districts, these districts

are having significant unusual high incidence rates of infection per the specific infections and over the study periods of 2010 to 2014. Although is the topmost cause of morbidity, the result with just four hotspot locations for Malaria implies that the health intervention approaches to address Malaria seemed to be relatively quite uniform as compared to the rest, especially Intestinal worms which had the most intensified clustering magnitude and with as far as three times the number of hotspots of Malaria. The elevated risk of infection at hotspot areas could be as a result of inherently high prevailing risk factors or inappropriate allocation of health resources and intervention programmes. These hotspots districts (most of which are found in Upper East, Brong-Ahafo and Western regions) will be requiring relatively greater or additional health intervention programmes to further reduce the elevated risks of infection to at least a comparative degree as in their neighbours. This may also demand possibly altering of disease specific prioritisation and intervention programmes per district to contain the infections.

Because the investigated diseases are among the infectious diseases of poverty, disproportionately affecting the poorest populations in the world (Amo-Adjei, 2013; Bhutta *et al*, 2014), spatial assessment of key determinants to the incidence of these infections was found to be very important. Identifying as many as 20 significant determinants and with a number of risk factors inter-correlating significantly clearly reiterated the interwoven and complexity of the factors contributing to transmission of infection. This is largely due to the fact that some risk factors to some extent explain others. Example as clearly seen in the analysis was that areas with high urbanisation level (less rural) were found to be overcrowded with the associated high insanitary practices, traditional or improvised housing units (due to inability to afford good houses), high level of non-owner occupancy, increase population mobility due to high commercial activities and so forth. These then lead to undue pressure on social and health facilities including lack of access to safe drinking water, increase rate of slums and/or squatter formation with the corresponding increase in unhygienic practices. There are also

low likelihood of the poor in the urban centres affording the costs of quality education and healthcare. The effect of non-owner occupancy could also be due to the level of care and maintenance of the household facilities or sanitation and environmental hygiene in general which might be comparatively better if the housing unit is occupied by the owners. The cumulative consequence of all these is the greater risk and rapid transmission of infection. These findings aligns with those established in the reviewed literatures and other countless scientific researches as known health disparity indicators defining differences in health outcomes. The use of technology reported in other studies, including Saker *et al* (2004) to be promoting health did not show correlation in this study. This could possibly mean that the use of technology (mobile phones, desk/laptops and internets) by Ghanaians are not significantly geared towards health promotion. The risk maps of the nationwide-adjusted significant health determinants indicated that classical risk factors such as demographics, built and social environmental factors, economic factors etcetera influencing disease occurrence and re-occurrence also varied geographically in the population at risk as observed in the infections. That is to state that as reported in previous works, including Waller and Gotway (2004), Ostfeld *et al* (2005) and elsewhere, mapping of determinants as the first step to assessing the spatial contribution of determinants to the incidence of infection highlighted the spatial variation in the underlying prognostic factors promoting the spatio-temporal dynamism of infection. This implies and agree with the spatial disease concept that diseases are likely to occur in some members of the population more than others (Dhama *et al*, 2013; Gazzinelli *et al*, 2012) as explained obviously in evaluating the spatial patterns and identification of hotspot areas of infection. A critical visual

75

comparison of spatial variations of determinants further reaffirmed the interconnectedness of most of the risk factors in driving the transmission of infection as outlined in the chain of infection (Fig 1.1) and high multicollinearity. As a rule of thumb in multivariate relationship, some of these significant determinants (independent or explanatory variables) are inherent predictors of incidence of infection (dependent or response variable) through other significant predictors. This was evident in the inclusion of a maximum of only four (Basic Education, Urbanisation level, Intermigration and Sex ratio) of the twenty significant risk factors as candidate predictors through the predictor selection approach. Among the three spatial regression models tested (see Appendix II for other two models), the models goodness-of-fit measure (Table 4.6) indicated and suggested regression with spatially lagged dependent variable for enhanced performance in predicting and monitoring the five leading infectious diseases in the districts of Ghana. A negative coefficient of a determinant in the model pointed out that increasing the determinant will decrease the rate of infection and vice versa but positive coefficient implies direct variation with the rate. Hence the Spatial Lag model is suggesting that increasing the proportion of the population attaining at least Basic education and then decreasing urbanisation level, intermigration and number of males per 100 females will decrease the rate of infection. This brings into line the findings in most previous and reviewed studies such as Saker *et al* (2004), Osei and Duker, (2008), Osei *et al* (2012), Bradshaw (2013) and many more. Thus in addition to the earlier explanation, these determinants, especially urbanisation due to migration and overpopulation promote conditions that enhance the emergence and rapid spread of a range of infectious diseases. Buttressed by the

mappings, urbanisation and migration are overt indication of gross inequity in the distribution of social and economic opportunities with the associated outcomes of which increasing rate of infection is paramount. Efforts to reduce ruralurban migration by fair distribution of socio-economic prospects and reduction in overpopulation will promote health. Even-though the spatial models indicated stationarity and non-missing of key predictors over the study area, GWR modeling located nonstationarity and possible missing of key variables in very few areas. This suggests that inclusion of few other explanatory ecological factors that might be contributing significantly to the incidence of infection and/or data quality will further improve upon the performance of the spatial predictive models. It could also be suggested as in other recent researches that it is of a grave necessity and benefits to take into account spatial patterns in the analysis of health and social science data to make them more informing for enhanced decision making (Ward and Gleditsch, 2007; Marek *et al*, 2015).
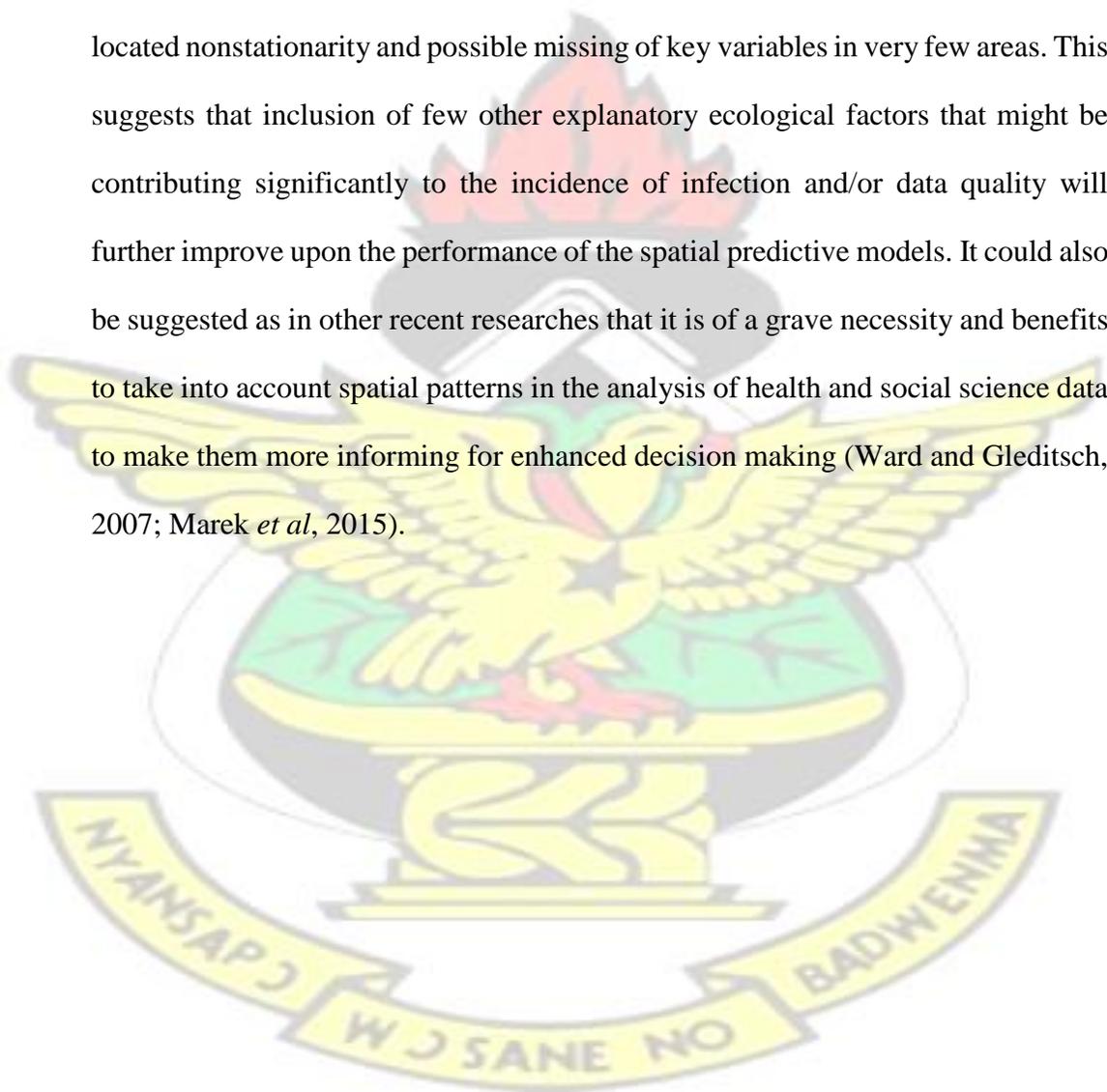
CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 SUMMARY OF FINDINGS

