SPATIAL ANALYSIS AND MAPPING OF CHOLERA CAUSING FACTORS IN KUMASI, GHANA.

JERRY ASAANA ANAMZUI-YA
March, 2012

SUPERVISORS:

First Supervisor: Prof. Dr. Ir. A. Stein
Second Supervisor: Dr. N. A. S. Hamm

SPATIAL ANALYSIS AND MAPPING OF CHOLERA CAUSING FACTORS IN KUMASI, GHANA.

JERRY ASAANA ANAMZUI-YA
Enschede, The Netherlands, March, 2012

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:

First Supervisor: Prof. Dr. Ir. A. Stein
Second Supervisor: Dr. N. A. S. Hamm

THESIS ASSESSMENT BOARD:

Dr. Ir. R. A. de By (chair)
Dr. A. Dilo (External Examiner)

ABSTRACT

Recent Global health reports show a continual vulnerability of large populations to infectious diseases such as cholera in relation to our environments. Of particular concern is the spatial distribution of cholera incidences and its associated environmental risks factors. This can be used by health officials and policy makers to make appropriate planning and resource allocation. Despite the availability of remotely sensed data in various formats for mapping, few studies have utilized the technology for mapping the environmental niche of *V. cholerae*. A key environmental factor which predisposes persons to cholera infection is sanitation. Two identified important measures of sanitation in an urban city, Kumasi are proximity to refuse dumps and water reservoirs within a community. To this end a RapidEye image is exploited to map the potential cholera reservoirs and compare with digitized reservoirs from a topographic map in spatial analysis. A spatial conditional autoregressive (CAR) modeling was carried out to determine the spatial dependency of cholera prevalence on (1) proximity to refuse dumps and digitized reservoirs and (2) proximity to refuse dumps and classified reservoirs from RapidEye image. The results showed that there is an inverse spatial relationship between cholera prevalence and proximity to both refuse dumps(p-Value < 0.0001) and classified reservoirs(p-Value < 0.001). The model with digitized reservoirs was found to be insignificant at $\alpha = 0.05$ (i.e, p-Value $= 0.07$). The results of the spatial models suggests a better fit with the classified reservoirs. Moran's I analysis showed a significant spatial association of cholera risk with neighboring communities. Probability and risks maps were also generated to characterize the spatial patterns of cholera prevalence in the Kumasi Metropolis.

Keywords

disease mapping, spatial statistics, remote sensing, cholera, conditional autoregressive modeling, environmental factors, GIS

ACKNOWLEDGEMENTS

I thank The Good Lord for His goodness, abundant grace and wonderful works in my life.

I acknowledge the Netherlands organization for international cooperation in higher education (nuffic) for providing the funds for my study. Thanks to my employers Bolgatanga Polytechnic for granting me study leave.

I am grateful to my supervisors Professor Alfred Stein and Dr. Nicholas Hamn for their constructive thoughts, critisms and remarks during my research work. My sincere thanks to Dr. Frank B. Osei for his advice and allowing me use his data.

Special thanks to my family for their support, prayers and encouragements; Obviously, i would not have come this far without them. Thank you so much Dad(who unfortunately passed away during my study). You have been a pillar and an inspiration in every aspect of my life. May the good Lord lay you to rest till we meet again. Thank you mum, uncle Eddy, Alhaji Osman, Rhoda, Ruth, Judith, Frank, Jonas and Joshua; I love you all.

Finally, i salute all my friends here at ITC, especially my classmates GFM 2 2010-12, and all the Ghanaian students who in one way or the other have touched my life, God bless you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 MOTIVATION AND PROBLEM STATEMENT

Recent Global health reports show a continual vulnerability of large populations to infectious diseases in relation to our environments[54, 80]. Governments and international establishments have extensively recognized the need to improve the well-being of its populations as well as to undertake prompt preventive and control measures. Infectious diseases are complex to control and prevent, leading to questions on how best to combat them through novel and creative solutions[35, 76]. Cholera is an epidemic and infectious disease which is of global and public health significance, hence the need to recognize and address it accordingly[81].

Cholera is an acute intestinal infection caused by a bacterium (*V. cholerae*) leading to intestinal infection and diarrhea. Infection is acquired chiefly by intake of contaminated water or food[16, 38]. Transmission is due to the fecal contamination of food and water as a result of poor sanitation. The bacterium can live naturally in any environment[52]. Thus, it remains a global threat especially in countries where access to clean safe drinking water and sufficient sanitation cannot be assured. Nearly every developing country faces cholera outbreaks or the risk of a cholera epidemic[47, 52].

The disease has been a public health burden in Ghana since 1970 when the first case was reported[61]. Between 1999 and 2005, a total of 26,924 cases and 620 deaths were reported officially to the World Health Organization (WHO). In addition to human suffering and loss of lives, a cholera outbreak causes panic, disrupts socio-economic activities and can hinder development in the affected areas. It is more prevalent in the Kumasi metropolis than other districts within the Ashanti region of Ghana. Various factors influence this, including overcrowding, urbanization, proximity and density of refuse dumps, proximity to water sources and poverty[31, 59, 60]. Of particular concern is the spatial distribution of cholera incidences and its associated risks factors. This can be used by health officials and policymakers to make appropriate planning and resource allocation. Also, it will be useful in limiting the severity and duration of an outbreak.

Spatial epidemiology is the study of the spatial distribution of disease incidence and its association to potential risk factors[34, 78]. Detail descriptions of spatial epidemiology and its applications can be found in[34, 62, 68, 78]. Most studies about cholera epidemiology use spatial statistical methods and geographical information systems (GIS) to map the disease[39, 33, 61, 72].

Satellite images can greatly enhance mapping of the environmental factors associated with cholera risk. Integrating satellite images, spatial statistics and GIS can provide public health officials with vital information needed to detect and manage cholera outbreaks. In order to correctly plan, manage and monitor any public health system, it is important to have up to date, relevant and complete information available to decision-makers.

Despite the availability of remotely sensed data in various formats for mapping, few studies have utilized the technology for mapping the environmental niche of *V. cholerae*. Even though the environmental variables have been identified[61], there is no map characterizing their distribution in the Kumasi Metropolis. This thesis focuses on combining field observations on cholera and data

retrieved on environmental factors by remote sensing such as refuse dumps and water reservoirs. Spatial statistical methods are used to map the disease.

1.2 RESEARCH IDENTIFICATION

RapidEye AG is a German geospatial information provider focused on assisting in management decision-making through services based on their remotely sensed imagery. The company owns a five satellite constellation producing 5 meter resolution imagery. These five identical Earth observation satellites allows for a continual capture of imagery. Any point on earth can be accessed, which enables rapid response for crop, environmental and emergency monitoring[64].

Identified environmental variables such as water reservoirs and refuse dumps can be derived from RapidEye images. Maps of these identified variables and cholera incidence data can be integrated in a GIS and analyzed using spatial statistical tools.

1.2.1 General objective

The main objective is to identify and characterize the spatial distribution of environmental factors that increase the risk of cholera infection in the Kumasi metropolitan area of Ghana using GIS, remote sensing and spatial statistical methods.

1.2.2 Specific objectives

1. To map out potential cholera causing factors in the study area from a RapidEye image.
2. Visualize the relations between cholera incidence, water bodies and refuse dumps using a GIS.
3. Determine the spatial relationship between cholera incidence and potential cholera reservoirs and refuse dumps using spatial statistics.