This study looked at using GIS-based spatial epidemiology to investigate the spatial dependency in the top five infectious diseases in Ghana. The geovisual exploratory data analysis through mapping of the smoothed rates provided a clear picture of a non-random distribution of the infections and their spatial dependency. Upon critical examination of the spatial pattern of infection with Moran's I, high significant clustering of the infections was confirmed for all diseases except URTI with insignificant scattering. Even-though increasing rates was in the order Typhoid fever, Intestinal worms, Diarrhoeal diseases, URTI and Malaria, increasing intensity of clustering of the diseases was URTI, Diarrhoeal diseases , Typhoid fever, Malaria and Intestinal worms (Table 4.1).Within the study periods, infection had moved from insignificant dispersion in 2010 to the greatest and significant clustering in 2014 (Table 4.2). Cluster-outlier analyses specifically identified a number of absolute locations with the significantly elevated or highly clustered incidence rates (hotspots) with respect to specific disease and the annual rates of infection. It was found out that 4, 8,8, 9 and 12 hotspot districts were identified for

Malaria, URTI and Diarrhoeal diseases, Typhoid fever and Intestinal worms respectively (Fig 4.13 to Fig 4.17 and Table 4.3). It was thus disclosed that albeit Malaria was the most leading cause of morbidity, Intestinal worms infection was found to be the most intensified clustered and with the highest number of hotspots. With respect to the specific periods of infection, the number of hotspot districts for the periods of 2010 to 2014 were 10,4,2,9 and 7 with specific identified hotspot locations showed in Fig 4.19 to Fig 4.23 and Table 4.4

Assessment of the thirty three (33) potential determinants indicated twenty (20) to be significantly (and with inter-correlation) contributing to the incidence of infection and with obvious geospatial variability across the country. A further probing of the significant prognostic factors unveiled that due to their inter-relationship, a maximum of four (4) of the risk factors viz Basic education, Urbanisation, Intermigration and Sex ratio could be used to significantly explain the impact of the other risk factors in the aetiology of the investigated infectious diseases. In the spatial modeling, Spatial Lag model emerged as the best spatial model among the three models in assessing and predicting how the four candidate determinants and incidence of infection share spatial relationships in the prognosis of the top five leading infectious diseases in Ghana. Although significant spatial dependency in the infection was found to exist in the entire study area, locally in very few areas, GWR modeling is proposing the possibility of missing of few key variables and non-stationarity of predictors whereby the model was over-predicting (red colour) and under-predicting (blue colour), Fig 4.40.

5.2 CONCLUSION

A GIS-based epidemiological approach was used to evaluate the spatial patterns, identify hotspot areas of infection and to assess the significance and spatial variability of determinants to the incidence of the top five infectious diseases in Ghana. After GIS database creation for the determinants and incidence of infection, spatial exploratory and spatial pattern analysis revealed spatial clustering in the incidence of infection. The observed spatial dependency could largely be due to social interaction through connected-neighbourhood effects on disease progression as reviewed in the previous studies. In terms of significant intensity of high clustering of the top five infectious diseases for the five periods; Intestinal worms, Malaria, Typhoid fever, Diarrhoeal diseases and URTI order was discovered (Table 4.1). This indicates the disease-specific prioritisation to be made to minimise the high clustering of the infections at a particular location to at least the levels in the surrounding districts while taking into account the disease status at the neighbouring areas as well. Table 4.3 showed the disease-specific hotspots districts requiring extra health policy intervention programmes or healthcare prioritisation to contain the infections to relative extent as their surrounding districts. Despite the generally high rates of infection across the country, most of the diseasespecific hotspot districts were located in Brong-Ahafo, Upper East and Western regions. Even among the five leading infectious diseases, it could be inferred that there is generally broader and quite an even health intervention programmes in addressing Malaria than the other infectious diseases, especially Intestinal worms was not receiving much and homogeneous attention. As many as twenty controllable health determinants emerged out of the 33 as significant risk factors with inter-relationship

and had proven to show geospatial variability in influencing the established geospatial incidence of infection. However, four of the significant determinants could be used to account for the prognostic effects of the other determinants on infection. Hence using the Spatial Lag model, it could be concluded from this spatial epidemiological study that modifying these significant determinants in such a way that the proportion of the population attaining at least Basic education is increased and decreasing the urbanisation level, intermigration and number of males per 100 females (sex ratio, relating to reproduction) will help decrease the rate of infection. The GWR modeling, however, indicated extremely low misspecification and non-stationarity of the predictors at very few locations where very few more key determinants should be included. Comparative visual observation of the disease and determinant maps pinpointed and provided relative information on the extent of space-time variation at which the top five infectious diseases and determinants were occurring. Hence the spatial relationship in the infections and determinants across the country and shown in the spatial models as well. The study therefore indicated areas of social health strategic planning and healthcare management approaches that need to be taken to further reduce infection especially in the hotspot areas. Secondarily, this study reaffirms that spatial epidemiology as data mining technique particularly in public health informatics is very essential. This redefine health data and make it epidemiologically more informing and communicating relatively enough, clearer and easily understandable information to almost all stakeholders and the general populace to enhance health promotion and public health administration and community-participatory interventions.

5.3 RECOMMENDATIONS

5.3.1 Recommendations for Stakeholders

The following recommendations are proposed in line with the findings from this research. The results suggested that despite generally high rates of infection across the country, there are a number of districts carrying much heavier burden of infection as compared to their neighbouring districts. The government, health authorities and other health related organizations should therefore consider the identified hotspot districts (most of them in Brong-Ahafo, Upper East and Western regions) to further strengthen and prioritise a specific disease control measures in these districts. It is also recommended that besides Malaria, more and uniform attention across districts should also be given to the other four leading infectious diseases especially the Intestinal worm and Typhoid fever. It is also suggested that policy and/ or social intervention to reduce high accumulation of people in a given locality, reduction in intermigration and increasing at least basic education attainment will significantly contribute to reducing the rate of infection across the country. The strategies to reduce these factors will greatly reduce a number of other significant factors including overcrowding and insanitary practices. Geo-health informatics components should be routinely included in health reports/researches for extensive spatio-temporal epidemiological assessment which will be of great benefits if basic logistics and personnel are provided. This provides health authorities with more easily understandable platform to identify focal areas at greater risks of infection (relying on annual incidence rates alone cannot clearly identify hotspots) to optimise public health prioritisation and specific interventions and optimal allocation of resources.

5.3.2 Recommendations for Further Studies

The GWR modeling indicated that few more potential risk factors, possibly hydrological, geological and climatic factors should be included in further researches. Inclusion of more risk factors, using remote sensing technologies to add more quality data and exploring more spatial modeling approaches could improve upon the capacities of spatial epidemiology for enhanced monitoring and surveillance system in the country. If the necessary addresses and smaller boundaries such as for sub-districts or smaller zones become available, future researches should consider conducting the GIS-based spatial epidemiology at individual, small groups or village levels to enhance decision making and disease surveillance at the local levels.

REFERENCES

Abelti, G., Brazzoduro, M. and Gebremedhin, B. (2001): Housing Conditions and Demand for Housing in Urban Ethiopia. In-depth studies from the 1994 Population and Housing Census in Ethiopia. Italian Multi-Bi Research Project Eth/92/P01.Central

Statistical Authority (CSA), Addis Ababa, Ethiopia, Institute for Population Research, National Research Council (Irp-Cnr), Roma, Italy

Adams, I., Darko, D. and Accorsi, S., (2004):The Burden of Diseases: Brief Overview of Disease Profile, Service Utilisation Patterns and Health Outcomes. *Bulletin of Health Information,* **1:**1

Adetokubo, O. L. and Gilles, H. M, (2003):Short Textbook of Public Health Medicine for the Tropics, 4th Edition. Published By Arnold, Hodder Headline Group, 338 Euston Road, London NW1 3BH

Alegana,V.A., Atkinson, P.M., Wright, J.A., Kamwi, R., Uusiku, P.,Katokele, S., Snow, R.W., and Noor, A.M.,(2013): Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial– temporal models. *Spatial and Spatio-temporal Epidemiology*, **7**. Pp 25–36 . Published by Elsevier Ltd. Available at http://dx.doi.org/10.1016/j.sste.2013.09.001

Almeida, A.S., and Werneck, G. L. (2014): Prediction of high-risk areas for visceral leishmaniasis using socioeconomic indicators and remote sensing data. *International Journal of Health Geographics*, **13**:13. Available at http://www.ij-

healthgeographics.com/content/13/1/13

Amo-Adjei, J. (2013): Views of health service providers on obstacles to tuberculosis

control in Ghana. Infectious Diseases of Poverty:**2**:9

http://www.idpjournal.com/content/2/1/9

Anamzui-Ya, J. A. (2012):*Spatial Analysis and Mapping of Cholera Causing Factors in

Kumasi, Ghana*. MSc. thesis submitted to the Faculty of Geo-information Science and

Earth Observation of the University of Twente. ITC, 7500 AA Enschede. The Netherlands

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*.

27, pages 93-115.

Anselin, L., Lozano, N., and Koschinsky, J. (2006): *Rate Transformations and

Smoothing*. Spatial Analysis Laboratory, Department of Geography, University of Illinois,

Urbana-Champaign Urbana, IL 61801

Anselin, L. (2005): Exploring Spatial Data with GeoDaTM : A Workbook. Center for

Spatially Integrated Social Science Spatial Analysis Laboratory Department of Geography,

University of Illinois, Urbana-Champ Urbana, IL 61801

Appiah, S.K., Mueller, U and Cross, J. (2011): Spatio-temporal modeling of malaria

incidence for evaluation of public health policy interventions in Ghana, West Africa. *19th

International Congress on Modeling and Simulation*, Perth, Australia.

http://mssanz.org.au/modsim2011

Arku, R.E., Bennett, J.E., Castro, M.C., Agyeman-Duah, K., Mintah, S.E., Ware, J.H., Nyarko, P., Spengler, J.D., Agyei-Mensah, S., and Ezzati, M. (2015). Spatial inequalities and social and environmental determinants of under-five mortality in Ghana in 2000 and 2010: Bayesian spatial analysis of census data In Poverty, Energy Use, Air Pollution And Health In Ghana: A Spatial Analysis. A degree of Doctor of Science dissertation submitted to the Department Environmental Health, The Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts.

Arie, C.,Guo,D.,Wang, F.,Wong, D.,Agouris, P.and Stefanidis, A. (2010). Spatial Data Analysis and Geoinformation Extraction. *Advanced Geoinformation Science*.

Atkinson, S.F., Sarkar, S.,Aviña, A., Schuermann, J.A, Williamson, P. (2014): A determination of the spatial concordance between Lyme disease incidence and habitat probability of its primary vector *Ixodes scapularis* (black-legged tick). *Geospatial Health* **9**(1), pages 203-212

Basommi,L. P (2011):*Spatial Analysis of Malaria Epidemiology in the Amanse West District*. MSc. Thesis submitted to the Department of Geomatic Engineering, College of Engineering. Kwame Nkrumah University of Science and Technology.

Bivand, R. (2014): Geographically Weighted Regression. Extract from Bivand, R. S., Pebesma, E. and Gómez-Rubio, V. (2013): Applied Spatial Data Analysis with R, Second edition. Springer-Verlag, New York..

Bradshaw, A. G. (2013): A Spatial Analysis of West Nile Virus in Texas, 2012. Published Master's Theses, Paper 464, University of Connecticut Graduate School http://digitalcommons.uconn.edu/gs_theses/464

Bonnefoy, X. (2007): Inadequate housing and health: an overview', *Int. J. Environment and Pollution*, Vol. 30, Nos. 3/4, pp.411–429.

Bonita, R., Beaglehole,R., Kjellström, T. (2006): *Basic epidemiology*, 2nd Edition. WHO Library Cataloguing-in-Publication Data, Printed in China
Boulos, M. (2004): Web GIS in practice: an interactive geographical interface to English Primary Care Trust performance ratings for 2003 and 2004. *International Journal of Health Geographics*, 28:1

Bouwmeester, H., Abele, S., Manyong, V.M, Legg, C., Mwangi, M., Nakato, V., Coyne, D., and Sonder, K., (2010), (Eds.: Dubois, T. et al); *The Potential Benefits of GIS Techniques in Disease and Pest Control: an Example Based on a Regional Project in Central Africa*. Proc. IC on Banana & Plantain in Africa. Acta Hort. 879

Buckeridge, D.L, Mason, R., Robertson, A., Frank,J., Glazier, R., Purdon,L., Amrhein, C.G., Chaudhuri, N., Fuller-Thomson, E., Peter Gozdyra, P., Hulchanski, D., Moldofsky, B., Thompson, M., &Wright, R. (2002): Making health data maps: a case study of a community/university research collaboration. *Social Science & Medicine* 55: 1189–1206.

Available at www.elsevier.com/location/socscimed

Bhutta, Z. A, Sommerfeld, J., Lassi, Z. S., Salam, R. A., and Das, J. K.(2014): Global burden, distribution, and interventions for infectious diseases of poverty. *Infectious Diseases of Poverty,* **3**:21. Doi:10.1186/2049-9957-3-21

Camara, G., Monteiro, A. M., Fucks, S. D. and Carvaho, M. S. (n.d): Spatial Analysis and GIS: A Primer. Image Processing Division, National Institute for Space Research,

Brazilian Agricultural Research Agency, National School for Public Health , Fundacao Oswaldo Cruz, Brazil. Pages 3-4, 7-8, 16

Cheong, Y.L., Leitão, P. J. and Lakes, T. (2014): Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees. *Spatial and Spatio-temporal Epidemiology*, 10. pages 75–84. Available at http://dx.doi.org/10.1016/j.sste.2014.05.002
Cianci, D., Hartemink, N., and Ibáñez-Justicia, A. (2015): Modeling the potential spatial distribution of mosquito species using three different techniques. *International Journal of Health Geographics*,**14**:10. DOI 10.1186/s12942-015-0001-0

Colston, J. and Saboyá, M. (2013): Soil-transmitted helminthiasis in Latin America and the Caribbean: modeling the determinants, prevalence, population at risk and costs of control at sub-national level. *Geospatial Health 7(2),* pages. 321-340

Congdon, P. (2015): Spatial variation in attributable risks. *Spatial and Spatio-temporal Epidemiology* **12**. Pp 39–52. Available at http://dx.doi.org/10.1016/j.sste.2015.02.002
DeGroote, J.P., Sugumaran, R. (2012): National and regional associations between human West Nile virus incidence and demographic, landscape, and land use conditions in the coterminous United States. *Vector Borne Zoonotic Dis* **12**, pages 657-665. DeGroote, J.P., Sugumaran, R and Ecker, M. (2014): Landscape, demographic and climatic associations with human West Nile virus occurrence regionally in 2012 in the United States of America. *Geospatial Health* **9**(1*),* pp. 153-168