1.2.3 Research questions

1. Which environmental factors relevant to cholera modeling and mapping can be extracted from a RapidEye image?
2. How can these environmental factors be extracted and what is their quality?
3. How can the derived remote sensing variables be combined with field data to produce maps of environmental factors?
4. How can maps of cholera risks be visualized in a GIS?
5. Which models are most effective for modeling the effects of environmental risk factors on cholera?

1.3 OUTLINE OF THESIS

The thesis comprises of seven(7) chapters.

Chapter 1 - Introduction

This chapter contains the motivation and problem statement, research objectives, research questions and the methodology.

Chapter 2 - Literature review

This chapter contains literature about cholera, statistical methods for spatial epidemiology, areal data and applying spatial regression (Conditional autoregressive models) in modelling areal data.

Chapter 3 - Study area, datasets and data preparation

This chapter consists of the study area, description of data sources and preparation.

Chapter 4 - Research methodology

This chapter describes the methods and and tools applied in the study to produce the outcome.

Chapter 5 - Results and analysis

Chapter 6 - Discussion

Chapter 7 - Conclusions and Recommendations

Chapter 2
Literature review

2.1 CHOLERA

Cholera is an acute diarrheal illness caused by a bacterium, *Vibrio cholerae* (*V.cholerae*) leading to intestinal infections[52, 13, 2]. There are numerous environmental strains of *Vibrio cholerae*, which are found mainly in brackish waters and marine environments, but only two strains are responsible for cholera epidemics in humans, serogroups O1 or O139[47, 67]. Cholera mainly affects the small intestines after ingestion of sufficient dose of the *V.cholerae* baterium through water and/or food that is contaminated[47, 48, 67].

Cholera infection may often appear to be mild or even without symptoms, but can sometimes be severe. Nearly one in 20 (5%) infected persons will have severe disease characterized by lots of watery diarrhea, vomiting, and leg cramps[69]. For these people, rapid loss of body fluids leads to dehydration and shock. Without immediate treatment, death can occur within a short time (few hours). The symptoms typically start suddenly, between one to five days after consumption of contaminated water or food. The watery diarrhea may have a fishy odor and an infected person may produce 10 to 20 litres (0.01 to 0.02 m^3) of diarrhea a day[69, 82]. Moreover, it causes intense thirst, loss of skin turgor, wrinkled skin of hands and feet, sunken eyes, pinched facial expression, thready or absent peripheral pulses, falling blood pressure, and inaudible hypoactive bowel sounds. If the severe diarrhea and vomiting are not aggressively treated it can, within hours, result in life-threatening dehydration and electrolyte imbalances[69, 48].

Though the bacterium can live naturally in any environment[52], studies have shown that *V. cholerae* exist as natural inhabitants of aquatic ecosystems[42, 67, 21]. They usually occur as part of flora of streams, riverine, brackish water, estuarine and coastal waters. They attach to surfaces provided by plants, filamentous green algae, copepods (zooplankton), crustaceans, and insects[21, 42]. Stagnant water and slow flowing water may also lead to an increasing exposure of the organism. According to Islam et al.[42], the baterium can survive in almost all kinds of aquatic environments including fresh water sources such as lakes, ponds, rivers and tanks. Cholera bateria can also survive in non-aquatic environments such as refuse dump sites, fruits, fresh vegetables, meat, cooked food[26], human and animal faecal waste, untreated or inadequately treated sewage.

2.1.1 Cholera Transmission

Sources of infection

Contaminated water with free-living *V. cholerae* cells are the main source of cholera, followed to a lesser extent by contaminated food[22, 57, 13], particularly seafood likecrab, oysters, and shellfish. Bad sanitation practices in highly populated areas harboring the bacteria are the source of intermittent outbreaks due to contamination of drinking water and/or improper food preparation. The source of the infection is usually other cholera sufferers when their untreated diarrheal discharge is allowed to get into waterways or drinking water supplies. Drinking any of the infected water and eating any foods washed in the water, as well as seafoods living in the affected

waterbodies, can cause a person to contract an infection. The infection is seldom spread directly from person to person.

Transmission mechanism

There are two routes of cholera transmission namely primary transmission and secondary transmission[61, 20]. According to Osei[61] citing Hartley et al.[36], primary transmission happens through exposure to an environmental reservoir of V.cholerae or contaminated water sources regardless of previously infected persons or faecal contamination. Therefore, aquatic environments are essential for the spread of cholera.

Secondary transmission route on the other hand occurs through exposure to fecally contaminated water sources, food or infected person. Osei[61] argued that this route reflect a complicated tranmission pattern, since multiple factors may play a role in the spread of the disease. For instance, faecal-oral transmission is increased by the degree of contamination of water supply as well as frequency of contacts of these waters[19], which is in turn influenced by local environmental factors, socioeconomic, demographic as well as sanitation conditions. Both routes are clearly important and useful to control epidemics.

2.1.2 Factors influencing spread

The cholera germ is passed in the stools of infected persons . It is widely spread by consuming food or water which has been contaminated by the fecal waste/stools of an infected person. This happens more often in developing countries. This is so because underdeveloped countries lack adequate clean water supplies for drinking and proper sewage disposal systems as well as also practices poor sanitation and poor food hygiene [40, 41, 45, 47]. Once cholera is introduced to a population in a specific location, numerous complex factors decisively influence its propagation and may lead to prolonged transmission[70, 18, 45, 41]. Socioeconomic, environmental, demographic and climatic factors enhances the vulnerability of a population to infection and contribute to the epidemic spread of cholera[40, 31, 14, 3, 56]. These factors include the following:

1. Poor sanitation

   Cholera is hypothesized as a disease of deficient sanitation[2, 41, 45]. The lack of adequate toiletry, cleaning, washing and drainage facilities results in sickness and increases the risk of transmission.

2. High poverty and low income level

   Borroto and Martinez-Piedra[14] and Talavera and Pérez[73] identified poverty as an important predictor of cholera. Low income levels result in poor diet, malnutrition, poor housing facilities and lack of access to education. Typically, the world's poorest people obtain drinking water from a river/streams or wells; in the absence of toilet facilities or public sewage systems, people defecate near these rivers and streams allowing human waste to mix with the same water used for drinking.

3. High migration

   This plays a role by introducing cholera into new populations[3, 31]

4. Cooking practices

   Light cooking of food which has been contaminated[45]

5. Overcrowding/high population

   High population will lead to overcrowding putting strain on existing sanitation systems, thereby putting the population at high risk[15, 71, 31, 51].

6. Lack of Clean drinking water

   Unsafe water supply/contaminated water will increase the risk of cholera infection[72, 45]

7. Proximity and density of refuse dumps

   According to Osei and Duker[59], there is a direct linear relationship between cholera prevalence and refuse dumps density and an inverse relationship with proximity to refuse dumps. Two explanations given were (1) high rate of contact with filth breeding flies;they argued that filth breeding flies serve as a carrier of the *V.cholerae* from refuse dump sites where all kinds of human garbage and excreta is disposed to humans and (2) flood water contamination; in the event of heavy rains, runoff from open spaced refuse dumps serves as a pathway for the distribution of the bacteria, washing infected excreta into wells, streams and surface water bodies.

8. Proximity to surface water sources

   Close proximity to contaminated drinking water bodies make inhabitants more prevalent to cholera[3, 72, 60]

9. Climatic conditions

   Studies have shown that there is direct correlation between cholera and sea surface temperature, sea surface height, precipitation and chlorophyll concentrations[55, 21, 50].