Dhama, K., Verma, A. K., Tiwari, R., Chakraborty, S., Vora, K., Kapoor, S., Deb, R., Karthik K., Singh, R., Munir, M. and Natesan, S. (2013): A Perspective on Applications of

Geographical Information System (GIS): an Advanced Tracking Tool for Disease Surveillance and Monitoring in Veterinary Epidemiology. Review *Article. Advances in Animal and Veterinary Sciences*. 1 (1): 14 – 24. Available at http://www.nexusacademicpublishers.com/journal/4

Ehlkes, L., Krefis, A.C., Kreuels, B., Krumkamp, R., Adjei, O., Ayim-Akonor, M., Kobbe, R., Hahn, A., Vinnemeier, C., Loag,W., Schickhoff, U., and May, J. (2014): Geographically weighted regression of land cover determinants of Plasmodium falciparum transmission in the Ashanti Region of Ghana. *International Journal of Health Geographics*, **13**:35. Available at http://www.ij-healthgeographics.com/content/13/1/35

Ekpo, U.F, Hürlimann, E., Schur, N., Oluwole, A.S., Abe, E.M., Mafe, M.A., Nebe, O.J, Isiyaku, S., Olamiju, F., Kadiri, M., Poopola, T.O.S, Eka I. Braide, E.I.,Saka,Y., Mafiana, C.F.,Kristensen,T.K., Utzinger, J., and Vounatsou, P. (2013): Mapping and prediction of schistosomiasis in Nigeria using compiled survey data and Bayesian geospatial modeling. *Geospatial Health* **7**(2), pages 355-366

Eisen, L. and Eisen, R. J. (2011): Using Geographic Information Systems and Decision SupportSystems for the Prediction,Prevention, and Controlof Vector-Borne Diseases. *Annual Review of Entomology,* 56:41–61. Doi:10.1146/annurev-ento-120709-144847

Gazzinelli, A., Correa-Oliveira, R., Yang, G.J., Boatin, B.A, and Kloos, H. (2012): A Research Agenda for Helminth Diseases of Humans: Social Ecology, Environmental Determinants, and Health Systems. *PLoS Negl Trop Dis* 6(4): e1603. Doi:10.1371/journal.pntd.0001603

Gesink, L. D. C., Bernstein, K.T., Serre, M.L., Schumacher, C. M., Leone, P. A., et al (2006): Modeling a syphilis outbreak through space and time using the Bayesian maximum entropy approach. *Ann Epidemiol*. 2006;16:797-804

Getis, A. (2004).The role of geographic information science in applied geography. *The GeoJournal Library.*

Global Burden of Diseases (2014): Global, regional, and national age–sex specifi c allcause and cause-specifi c mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* ; 384: 980–1004 .

http://dx.doi.org/10.1016/S0140-6736(14)61682-2
Ghana Statistical Service (2013): 2010 Population & Housing Census. Regional Analytical Reports

Ghana Statistical Service (2012): 2010 Population & Housing Census. Summary Report of Final Results

Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF Macro. (2009): *Ghana Demographic and Health Survey 2008*. Accra, Ghana: GSS, GHS, and ICF Macro.

Gudes, O., Kendall, E., Yigitcanlar, T., Pathak V., and Baum, S. (n.d): Rethinking health planning: a framework for organising information to underpin collaborative health planning

Hansen K., VanOsdel N. (n.d): *GIS Applications in Health: An Introduction to GIS.* University of Nebraska Medical Center, College of Public Health. Northern Plains Comprehensive Cancer Control Program (NPCCCP) Northern Plains Tribal Cancer Data

Improvement Initiative (NPTCDI). Aberdeen Area Tribal Chairmen's Health Board. Available at http://www.aatchb.org/epi/ccc.htm

Hay, S.I., Guerra, C. A., Gething, P. W., Patil A. P., Tatem, A. J, Noor, A. M., et al. (2009): Correction: A World Malaria Map: Plasmodium falciparum Endemicity in 2007. *PLoS Med* 6(10): doi:10.1371/annotation/a7ab5bb8-c3bb-4f01-aa34- 65cc53af065d

Health Metrics Network (2005): Ghana health information system: Profile and assessment.

Hönig,V., Švec, P., Masař, O., Grubhoffer, L. (2011):Tick-Borne Diseases Risk Model for South Bohemia (Czech Republic). GIS *Ostrava***1**, pp 23. – 26.

Jebara, K. B. (2007): *The Role Of Geographic Information System (GIS) In The Control And Prevention Of Animal Diseases.* Head, Animal Health Information Department, OIE, World Organisation for Animal Health Conf. OIE 2007, 175-183

Jeefoo, P., Tripathi, N. K., Souris, M. (2011): Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao province, Thailand. *Int J Environ Res Public Health*;8(1):51-74. Doi: 10.3390/ijerph8010051.

Jeffery, C., Ozonoff, A., and Pagano, M., (2014):The effect of spatial aggregation on performance when mapping a risk of disease. *International Journal of Health Geographics***13**:9. http://www.ij-healthgeographics.com/content/13/1/9

Jenkins, C. D. (2003): *Building better health: a handbook for behavioral change* PAHO, Scientific and Technical Publication, Washington D.C.

Jewell, N. P (2009): *Statistics for epidemiology*. Texts in statistical science series; 58. Published in the Taylor & Francis e-Library, USA.

Johnson, C. P. and Johnson, J (2001): GIS: A Tool for Monitoring and Management of Epidemics. *Map India 2001 Conference*, New Delhi.

Kaushal, A., Johnson, C. P., (2003): Disease Surveillance Using GIS and Remote Sensing. Indo-French Workshop on. *Tele-Epidemiology of Dengue.* Pune

Khalid, A. and Al-Zahrani. A. (2013).Spatial Autocorrelation of Cancer Incidence in Saudi Arabia. *International Journal of Environmental Research and Public Health*. Kolivras KN. (2006): Mosquito habitat and dengue risk potential in Hawaii: a conceptual framework and GIS application. *Professional Geographer*, 58(2): 139-154

Krivoruchko, K., Gotway,C. A., and Zhigimont, A. (2003): Statistical Tools for Regional Data Analysis Using GIS. New Orleans, Louisiana, USA. ACM 1-58113-730-3/03/0011 Kwong, J.C, Crowcroft, N.S., Campitelli, M. A., Ratnasingham, S., Daneman, N., Deeks, S. L. and Manuel, D. G. (2010): Ontario Burden of Infectious Disease Study Advisory Group; *Ontario Burden of Infectious Disease Study* (ONBOIDS): An OAHPP/ICES Report. Toronto: Ontario Agency for Health Protection and Promotion, Institute for Clinical Evaluative Sciences. P13.

Law, D.C.G., Serre, M,L., Christakos,G., Leone, P.A. and Miller, W.C. (2004):Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies. *Sex Transm Infect*. 2004;80:294-299.

Lumbala, C., Simarro, P.P., Cecchi, G., Paone, M., Franco, J.R., Mesu, Makabuza, J., Diarra, A., Chansy, S., Priotto, G., Mattioli, R. C.and Jannin, J.G. (2015), Human African trypanosomiasis in the Democratic Republic of the Congo: disease distribution and risk *International Journal of Health Geographics.*Doi:10.1186/s12942-015-0013-9

MacMillan, K., Monaghan, A.J., Apangu, T., Griffith, K.S., Mead, P.S., Acayo, S., Acidri, R., Moore, S.M., Mpanga, J.T., Enscore, R. E., Gage, K. L., and Eisen, R. J.