10. Poor personal hygienic standards

    Poor personal hygiene increases cholera propagation within a given environment.

Figure 2.1 is a table of a summary of cholera risk factors taken from Collins et al[20].

2.1.3 The burden of cholera in Ghana

The disease has been a public health burden in Ghana since 1970 when the first case was reported[61]. Between 1999 and 2005, a total of 26,924 cases and 620 deaths were reported officially to the World Health Organization (WHO). In addition to human suffering and loss of lives, cholera outbreaks causes panic, disrupt socio-economic activities and can impede development in the affected communities.

2.2 SPATIAL EPIDEMIOLOGY

"Spatial Epidemiology is the description and analysis of the geographic, or spatial, variations in disease with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors"[24]. The spread of infectious diseases is closely associated with the concepts of spatial and spatio- temporal proximity, as individuals who are linked in a spatial and a temporal sense are at a higher risk of getting infected[62]. Proximity to environmental risk factors is therefore important. Thus knowledge of the spatial and temporal variations of diseases and characterizing its spatial structure is essential for the epidemiologist to understand better the population's interactions with its environment[61].

<table>
<tr><th>Underlying risks</th><th colspan="2">Proximate risks</th><th>Risk behaviors</th><th>Risk enhancers</th></tr>
<tr><td rowspan="11">Poverty
Poor housing
Overcrowding

Poor education levels, especially for females
Malnutrition

Poor drainage systems

And/or poor municipal sanitation systems. No or poor municipal piped water supply systems

Lack of community support

Low Human Development Index (HDI)

Changes in ambient conditions, including climate, which increase proliferation of plankton index</td><td rowspan="6">Faeces in the environment-Secondary transmission</td><td>Faeces in the environment</td><td>Open defecation, Poor maintenance or bad construction of toilets</td><td>Lack of toilets
–lack of resources
–dislike, distrust of toilets</td></tr>
<tr><td>Faeces on hands</td><td>Not washing hands with soap after defecation or after dealing with child feaces</td><td>No soap available/unable to afford soap</td></tr>
<tr><td>Contaminated food</td><td rowspan="2">Not washing raw fruits and vegetables</td><td>Lack of clean water available Live far from water source</td></tr>
<tr><td>Contaminated soil</td><td>Storing food at ambient temperatures allows pathogens to multiply</td></tr>
<tr><td>Contaminated flies</td><td>Not covering food Uncovering refuse or open toilets</td><td>Lack of refuse disposal Environmental unsuitable for good refuse disposal</td></tr>
<tr><td rowspan="2">Contaminated water</td><td>Drinking contaminated water</td><td>Easily contaminated water sources, e.g. open water source</td></tr>
<tr><td rowspan="2">Aquatic reservoir of Vibrio cholerae-Primary transmission</td><td>Washing food with contaminated water</td><td>Poor storage of water
No disinfection of water(chemical, filtering, Boiling UV)
Live far from water source</td></tr>
<tr><td>Contaminated bathing water</td><td>Bathing, washing and gargling in contaminated water points</td><td>Lack of alternative bathing and washing facilities</td></tr>
<tr><td>Contaminated shell fish</td><td rowspan="2">Eating undercooked/raw fish</td><td>Coastal-based livelihoods</td></tr>
<tr><td></td><td>Lack of fuel</td></tr>
</table>

Source: Collins et al[20]

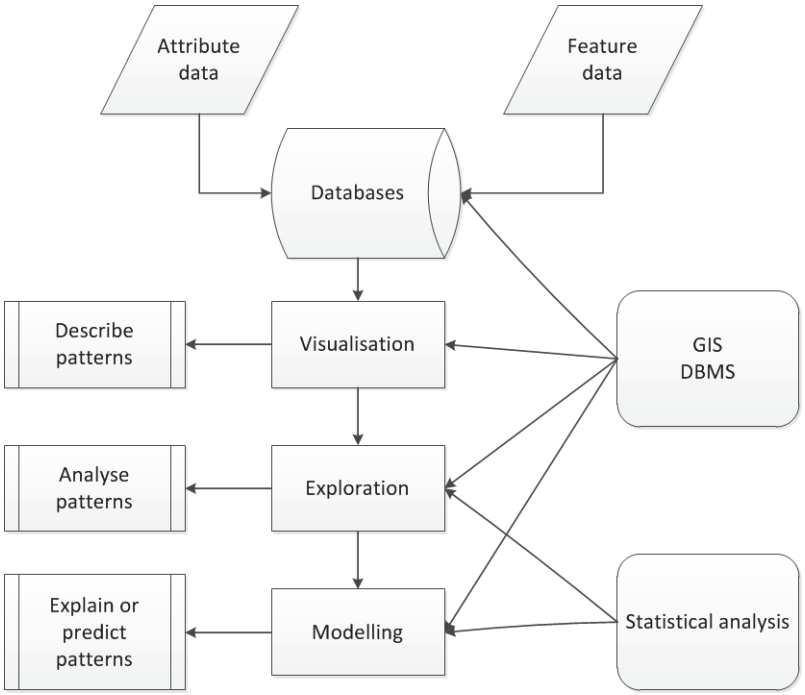
Figure 2.1: Cholera risk factors

Spatial epidemiology dates back to the 1800s, when maps of disease rates in different countries began to emerge to characterize the spread and possible causes of outbreaks of infectious diseases such as yellow fever and cholera[79]. Spatial analysis in the nineteenth and twentieth century was mostly employed by plotting the observed disease cases or rates. For example, Snow[72] mapped cholera cases together with the locations of water sources in London, and showed that contaminated water was the major cause of the disease. Recent advances in technology now allow not only disease mapping but also the application of spatial statistical methods[23, 44], satellite derived data[37] and Geographical information systems(GIS)[3, 60].

2.2.1 Framework for spatial analysis

Spatial epidemiology comprises of a wide range of methods. Determining which ones to use can be challenging[62]. Fig. 2.2 is a diagrammatic representation of a spatial analysis framework taken from Pfeiffer et. al[62] adapted from Bailey and Gatrell[9]. Pfeiffer et. al[62] identified

four groups as illustrated in Fig. 2.2 that can be used to define a logical, sequential process for conducting spatial analysis:



Source: Pfeiffer et. al[62], adapted from Bailey and Gatrell[9]

Figure 2.2: Conceptual framework of spatial epidemiological data analysis

1. Data

   "The objectives of spatial epidemiological analysis are the description of spatial patterns, identification of disease clusters, and explanation or prediction of disease risk"[62]. Central to these objectives is the need for data. Geographic data systems include georeferenced feature data and attributes, be they points or areas. These data are obtained by taking field surveys, remotely sensed imagery or use of existing data generated either by government organizations or those closely linked to government such as cadastral, postal, meteorological or national census statistics and health organizations.

2. GIS and DBMS

   Management of the data is performed using GIS and database management systems (DBMS), and is of relevance throughout the various phases of spatial data analysis. GIS provide a platform for managing these data, computing spatial relationships such as proximity to source of infection, connectivity and directional relationships between spatial units, and visualizing both the raw data and results from spatial analysis within a cartographic context[62].

3. Visualization and exploration

   Visualization and exploration cover techniques that focus solely on examining the spatial dimension of the data. Visualization tools are used resulting in maps that describe spatial patterns and which are useful for both stimulating more complex analyses and for communicating the results of such analyses. Exploration of spatial data involves the use of statistical

methods to determine whether observed patterns are random in space. However there is some overlap between visualization and exploration, since meaningful visual presentation will require the use of quantitative analytical methods[53].