(2012).Climate Predictors of the Spatial Distribution of Human Plague Cases in the West Nile Region of Uganda**. *Am J Trop Med Hyg*. Mar 1; 86(3): 514–523. doi: 10.4269/ajtmh.2012.11-0569

Marek, L., Tuček, P., and Vít Pászto, V. (2015): Using geovisual analytics in Google

Earth to understand disease distribution: a case study of campylobacteriosis in the Czech Republic (2008–2012). *International Journal of Health Geographics*, **14**:7, http://www.ij-healthgeographics.com/content/14/1/7

Masters, S.H.,Burstein, R., Amofah, G., Abaogye, P., Kumar, S. and Hanlon, M. (2013):Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis. *Environ Monit Assess*.185(4):3561-79. Doi: 10.1007/s10661-012-

2810-y

McCann, R.S., Messina, J.P.,MacFarlane, D.W., Bayoh, M.N., Vulule,J.M., Gimnig, J.E and Walker, E.D (2014): Modeling larval malaria vector habitat locations using landscape features and cumulative precipitation measures. *International Journal of Health Geographics*, **13**:17. http://www.ij- healthgeographics.com/content/13/1/17

Meza J. L. (2003): Empirical Bayes estimation smoothing of relative risks in disease mapping. Journal of Statistical Planning and Inference 112 (2003) 43 – 62. www.elsevier.com/locate/jspi

Mika, J.P.R. (2004): Not All Maps Are Equal: GIS and Spatial Analysis In

Epidemiology. National Public Health Institute, Department of Epidemiology and Health Promotion, Diabetes and Genetic Epidemiology Unit, Helsinki, Finland. *International Journal of Circumpolar Health* **63***:1*

Mitchell, A. (2015). The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics (Kindle Locations 985-986). Esri Press. Kindle Edition.

Molla, Y.B., Wardrop, N.A., Blond, J.S.L., Baxter,P., Newport, M.J., Atkinson, P.M., and Gail Davey, G. (2014): Modeling environmental factors correlated with podoconiosis: a geospatial study of non-filarialelephantiasis. *International Journal of Health Geographics*, **13**:24. http://www.ij- healthgeographics.com/content/13/1/24

Munyaneza, F., Hirschhorn, L.R., Amoroso,C.L., Nyirazinyoye, L., Birru, E., Mugunga, J.C., Murekatete, R.M., and Ntaganira, J. (2014): Leveraging community health worker system tomap a mountainous rural district in low resourcesetting: a low-cost approach to expand use of geographic information systems for public health. *International Journal of Health Geographics*. **13**:49. http://www.ij-healthgeographics.com/content/13/1/49 Murack, J. (2013):Spatial Autocorrelation using GIS. Spatial autocorrelation

presentation.

 Murack, J. (n.d): Regression Analysis Using ArcGIS

Murad, A. A. (2014): Using geographical information systems for defining the accessibility to health care facilities in Jeddah City, Saudi Arabia. *Geospatial Health*

8(3), pp. S661-S669

National Population Council (2011): Ghana Population Stabilisation Report

Norstrøm, M. (2001): Geographical Information System (GIS) as a Tool in Surveillance and Monitoring of Animal Diseases. *Acta vet. scand*. *Suppl. 94:* pages 79-85.

Odoi, A., Martin, S.W., Michel, P., Holt, J., Middleton, D. and Wilson, J. (2003): Geographical and temporal distribution of human giardiasis in Ontario, Canada. *International Journal of Health Geographics*, **2**:5. Available at http://www.ijhealthgeographics.com/content/2/1/5

Onwuemele, A. (2014): An assessment of the spatial pattern of malaria infection in Nigeria. International Journal of Medicine and Medical Sciences. Vol. 6(2), pages 80-86. Doi: 10.5897/IJMMS2013.1006

Ord, J.K. and Getis, A. (1995): Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis, Vol. 27, No.* **4 .** Ohio State University Press

Osei-Assibey, E and Grey, S.K. (2013): Millennium Development Goals in Ghana from2010 PHC report. Ghana Statistical Service

Osei, F. B. and Duker, A. A. (2008): Spatial and demographic patterns of Cholera in Ashanti region –Ghana. *International Journal of Health Geographics***7**:44. Doi:10.1186/1476-072X-7-44

Osei, F. B. (2010): *Spatial Statistics of Epidemic Data: The case of Cholera Epidemiology in Ghana*. PhD dissertation, number 177 submitted to the Faculty of Geoinformation Science and Earth Observation of the University of Twente. ITC, 7500 AA

Enschede. The Netherlands

Osei, F.B., Duker, A.A. and Stein, A. (2012): Evaluating Spatial and Space-Time

Clustering of Cholera in Ashanti-Region-Ghana, Cholera, Dr. Sivakumar Gowder (Ed.),

ISBN: 978-953-51-0415-5,InTech,Available from:

http://www.intechopen.com/books/cholera/evaluating-spatial-and-space-time-

clusteringofcholera-in-ashanti-region-ghana

Ostfeld, R.S, Glass, G. E. and Keesing, F. (2005): Spatial epidemiology: an emerging (or

re-emerging)discipline. Review. *TRENDS in Ecology and Evolution*.**20**:6

Doi:10.1016/j.tree.2005.03.009

Otto H, Rolf A, (Eds.) (2009): Principles of Geographic Information Systems: An

Introductory textbook, fourth Edition. International Institute for Geo-Information Science

and Earth Observation, Enschede, The Netherlands.

Potter T. (n.d): GIS to Assist in Early Detection of Infectious Diseases. *Advancing Public

Health Outcomes Through Information Technology*

Rai, P. K., Nathawat, M. S. and Mishra, A. (n.d): *Roleof GIS and GPS in Vector Born

Disease Mapping: A Case Study*. Department of Geography, Banaras Hindu University,
Varanasi, Department of Remote Sensing, Birla Institute of Technology, Mesra,

Ranchi.

Rajiv, G. and Dee, J. and Rajni J. (2003): Geographic Information Systems for the Study

and Control of Infectious diseases. Birla Institute of Technology & Science, Pilani.

Garauv Clinic, Ashram, New Delhi, Civil Engineering Group, BITS, Pilani (Raj.) 333 031,

India. Map India Conference

Rayed, C. A. (2012): Using GIS for Modeling a Spatial DSS for Industrial Pollution in Egypt. Computer and Information System dept. Sadat Academy for Management Science Cairo, Egypt   American Journal of Geographic Information System, 1(3): 33-38 DOI: 10.5923/j.ajgis.20120103.01

Saker L., Lee K., Cannito B., Gilmore A. and Campbell D. (2004): Globalisation and Infectious Diseases: A Review of the linkages. Special Topics No 3. WHO Special Programme for Research and Training in Tropical Diseases.

Santana, K. S, Bavia, M. E., Lima, A. D., Guimarães, I. C., Soares, E. S., Silva, M. M., Mendonça, J. and Martin, M. S. (2011): Spatial distribution of triatomines (Reduviidae: Triatominae) in urban areas of the city of Salvador, Bahia, Brazil**. *Geospat Health*,5(2):199-203.

Sankoh, O. A., Berke, O., Simboro, S. and Becher, H. (2002): Bayesian and GIS Mapping of Childhood Mortality in Rural Burkina Faso, *Control of Tropical Infectious Diseases*, Uni-Heidelberg

Sanson, R. L, Ster, M.W. and Morris, R.S. (1994): Interspread-A spatial stochastic simulation model of epidemic foot-and-mouth disease. The Kenyan Veterinarian. 18(2):

493-495.
Schimmer, B., Schegget, R., Wegdam, M., Züchner, L., Bruin, A., Schneeberger, P. M., Veenstra, T., and Vellema, P. and Hoek, W. V. D. (2010): The use of a geographic information system to identify a dairy goat farm as the most likely source of an urban Qfever outbreak. *BMC Infectious Diseases*, **10**:69. http://www.biomedcentral.com/1471-

2334/10/69

Schwanke Khilji, S.U, Rudge, J.W., Drake, T., Chavez, I., Borin, K., Touch, S. and Coker, R. (2013): Distribution of selected healthcare resources for influenza pandemic response in Cambodia. *Int J Equity Health.***4**;12:82. Doi: 10.1186/1475-9276- 12-82.