4. Modeling

   Analytical procedures that simulates real-world conditions within a GIS using the spatial relationships of geographic features. Modeling introduces the concept of cause-effect relationships using both spatial and non-spatial data sources to explain or predict spatial patterns[62].

However, this is not a linear process, as presenting the results from exploration and modeling requires a return to visualization.

2.2.2 Statistical methods for spatial epidemiology

There exist various epidemiological inquiry which include disease mapping, geographic correlation studies and clustering/cluster detection.

Disease mapping

Disease mapping provide information on a measure of disease occurrence across a geographic space. Disease maps are able to provide us a rapid visual summary of complex geographic information. These maps may also identify subtle patterns in epidemic/health data that are sometimes missed in tabular presentations[24]. The aims of disease mapping include:

- Simple description by showing or displaying a visual summary of geographical risk for example, the map of Snow[72] in fig. 2.3.
- Hypothesis generation by giving clues to causes of diseases and or factors that influence spread by informal examination of maps with exposure maps, components of spatial versus non-spatial residual variability may also provide clues to source of variability. The formal examination is carried out via spatial regression.
- Provide estimates of risk by area to inform public health resource allocation.
- Estimation of background variability in underlying risk in order to place epidemiological studies in context.

Geographic correlation studies

The objective of geographic correlation studies is to examine geographic disparities across inhabitants in an exposure to environmental variables which may be measured in air, water, or soil, socioeconomic and demographic measures such as race and income, or lifestyle factors such as smoking and diet in relation to health outcomes measured on a geographic scale[24]. Correlation studies also aims at:

- Examination of the association between disease outcome and explanatory variables, in a spatial setting, using regression models.
- Conventional modeling approaches such as logistic regression for point data, and loglinear models for count data.
- the examination of risk with respect to a specific point or line putative source of pollution.

- For count data, disease mapping models can be extended to incorporate a regression component.

Correlation studies deals with the association between disease risk and exposures of interest. we examine the association between risk and exposures at the area level via ecological regression using Poisson regression as a framework for areal data and logistic regression for point data.

Clustering/Cluster detection

Clustering examines tendency for disease risk to exhibit "clumpiness", while the Cluster detection refers to on-line surveillance or retrospective analysis, to reveal "hot spots". The aim is to investigate disease clusters and disease incidence near a point source[46].

2.2.3 Spatial epidemiology of cholera

John Snow was the first to map cholera[72]. In his study, Snow was able to assess the spatial pattern of cholera cases in relation to potential risk factors, in this instance the locations of water pumps. He furthermore made a solid use of statistics to demonstrate the connection between the quality of the source of water and cholera incidence and used a dot map to illustrate how cases of cholera clustered around the Broad Street water pump in London (See fig. 2.3).

After Snow's work, some epidemiological studies on cholera have focused on pathogenesis and biological characteristics of *V.Cholerae*[32, 67]. These studies have been useful, in understanding the environments that are most suitable for the bacteria.

To be able to identify and map environmental factors that impact risk of cholera, spatial epidemiological tools have to be applied in cholera studies. Understanding the spatial relationship between cholera and environmental risk factors have been a challenge for long. Recent studies have used GIS based and statistical methods in mapping the disease[3, 60, 25, 29]. Osei and Duker[59] mapped locations of all an environmental risk factor (refuse dumps) using a Global Positioning System (GPS) in Kumasi, Ghana. They created two spatial factor maps using GIS and spatial analysis, a spatial distance surface showing proximity of distances from community to refuse dumps and a kernel density surface showing the number of refuse dumps per density area. Two spatial covariates were derived and used as explanatory variables in spatial regression model to relate cholera incidence to refuse dumps in Kumasi. In a related study, Osei et. al[60], potential cholera reservoirs (rivers and streams) and elevation were digitized from a topographic map. Spatial distance factor maps of nearest reservoirs to communities were created and used as covariates in spatial regression modeling.

2.3 AREAL DATA AND SPATIAL AUTOCORRELATION

Area data are observations associated with a fixed number of areal units. The areas may form a regular lattice, as with remotely sensed images, or be a set of irregular areas or zones, such as countries, districts and census zones[27]. Data about individuals are often available only at an aggregated areal level in order to protect personal information. For example, average income levels for census tracts are readily available, but the income of an individual person in that census tract is usually not available. Similarly, the total number of people with cholera in a health service area might be known, but not each person's individual location within that area.

Spatial autocorrelation statistics are used to measure and analyze the degree of spatial correlation/dependency among observations in a geographic space[28]. The principle underlying the analysis of spatial data is the proposition that values of a variable in near-by locations are more similar or related than values in locations that are far apart. This inverse relation between value

Source: Snow[72]

Figure 2.3: John Snow's 1854 cholera-outbreak map of London (deaths shown as dots, water pumps as crosses)

association and distance is summarised by Tobler's first law that *"everything is related to everything else, but near things are more related than distant things"*[74].

2.3.1 Spatial weights and neighborhoods

An important aspect of defining spatial association is the determination of the relevant neighborhood of a given area, that is, those areal units surrounding a given data point (area) that would be considered to influence the observation at that data point. This is a necessary step in using areal data[27]. These neighboring areas are spatial units that interact in a meaningful way. This interaction could relate, for example, to spatial spillovers and externalities[46].

Spatial autocorrelation measures require a weights matrix that defines a local neighborhood around each geographic area/unit[5]. The value at each areal unit is compared with the weighted average of the values of its neighbors. A weighting system is chosen and assigned to the neighborhoods. Weights can be constructed based on either contiguity to the polygon boundary (shape) files, or calculated from the distance between points (points in a point shape file or centroids of polygons)[12, 78, 5]. The weights are usually row-standardized to ensure that the row members for each observation sum to 1, with zero on the diagonal and some non-zero off-diagonal

elements. The formula for each weight is:

$$w_{ij} = \frac{C_{ij}}{\sum_{j=1}^{N} C_{ij}} \tag{2.1}$$

With

$$C_{ij} = 1$$

when i is linked to j, otherwise

$$C_{ij} = 0$$

The spatial weights reflects the strength of the geographic relationship between observations in a neighborhood, e.g., the distances between neighbors, the lengths of shared border, or whether they fall into a specified directional class such as "North".

2.3.2 Spatial autocorrelation test

Spatial autocorrelation statistics include Global Moran's I and Local Moran's I (LISA). These measures compare the spatial weights to the covariance relationship at pairs of locations. A spatial autocorrelation value observed to be positive than expected from random shows there is clustering of similar values across geographic space, while significant negative spatial autocorrelation indicates that neighboring values are more dissimilar than expected by random, suggesting there is a spatial pattern similar to that of a chess board[5].

2.3.3 Global Moran's I

Moran statistics are one class of measures of spatial autocorrelation. Global autocorrelation statistics provide a single measure of spatial autocorrelation for an attribute in a region as a whole[5].

$$I = \frac{N}{\sum_i \sum_j W_{ij}} \times \frac{\sum_i \sum_j W_{ij}\,(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \tag{2.2}$$

where there are N units, the attribute value for each unit i is y_i, and W_{ij} is the weight (or connectivity) for units i and j. The locational information for this formula is found in the weights. Therefore, for non-neighboring tracts, the weight is zero, so these are not used in the calculation of correlation. The expected value of Moran's I is

$$E(I) = \frac{-1}{(n-1)} \tag{2.3}$$

A Moran I of $+1$ indicates strong positive spatial autocorrelation (i.e., clustering of similar values), 0 indicates random spatial ordering, and -1 indicates strong negative spatial autocorrelation (i.e., a checkerboard pattern)[4].

2.3.4 Local indicators of spatial autocorrelation(LISA)

Local spatial autocorrelation statistics provide a measure, for each unit in the region, of the unit's tendency to have an attribute value that is correlated with values in nearby areas.

$$I_i = z_i \sum_j w_{ij} z_j \tag{2.4}$$

Where z_i andz_j are standardized scores of attribute values for unit i and j, and j is among the identified neighbors of i according to the weights matrix w_{ij}. The local spatial autocorrelation analysis is based on LISA statistics and computes a measure of spatial association for each individual location[4].

2.4 MODELING AREAL DATA

Analysing public health data involve disease counts, proportions, or rates. These counts or rates are not continuous like the continuous outcomes familiar in linear regression. Whereas large counts or rates may roughly follow the assumptions of linear models, spatial analyses often focus on counts from small areas with relatively few subjects at risk and few cases expected during the study period. Such instances require models appropriate for count or rate outcome[27].

Modeling spatial interactions that arise in spatially referenced data is commonly done by incorporating the spatial dependence into the covariance structure either explicitly or implicitly via an autoregressive model. In the case of lattice (areal) data, two common autoregressive models used are the conditional autoregressive model (CAR) and the simultaneously autoregressive model (SAR). Both of these models produce spatial dependence in the covariance structure as a function of a neighbor matrix, W and often a fixed unknown spatial correlation parameter[78, 77].

2.4.1 Spatial regression models

Spatial regression models are statistical models that account for the presence of spatial effects, i.e., spatial autocorrelation or spatial dependence and/or spatial heterogeneity. This is not the case with the standard linear regression model[7]. The standard linear regression using least squares (OLS) is used to find a linear relationship between a dependent variable and a set of explanatory variables. It is written in vector form as

$$y = X\beta + \varepsilon \tag{2.5}$$

where $\varepsilon \sim\ N\left(0, \sigma^2\, I\right)$ which can be written as:

$$y \sim\ N\left(X\beta, \sigma^2\, I\right) \tag{2.6}$$

But with spatial data, the assumption that the error terms are independent and normally distributed may not hold. For spatial regression, the error terms are assumed to be correlated. Moreover it includes spatial dependency in regression analysis, in which case a general model adopted is[30]:

$$y \sim\ N\left(X\beta, A\right) \tag{2.7}$$

where A is a positive definite symmetric covariance matrix which allows non-zero covariances amongst the error terms. Now, A is chosen so that elements of y that are closer to each other in space also have higher covariance. Denoting $X\beta$ as μ, (2.6) can be simplified as:

$$y \sim\ N\left(\mu, A\right) \tag{2.8}$$

This represents a generic assumption for spatial regression models. Two ways in which such models are specified are the conditional autoregressive(CAR) and the simultaneous autoregressive (SAR) models. These Spatial autoregressive models were developed primarily for use with geographically aggregated spatial data where measurements could be taken at any location in the study area in contrast to the geostatistical models developed for spatially continuous data[78].

2.4.2 Conditional autoregressive (CAR) models

The primary purpose of CAR models is to provide a modeling mechanism to account for residual spatial correlation i.e., spatial trends not explained by spatial patterns in covariate values [78]. For

CAR models, the y-variable is assumed to be dependent not only on explanatory x-variables but also on other nearby y-variables. The model is specified as

$$y_i|\{y_j : j \ \ i\} \sim \ N\left(\mu_i + \ \sum_{j=1}^{n} c_{\,ij}(y_i - \mu_j), \lambda^2\right) \tag{2.9}$$

That is, the distribution of y_i conditional on all the other y-values is normal. The distribution is expressed in terms of $y_i - \mu_i$, the difference between the observed y_i and the expected value of y_i obtained when considering the x-variables. Some restrictions are imposed on the spatial weight values, the c_{ij} values. First, c_{ij}=c_{ji} and second,$c_{ii} = 0$. The former restriction means the weights must be symmetrical; the latter restriction simply means that the conditional distribution of y_i cannot depend on y_i itself - only on other y-values. Typically, the c_{ij}s are chosen to reflect the spatial structure of the data. If the data are associated with a set of zones, then c_{ij} might be defined as 1 if zones i and j are contiguous, and 0 otherwise. For point data, c_{ij} might be defined as a continuous function of distance such as $kd_{ij}^{-\alpha}$ where α =1 or 2 and d_{ij} is the distance between points i and j (assuming that there are no coincident points). This latter scheme could also be applied to zonal/area data using distances between zone centroids.

Equation (2.9) can be rewritten as:

$$y \sim \ N\left(\mu, (I - C)^{-1}\lambda^2\right) \tag{2.10}$$

where λ^2 is the conditional variance, and the variance-covariance matrix $C = c_{ij}$, from which it can be seen that, restricting $c_{ij} = c_{ji}$ stops the matrix $I - C$ from becoming ill-defined.

2.4.3 Simultaneous autoregressive (SAR) models

A simultaneous autoregressive model can be defined as

$$y \sim \ N\left(\mu_i + \ \sum_{j=1}^{n} b_{ij}(y_i - \mu_j), \tau^2\right) \tag{2.11}$$

This differs from the CAR model because the distribution of y_i is not conditional. In this case the marginal distributions for all the y_is are specified as a system of simultaneous equations. Again, the restriction on the spatial weights that b_{ii}= 0 is imposed but there is no longer a symmetry constraint on the weight matrix. Equation (2.11) can also be re-written as;

$$y \sim \ N\left(\mu, \tau^2(I - B)^{-1}(I - B^T)^{-1}\right) \tag{2.12}$$

similarly to equation (2.10), $B = b_{ij}$

Differences between SAR and CAR

The structure of B and C is usually specified by the shape of the lattice. One common way to construct B or C is with a single parameter that scales a user defined neighborhood matrix W that indicates whether the regions are neighbors or not as described in Equation (2.1). Thus, for the SAR model $B = \lambda_s W$ and for the CAR model $C = \lambda_c W$ where λ_s and λ_c are often referred to as "spatial correlation or spatial dependence" parameters and are left to be estimated.

- Marginal variances differ, even if $\lambda_c = \lambda_s$, all variances are location independent and are the same.
- The CAR model with constant variance requires that the matrix W is symmetric
- OLS is inconsistent for the SAR model

For this study, the CAR model was considered and used.

Chapter 3

Study area, datasets and data preparation

3.1 THE STUDY AREA

The study area is the Kumasi metropolitan, which is the second largest urban center of Ghana. Kumasi metropolis is one of 18 districts and capital of the Ashanti Region (See Figure 3.1). Ashanti Region is centrally located in the middle belt of Ghana. The metropolis lies at the intersection of latitude 6.04^0N and longitude 1.28^0W, covering an area of about $220km^2$[31, 83]. According to Osei and Duker[31], Kumasi has a population of about 1.2 million which accounts for just under a third (i.e. 32.4%) of Ahanti's region population. The metropolis is subdivided in communities. Unfortunately, these communities have no established boundaries. For this study, 68 communities were used.