Scott, L.M and Janikas, M.V. (2010): Spatial Statistics in ArcGIS. In Fischer, M.M and Getis, A., (eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, pp 27- 41. Doi: 10.1007/978-3-642-03647-7_2, Springer-Verlag Berlin Heidelberg

Sen, K. and Bonita, R. (2000): Global health status: two steps forward, one step back. *Lancet* ;356: 577-582

Shaddick, G. (n.d): *Spatial epidemiology Four tutorial style BRG seminars*,Department of Mathematical Sciences, University of Bath, UK

Shittu,O.B., Akpan, Popoola, T.O.S., Oyedepo, J. A. and Ogunshola, E. O. (2010): Epidemiological features of a GIS-supported investigation of cholera outbreak in

Abeokuta, Nigeria Journal of Public Health and Epidemiology Vol. 2(7), pages. 152-162. Available online at http://www.academicjournals.org/jphe.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A, Ruiz-Postigo, J.A., Mattioli, R.C.,and Jannin, J.G. (2014): Mapping the capacities of fixed health facilities tocover people at risk of gambiense human Africantrypanosomiasis. *International Journal of Health Geographics*, **13**:4. Available at http://www.ij-

healthgeographics.com/content/13/1/4

Toprak, D. and. Erdoğan, S. (2008): Spatial Analysis of the Distribution of Typhoid Fever in Turkey.*The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol. XXXVII. Part B8. Beijing

Tsai, P.J. (2012): Application of Moran's Test with an Empirical Bayesian Rate to Leading Health Care Problems in Taiwan in a 7-Year Period (2002–2008). Global Journal of Health Science; Vol. 4, No. 5. Published by Canadian Center of Science and Education. Doi:10.5539/gjhs.v4n5p63

Tsai, P. J., Lin, M. L., Chu, C. M. and Perng, C. H. (2009): Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health, 9*, 464. http://dx.doi.org/10.1186/1471-2458-9-464

Tsegaw, T., Gadisa, E., Seid, A., Abera, A., Teshome, A., Mulugeta, A., Herrero, ., Argaw, D., Jorge, A. and Aseffa, A. (2013): Identification of environmental parameters and risk mapping of visceral leishmaniasis in Ethiopia by using geographical information systems and a statistical approach. *Geospatial Health 7(2),* pp. 299-308

Um, S., Kim, N.H, Lee, H. K., Song, J. S. and Kim, H. C. (2014): Spatial epidemiology of dry eye disease: findings from South Korea. *International Journal of Health Geographics.***13**:31. Available at: http://www.ij-healthgeographics.com/content/13/1/31

Usman, A. K. and Ahmed, M. (2013): Distribution of Primary Health Care Facilities in Kano Metropolis Using Geographic Information System**.** *Research Journal of Environmental and Earth Sciences***5**(4): 167

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (2012): *Principles of Epidemiology in Public Health Practice: An*

Introduction to Applied Epidemiology and Biostatistics, Third edition). Office of

Workforce and Career Development (OWCD), Career Development Division (CDD)Atlanta, Georgia 30333

Verma, S., and Gupta, R.D. (2014): Spatial and Temporal Variation of Japanese

Encephalitis Disease and Detection of Disease Hotspots: A Case Study of Gorakhpur

District, Uttar Pradesh, India. ISPRS Annals of the Photogrammetry, Remote Sensing and

Spatial Information Sciences, Volume II-8, doi:10.5194/isprsannals-II-8-1-2014 Verutes,

G.M., Fiocco, M.B., Weeks, J.R. and Coulter, L.L. (2012): Health, Poverty, and

Place in Accra, Ghana: Mapping Neighborhoods. *J Maps*. **8**(4):369-373.

Waller, L. A. (n.d): Spatial Epidemiology. Emory University, Atlanta, Georgia

Waller, L. A., Gotway C. A. (2004): Applied Spatial Statistics for Public Health Data.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in

Canada.

Waller, L.A (2007). Spatial Epidemiology, Statistical Advances in the Biomedical
Sciences.

Ward, M.D. and Gleditch, K. S (2007): *An Introduction to Spatial Regression Models in the Social Sciences.* Barcelona, Seattle, San Diego, Oslo, & Colchester.

World Health Organization and UNICEF (2013): *Progress on sanitation and drinkingwater*. WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27,

Switzerland.

World Health Organization (2013): *The economics of the social determinants of health and health inequalities: a resource book*. WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland.

World Health Organization and UNICEF (2010): Progress on Sanitation and Drinkingwater: WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27,

Switzerland

World Health Organization (2010a): Ghana Factsheets of Health Statistics

World Health Organization (2010b): Ghana Country Profile. Preventive Chemotherapy and Transmission Control, Department of Control of Neglected Tropical Diseases World

Health Organization, 20, avenue Appia, 1211 Geneva 27, Switzerland

World Health Organization (2009): Global health risks: mortality and burden of disease attributable to selected major risks

World Health Organisation (2001)**:**Infections and infectious diseases. A manual for nurses and midwives in the WHO European Region

Yrigoyen C. C (2007): Spatial autocorrelation tests. Course on Spatial Econometrics with Applications. Session3, Universidad Autónoma de Madrid Lugar: Universidad Politécnica de Barcelona.

APPENDICES

I. **Description of District locations**

Table1. The 170 Districts of Ghana as at 2010 PHC

District code	District Name	Region
1	Jomoro	Western
2	Ellembelle	Western
3	Nzema East Municipal	Western
4	Ahanta West	Western
5	Sekondi- Takoradi Metro	Western
6	Shama	Western
7	Mpohor-Wassa East	Western
8	Tarkwa Nsueam Municipal	Western
9	Prestea Huni Valley	Western
10	Wassa Amenfi East	Western
11	Wassa Amenfi West	Western
12	Aowin -Suaman	Western
13	Sefwi Akontombra	Western
14	Sefwi Wiawso	Western
15	Sefwi Bibiani-Anhwiawso Bekwai	Western
16	Juabeso	Western
17	Bia	Western
18	Komenda-Edina-Eguafo-Abirem (KEEA)	Central
19	Cape Coast Metro	Central
20	Abura-Asebu-Kwamankese	Central
21	Mfantsiman Municipal	Central
22	Ajumako-Enyan-Essiam	Central
23	Gomoa West	Central
24	Effutu Municipal	Central
25	Gomoa East	Central
26	Awutu Senya	Central
27	Agona East	Central
28	Agona West Municipal	Central
29	Asikuma-Odoben- Brakwa	Central
30	Assin South	Central

31	Assin North Municipal	Central
32	Upper Denkyira East Municipal	Central
33	Upper Denkyira West	Central
34	Twifo-Hemang-Lower Denkyira	Central
35	Ga South Municipal	Greater Accra

36	Ga West Municipal Greater Accra
37	Ga East Municipal Greater Accra
38	Accra Metropolis Greater Accra
39	Adenta Municipal Greater Accra
40	Ledzokuku/Krowor Municipal (LeKMA) Greater Accra
41	Ashaiman Municipal Greater Accra
42	Tema Municipal Greater Accra

43	Dangme West/Shai Osudoku Greater Accra	
44	Dangme East Greater Accra	
45	South Tongu Volta	
46	Keta Municipal Volta	
47	Ketu South Volta	
48	Ketu North Volta	
49	Akatsi Volta	
50	South Dayi Volta	
51	Hohoe Municipal Volta	
52	Biakoye Volta	
53	Jasikan Volta	
54	Krachi East Volta	
55	Krachi West Volta	
56	Nkwanta South Volta	
57	Nkwanta North Volta	
58	North Tongu Volta	
59	Adaklu Anyigbe Volta	
60	North Dayi Volta	
61	Birim South Eastern	
62	Birim Central Municipal Eastern	
63	West Akim Municipal Eastern	
64	Suhum-Kraboa Coaltar Eastern	
65	Akwapem North Eastern	
66	New Juaben Municipal Eastern	
67	Yilo Krobo Eastern	
68	Lower Manya Krobo Eastern	
69	Asuogyaman Eastern	
70	Upper Manya Krobo Eastern	

71	Fanteakwa Eastern
72	East Akim Municipal Eastern
73	Kwaebibirem	Eastern
74	Akyemmansa	Eastern
75	Birim North	Eastern
76	Atiwa	Eastern
77	Kwahu West Municipal	Eastern
78	Kwahu South	Eastern
79	Kwahu East	Eastern
80	Kwahu North	Eastern
81	Akwapem South	Eastern
82	Atwima Mponua	Ashanti
83	Amansie West	Ashanti