There are two major seasons, the rainy and dry season. The rainfall pattern is bimodal with long rainy season from April to July, sometimes with peaks in May/June and a short season September and Mid-November[60]. As described by Osei[61], approximately 82% of the inhabitants in Kumasi have access to portable, pipe-borne water, however surface water from rivers and streams is still used largely for cooking, bathing and washing utensils due to rampant water shortages. According to Osei et al.[60], the coverage rate of safe house-to-house collection of solid waste is very low as such a greater proportion of household, approximately 81.2%, dispose of solid waste at open space refuse dumps[59]. Furthermore, most demarcated areas for public sanitation and waste disposal facilities have been sold out due to high demand for land compelling inhabitants to defecate at open space refuse dumps[61].

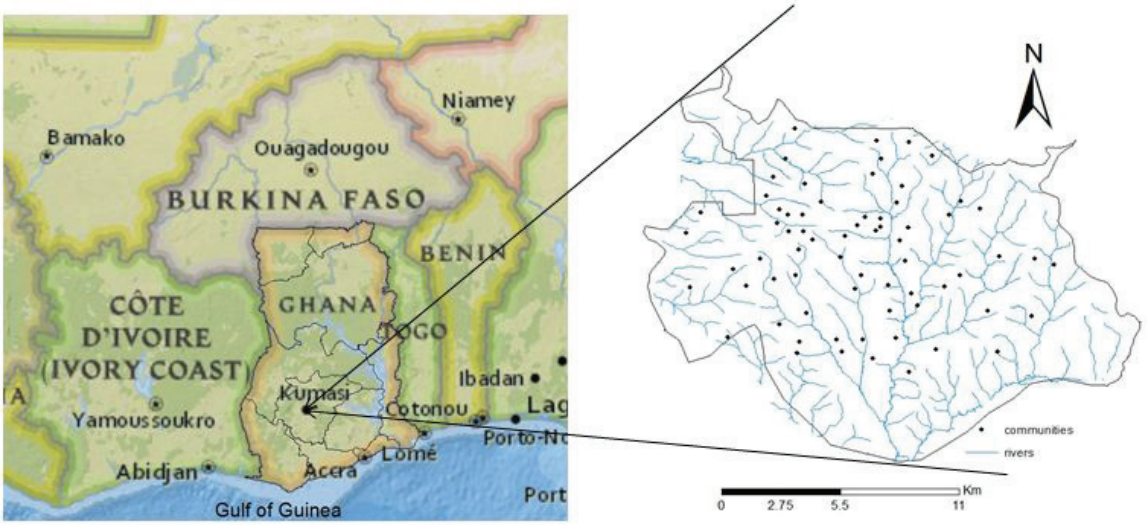


Figure 3.1: Regional map of Ghana(left) and Kumasi(right)

3.2 DATASETS

3.2.1 Cholera data

The cholera data was obtained from Osei[61]. The data was collected from the Kumasi Metropolitan Disease Control Unit (DCU) in the year 2005 during which there were severe outbreaks of cholera in Kumasi, Ghana. According to Osei and Duker[59], the outbreak lasted for 72 days, which was within the rainy season. The data consist of number of reported cases per community (spatial unit for reporting). Each community is represented as point shapefile feature with X, Y coordinates in meters and has number of cholera cases reported in 2005, population estimates for 2005 and Raw rates as attributes. The population estimates, obtained from the Ghana Statistical service(GSS), were used in calculating the raw rates. Raw rates were calculated as number of cholera cases in each community divided by the estimated population in 2005 and rescaled by multiplying it by a factor of 10,000 to express the raw rates as per 10,000 people more intuitively. This study utilized only cholera cases reported during the 2005 outbreak.

3.2.2 Refuse dumps data

The composition of waste in Kumasi is predominantly made of biodegradable(organic) materials and a high percentage of inert materials as well as small amounts of paper,plastic and metals[8]. The inert material is mostly made of wood ash, sand and charcoal. Solid waste management is contracted to a number of private companies by the Waste Management Department in Kumasi (WMD). The collection system of the waste management in the metropolis is based on two systems which are house-to-house waste collection and communal solid waste collection[75, 85]. The communal waste collection system consists of containers placed throughout the city (See Figure 3.2). The containers are being emptied by waste collection companies and transported to landfill sites located in the outskirts of the metropolis in a regular basis.

Source: Wikner[85]

Figure 3.2: A communal waste container, at KNUST Kentinkrono, Kumasi.

With house-to-house waste collection, the waste is collected at the yard or door at the households. At least 5 out of 10 household dispose of their waste right besides their houses, instead of finding the nearest waste dump[59]. Waste that is not being collected is being indiscriminately dumped in rivers (See Figure 3.3) and gutters/drains as shown in (Figure 3.4) or burned.

Source: Wikner[85]
Figure 3.3: Solid waste dumped in a river

Source: Wikner[85]
Figure 3.4: Solid waste dumped in a gutter

The refuse dumps data were obtained from Osei and Duker[59]. The data was collected in a field survey in 2005 using a Global Positioning System(GPS). A total of 124 refuse dumps were mapped. The refuse dumps data consist of only point shapefiles showing only the X and Y coordinates in meters.

3.2.3 Rivers data

Included in the data obtained from Osei[61] is a layer of stream segments which were digitized from a topographic map in 2005.

3.2.4 RapidEye image data

This research made use of a RapidEye sensor. The RapidEye sensor which captured the data used for classification of the potential cholera reservoirs is briefly described in the section below.

RapidEye sensor and image capture

RapidEye AG is a German geospatial information provider focused on assisting in management decision-making through services based on their own Earth observation imagery. RapidEye has five satellite constellation producing 5 meter resolution imagery that was designed and implemented by MacDonald Dettwiler (MDA) of Richmond, Canada. Each of the five satellites contain identical sensors, equally calibrated and travel on the same orbital plane (at an altitude of 630km). Together, the 5 satellites are capable of collecting over 4 million km^2 of 5 m resolution, 5-band color imagery every day[66, 84]. A summary of the image specification is shown in Table 3.1.

Table 3.1 RapidEye product specifications

Source: RapidEye[65]

Digital Data Product Specifications	Information
Spectral Bands	Capable of capturing any subset of the following spectral bands: Blue 440 - 510 nm Green 520 - 590 nm Red 630 - 685 nm Red Edge 690 - 730 nm NIR 760 - 850 nm
Ground sampling distance (nadir)	6.5 m
Pixel size (orthorectified)	5 m
Swath Width	77 km
On board data storage	1500 km of image data per orbit
Equator crossing time	11:00 am (approximately)

RapidEye image data of study area

Four image tiles covering the study area captured on 9th November 2009 were acquired. These images were RapidEye ortho-Level 3A specification. Level 3A offers the highest level of processing available. Hence, radiometric, sensor and geometric corrections have been applied to the data. It is a 16 bit data and consist of an image file, a metadata file, a browse image file and Unusable Data Mask(UDM) file. It has a UTM projection and WGS84 Horizontal datum system. The images were first mosaicked into one large image using ERDAS Imagine 2011 and the study area of interest subsetted from the larger image. A subset of the image is shown in Figure 3.5.