84 Amansie Central Ashanti 85 Adansi South Ashanti 86 Obuasi Municipal Ashanti 87 Adansi North Ashanti 88 Bekwai Municipal Ashanti 89 Bosome Freho Ashanti 90 Asante Akim South Ashanti 91 Ejisu Juaben Municipal Ashanti 92 Bosumtwe Ashanti 93 Atwima Kwanwoma Ashanti 94 Kumasi Metropolis Ashanti 95 Atwima Nwabiagya Ashanti 96 Ahafo Ano South Ashanti 97 Ahafo Ano North Ashanti 98 Offinso Municipal Ashanti 99 Afigya Kwabre Ashanti 100 Kwabere (Kwabere East) Ashanti 101 Afigya Sekyere (Sekyere South) Ashanti 102 Asante Mampong Muni Ashanti 103 Sekyere East Ashanti 104 Sekyere Afram Plains Ashanti 105 Sekyere Central Ashanti 106 Ejura-Sekyedumase Ashanti 107 Offinso North Ashanti 108 Asante Akim North Ashanti 109 Asunafo South Brong Ahafo 110 Asunafo North Municipal Brong Ahafo 111 Asutifi Brong Ahafo	

112   Dormaa Municipal Brong Ahafo
113   Dormaa East Brong Ahafo
114   Tano South   Brong Ahafo
115   Tano North   Brong Ahafo
116   Sunyani Municipal Brong Ahafo
117   Sunyani West Brong Ahafo 118 Berekum Municipal Brong Ahafo
119   Jaman South   Brong Ahafo
120   Jaman North   Brong Ahafo
121   Tain   Brong Ahafo
122   Wenchi Municipal Brong Ahafo
123   Techiman Municipal Brong Ahafo
124   Nkoranza South Brong Ahafo

125 Nkoranza North Brong Ahafo
126 Atebubu Amantin Brong Ahafo
127 Sene Brong Ahafo
128 Pru Brong Ahafo
129 Kintampo South Brong Ahafo
130 Kintampo North Municipal Brong Ahafo
131 Bole Northern
132 Sawla-Tuna-Kalba Northern
133 West Gonja Northern
134 Central Gonja Northern
135 East Gonja Northern
136 Kpandai Northern
137 Nanumba South Northern
138 Nanumba North Northern
139 Zabzugu Tatale Northern
140 Yendi Municipal Northern
141 Tamale Metropolis Northern
142 Kumbungu Northern
143 Savelugu Nanton Northern
144 Karaga Northern
145 Gushiegu Northern
146 Saboba Northern
147 Chereponi Northern
148 Bunkpurugu-Yunyoo Northern
149 East Mamprusi Northern
150 West Mamprusi   Northern
151 Builsa Upper East
152 Bolgatanga Municipality Upper East

153	Talensi Nabdam	Upper East
154	Bongo	Upper East
155	Bawku West	Upper East
156	Garu Tempane	Upper East
157	Bawku (East) Municipal	Upper East
158	Wa West	Upper West
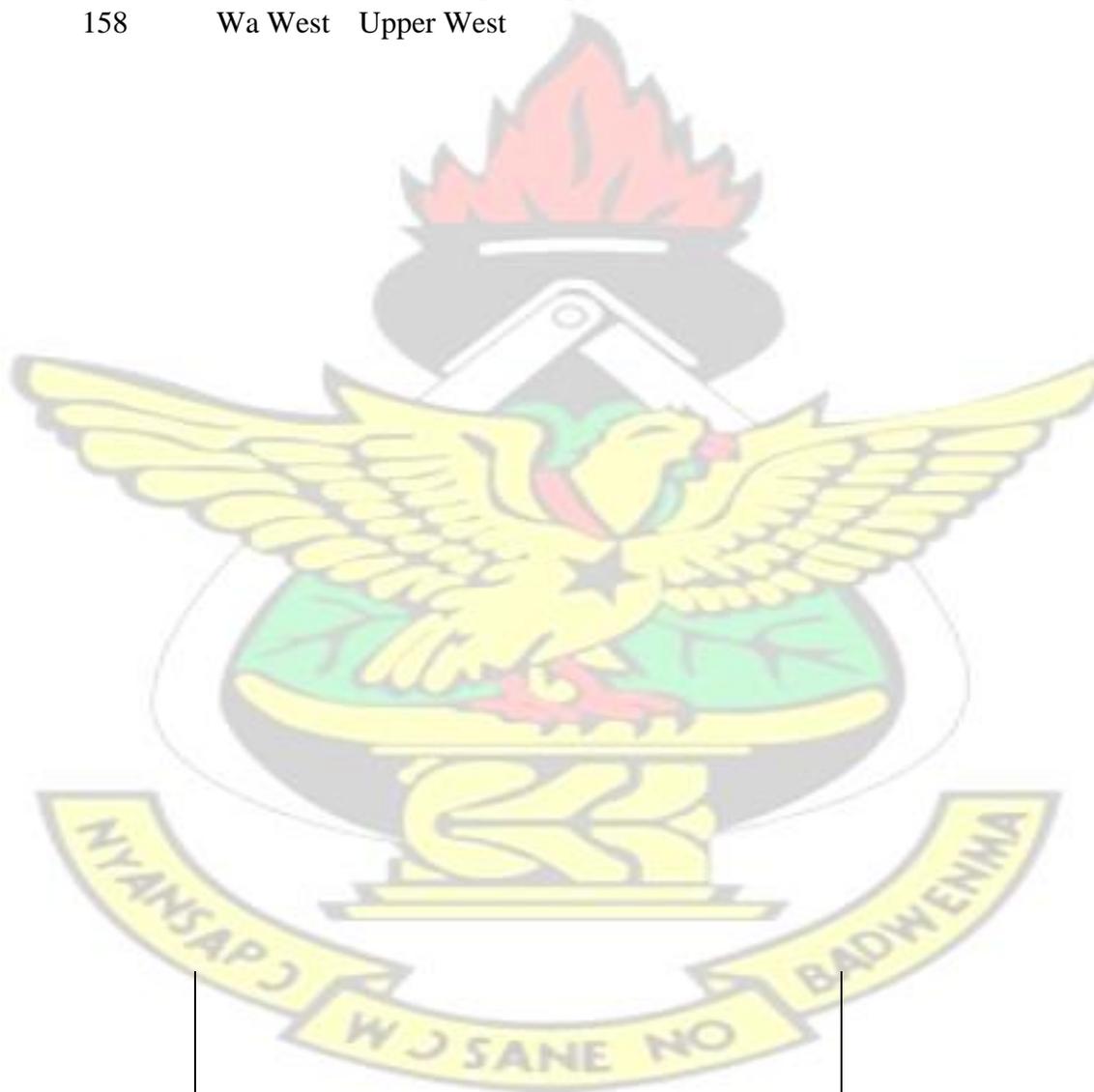
159	Wa Municipal	Upper West	
160	Wa East	Upper West	
161	Sissala East	Upper West	
162	Nadowli	Upper West	
163	Jirapa	Upper West	
164	Sissala West	Upper West	
165	Lambussie Karni		Upper West
166	Lawra		Upper West
167	Kasena Nankana East		Upper East
168	Kasena Nankana West		Upper East
169	Kadjebi		Volta
170	Ho Municipal		Volta

II. Modeling spatial relationship

a. Summary of model estimations

i. Ordinary Least Squares (OLS) model

Number of observation: 170   Degree of freedom: 165

Mean dependent variable : 18742.1   Number of parameter estimates: 5

S.D. dependent variable : 9445.01   S.E of regression : 8515.63

Table 2. Ordinary Least Squares (OLS) model parameter estimations

Variable	Coefficient	Std. Error	t-Statistic	P-value
Constant	48113.79	11235.93	4.2821	0.00003*
Sex ratio	-250.0656	115.3896	-2.1671	0.03166 *
Basic education	-25755.49	4827.879	-5.3347	0.00000*
Urbanisation level	13262.02	3013.719	4.4005	0.00002 *
Inter-migration	4678.116	1972.485	2.3717	0.01886 *

ii. Spatial Lag model

Number of observation: 170   Degree of freedom: 164

Number of parameter estimates: 6   Mean dependent variable : 18209.2

S.D. dependent variable : 5989.4   S.E of regression : 4157.8

Table 3. Spatial Lag model-maximum likelihood parameter estimations

Variable	Coefficient	Std. Error	z-value	P-value
Lag coeff. (Rho)	0.713833	0.0573	12.4540	0.00000*
Constant	3385.134	5536.44	0.6114	0.54092
Sex ratio	22.9913	56.4371	0.4074	0.68373
Basic education	-2391.13	2367.834	-1.0099	0.31257
Urbanisation level	384.0142	963.1658	0.3987	0.69011
Inter-migration	1407.215	1477.828	0.9522	0.34099

iii. Spatial Error model

112

Number of observation: 170   Degree of freedom: 165

Number of parameter estimates: 5 Mean dependent variable : 18742.052

S.D. dependent variable : 9445.0078   S.E of regression : 8285.79

Table 4. Spatial Error model-maximum likelihood parameter estimations

Variable	Coefficient	Std. Error	z-value	P-value
Lag coeff. (Lambda)	0.185317	0.1070157	1.7317	0.08333
Constant	48507.49	11891.22	4.0793	0.00005*
Sex ratio	-256.1356	122.8256	-2.0854	0.03704*
Basic education	-24712.86	5080.963	-4.8638	0.00000*
Inter-migration	3843.55	1901.16	2.021687	0.04321*
Urbanisation level	12848.4	3166.587	4.057491	0.00005*

b. Spatial Regression Model Diagnosis

i. OLS model diagnostics

Table 5 Diagnosis of OLS model

Model diagnostic measure	Value	P-value	Model assessment
R_2	0.211024	-	Performance
Adjusted-R^2	0.191897	0.0000*	Performance
Corrected Akaike info criterion, AICc	3564.25	-	Performance
Log likelihood	-1777.12	-	Performance

Moran's I (error)	2.1049	0.03530*	Spatial dependence
Jarque-Bera (df = 2)	107.7229	0.0000*	normality of errors (model bias)
Breusch-Pagan test (df = 4)	3.8145	0.43169	Heteroskedasticity (nonstationary of predictors)
Moran's I on residuals	0.0878	0.05604	Misspecification

ii. Spatial Lag model diagnostics

Table 6. Diagnosis of Spatial Lag model

Model diagnostic measure	Value	P-value	Model assessment
R^2 (pseudo)	0.518094		Performance
Adjusted - R^2			Performance
Corrected Akaike info criterion, AICc	3352.05		Performance
Log likelihood	-1670.03		Performance
Likelihood Ratio Test (df = 1)	83.481	0.00000*	Spatial dependence

Breusch-Pagan test (df = 4)	8.6989	0.06908	
			Heteroskedasticity (non-stationary of predictors)
Moran's I on residuals	-0.1009	0.0377*	Misspecification and bias

iii. Spatial Error model diagnostics

Table 7. Diagnosis of Spatial Error model

Model diagnostic measure	Value	P-value	Model assessment
R^2 (pseudo)	0.230403		Performance
Adjusted - R^2	—		Performance
Corrected Akaike info criterion, AICc	3561.28		Performance
Log likelihood	-1775.64		Performance
Likelihood Ratio Test (df = 1)	2.9676	0.08495	Spatial dependence
Breusch-Pagan test (df = 4)	4.7826	0.31034	Heteroskedasticity (non-stationary of predictors)
Moran's I on residuals	-0.0140	0.8695	Misspecification and model biasness

* Statistically significant at p-value < 0.05

-: Not applicable/Available

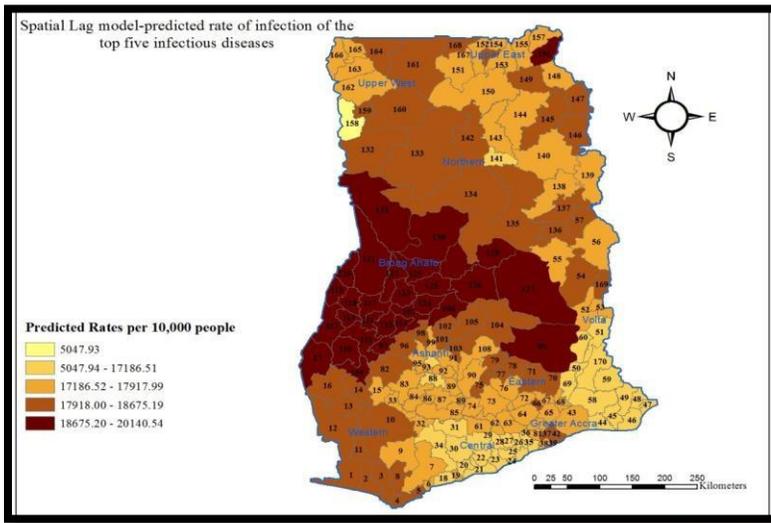**c. Predicted maps of the selected model (Spatial Lag model)**

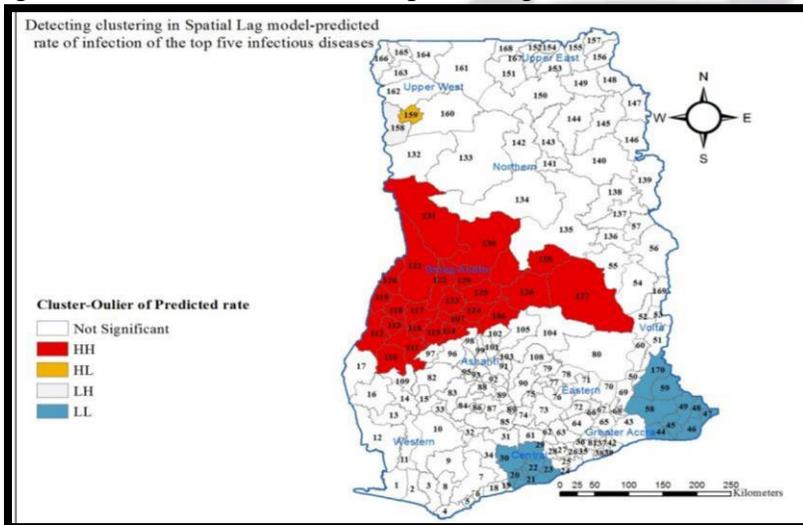Fig 1 Predicted incidence from Spatial Lag model

Fig 2: Cluster-Outlier detection in Spatial Lag model-predicted incidence

d. Spatial Regression Modeling Equations

i. *Ordinary Least Squares regression*

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

Y = dependent variable (to be predicted or understood) X = value of the explanatory/ independent

variables (predictors) β = coefficients computed by the regression tool (representing the strength

and type of relationship between X and Y   ε = random error term/ residuals n = number of exploratory variables ii. *Spatial Lag Model*

$$y = (\rho)W_y + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

Y = dependent variable

Wy = spatially lagged dependent variable for connectivity weights matrix W_{ij}
W_{ij}= the element in the spatial weight matrix corresponding to the observation pair of districts i,

j to define the spatial interaction among the districts in the study area by wi·yi with the connectivity

vector wi· The W connectivity matrices are row-standardized such that each row sums to 1 (Σwij

= 1) if districts i and j share a common boundary; otherwise wij = 0, for non-neighbouring districts.

X = matrix of observations on the explanatory variable ε = vector of error terms

ρ(rho) and β(beta) are estimated parameters

NB: Spatial lag variable averages the neighboring values of a location y is

dependent on its neighbors through the weights matrix Weights matrix

accounts for the autocorrelation in the model *iii. Spatial Error Model*

This also include spatial autoregressive error term to accounts for autocorrelation in the error

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + (W)\varepsilon + \mu$$

Y = dependent variable

X = matrix of observations on the explanatory variables W= spatial

weights matrix ε = is a vector of spatially autocorrelated error terms u = is

a vector of i.i.d. (independent identically distributed) errors

(lambda) and β are estimated parameters

KNUST