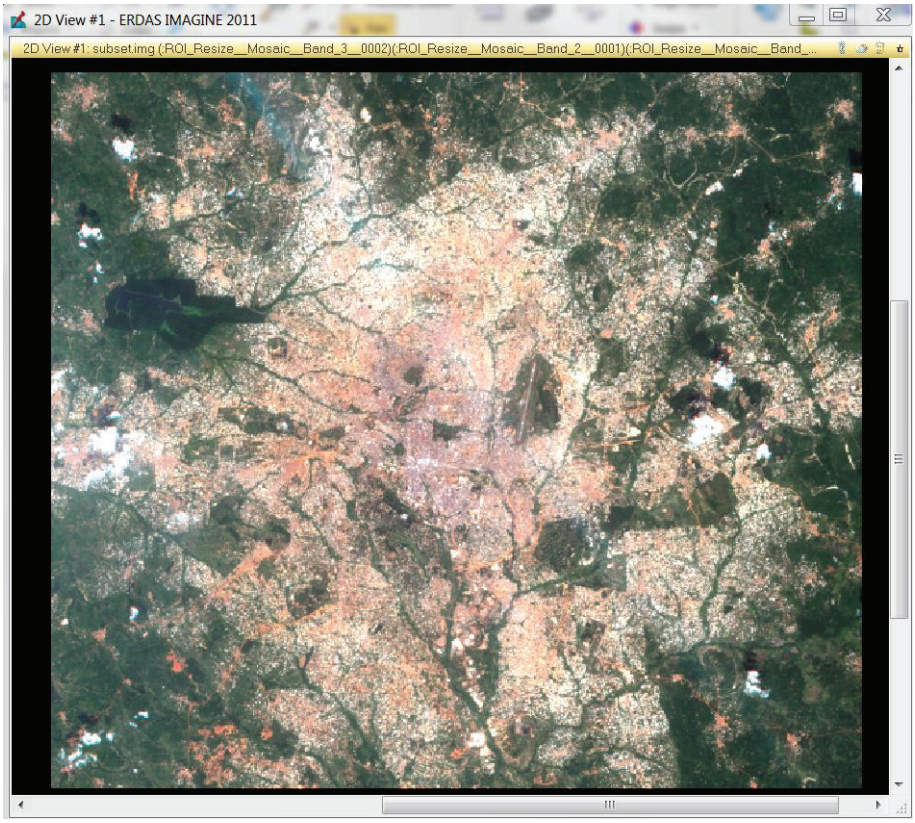


Figure 3.5: True colour subset of RapidEye image

3.3 DATA PREPARATION

The shapefiles for open space refuse dumps, community centroids and rivers/streams was adopted from Osei[61]. All three feature layers cover the whole of the Kumasi Metropolitan. The shapefiles when loaded into ArcMap had no coordinate system nor a reference ellipsoid. However, the data description from Osei[61] revealed that they were in the Ghana Transverse Mercator System(GTM). They were then reassigned to Accra Ghana Grid (Ghana coordinate system) and later converted to Universal Transverse Mercator UTM_WGS 1984 and then overlaid with the RapidEye image during image processing.

3.4 SOFTWARE

The following softwares were used for the study:

- ERDAS imagine 2011 for image processing/classification
- ArcGIS (ArcMap 10) for database creation, geovisualization and geospatial analysis
- OpenGeoDa, version 1.0.1 [7]for exploratory data analysis and geovisualization
- R-software, version 2.13.2[63] for statistical analysis. The R packages used in this study include; spdep[11], maptools[49] rgdal[43] and the RColorBrewer[58].

Chapter 4
Research methodology

4.1 INTRODUCTION

The image of the study area was classified to obtain a land cover map. Classified rivers and water-bodies were extracted and integrated with refuse dumps layer and cholera data and visulized using GIS. Spatial analysis and regression modeling was then carried out on the data. The flow chart for the research is shown in Figure 4.1.

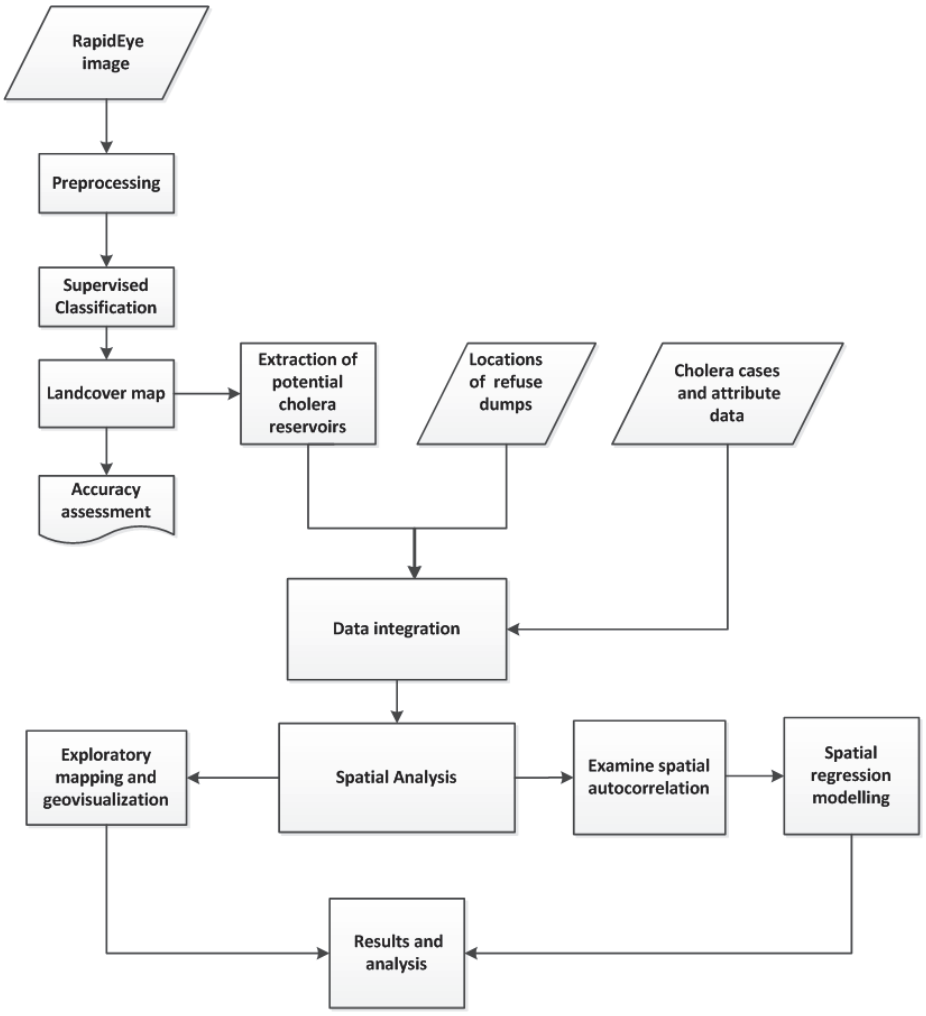


Figure 4.1: Flow chart of Methodology in the study

4.2 IMAGE ANALYSIS

4.2.1 Image pre-processing

The RapidEye satellite image was imported into ERDAS imagine. The image has been orthorectified with radiometric, geometric and terrain corrections in WGS1984(UTM Zone 30^0N) projection from the producer[66]. This implies, all pixels on the image has been assigned to real world coordinates in the WGS1984(UTM Zone 30^0N) projection system. This projection system was maintained, since Ghana uses the WGS1984(UTM Zone 30^0N) projection system.

4.2.2 Image classification

A pixel based approach (supervised classification) was used in classifying the image. The image classification process apportions the pixels of an image to exact spectral behavior of the ground data. Pixels are sorted into a finite number of individual classes, or categories of data, based on their data file values. If a pixel satisfies a certain set of criteria, then the pixel is assigned to the class that corresponds to that criteria. This process converts image data to thematic data. The landuse/landcover of the study area was classified to identify water reservoirs using the RapidEye image of 2009 using the maximum likelihood algorithm.

The results of the image classification were validated in order to assess their accuracy. For this study, a random sampling scheme was used to select 85 points (pixels) from the classification output and compared with the reference data. Comparison was done by creating an error matrix. The image classification and accuracy assessment were done in ERDAS imagine 2011.

4.3 DATA INTEGRATION AND VISUALIZATION

The communities do not have established boundaries between them, though the cholera cases were collected at the community level. Hence to analyze the data at this level, arbitrary boundaries were created with the aid of the communities centroids and the RapidEye image. This was done visually and arbitrary keeping in mind, natural boundaries such as rivers and that these communities form clusters at certain intervals.

The classified water reservoirs were extracted and converted into shapefiles. Two sources of water reservoirs are used in this study. They include water reservoirs obtained from the classification of the RapidEye image and digitized rivers/streams data that were adopted from Osei et.al[60]. They are analyzed separately in this study since they are measured on different scales. Maps of refuse dumps, community centroids and water reservoirs were overlaid on each other in ArcMap 10 and Spatial factor maps were created and visualized. A description of how the spatial factor maps were generated is shown in the next section below.

4.3.1 Spatial factor maps

Based on the hypothesis that cholera is a disease of deficient sanitation and assuming all water reservoirs (rivers/streams) are potential cholera reservoirs, the following predictions are made[59] :

- Inhabitants living in close proximity to open-spaced refuse dumps should have higher prevalence than those farther
- Areas with high density of open-spaced refuse dumps should have higher cholera prevalence than areas with lower density

- Inhabitants living in close proximity to to potential cholera reservoirs should have higher prevalence than those farther

Therefore to determine the spatial relationships between cholera prevalence per community and (a) distances to refuse dumps (b) density of refuse dumps and (c)distance to potential cholera reservoirs, four spatial factor maps were created using ArcMap. The spatial factor maps were overlaid with point map (centroids) of communities to create four explanatory variables (independent variables);

1. *Proximity to refuse dumps*; spatial distance surface, showing distances of each point or pixel to the nearest refuse dumps
2. *Density of refuse dumps*; kernel density surface, showing the number of refuse dumps per unit area
3. *Proximity to digitized reservoirs*; spatial distance surface, showing distances of each point or pixel to the nearest reservoir that describes all digitized streams and rivers from Osei et.al[60].
4. *Proximity to classified reservoirs*; spatial distance surface, showing distances of each point or pixel to the nearest reservoir that describes all water reservoirs extracted from RapidEye image.

The difference between (3) *Proximity to digitized reservoirs* and (4) *Proximity to classified reservoirs* is that (3) uses information of reservoirs that were digitized from a topographic map as adopted from Osei et.al[60] whilst (4) uses information of reservoirs that was classified from the RapidEye image. The reservoirs are from the same study area but measured separately. This way we can compare their effects on cholera to see if there are any significant differences.

Finally, a database for cholera was generated for the study area consisting of polygon boundaries of communities and points of community centroids. The attributes of each community were; number of cholera cases, cholera raw rate (per 10,000 people), population and the four explanatory variables discussed above. A description of how each of the explanatory variables were derived is discussed below.

Proximity to refuse dumps

Using Spatial Analyst extension and the Distance toolbox, Euclidean distance surface was generated in ArcMap, with refuse dumps layer selected as input feature source. This calculates for each cell, the euclidean distance to the closet source. In this study, a search radius of 1km was used. From the Analysis Tools extension, the Near tool from the Proximity toolbox was then used to determine the distance from each community centroid to the nearest refuse dump. This procedure adds a new field NEAR_DIST to the input community centroids attribute table and stores the paths of the centroids that contain the nearest refuse dumps. A map showing the distances from each pixel to the nearest potential cholera source is shown in Figure 4.2.

Density of refuse dumps

A kernel density surface, showing the number of refuse dumps per unit area (Figure 4.3) was created. This was done using the Density toolbox. Kernel density calculates the magnitude per unit area from point features using a kernel function to fit a smoothly tapered surface to each point. The surface value is highest at the location of the point and decreases with increasing distance from the point, reaching zero at a search radius distance from the point. A radius of 1km

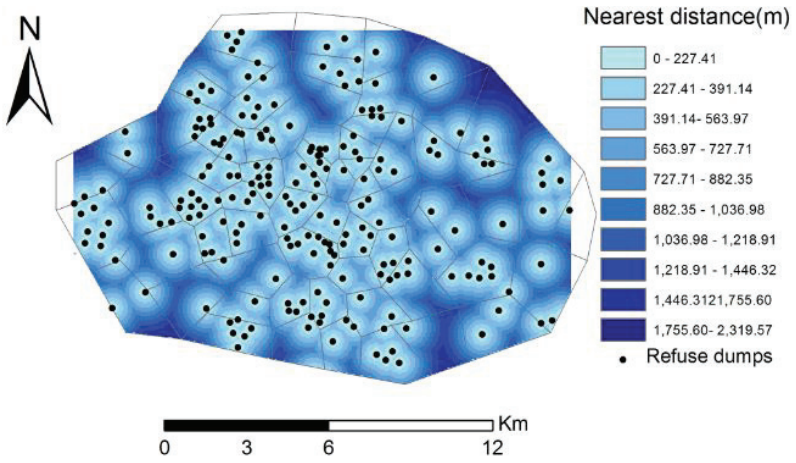


Figure 4.2: Distance surface of nearest refuses dumps

was used. The Extract Values to Points tool under the Extraction toolbox in the spatial analyst extension was used to extract the density values of community centroids and recorded into the attributes of the community feature class.

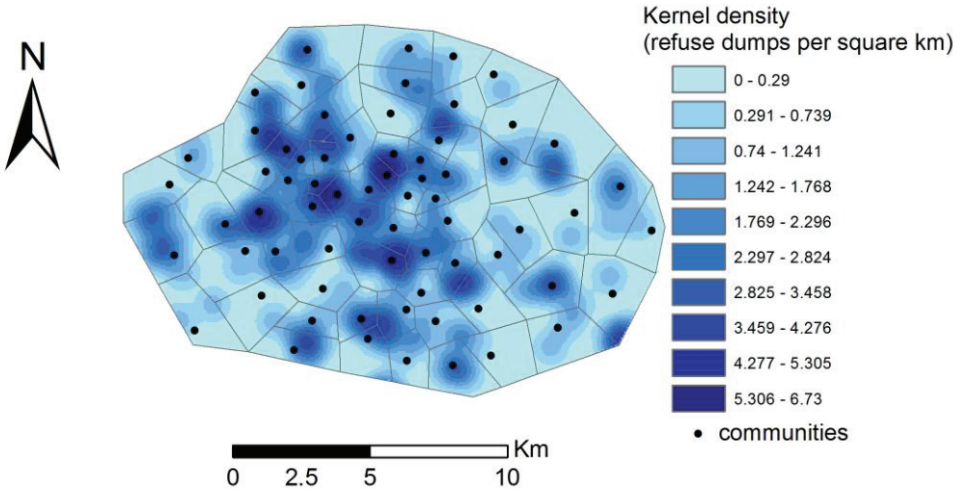


Figure 4.3: Kernel density surface

Proximity to digitized reservoirs

The same approach as discussed in section 4.3.1, subsection **Proximity to refuse dumps** was applied in this case. Here the input layer used are the digitized streams adopted from Osei et.al[60]. Figure 4.4 is the output distance surface map showing the distance from each pixel to the nearest potential cholera reservoir.

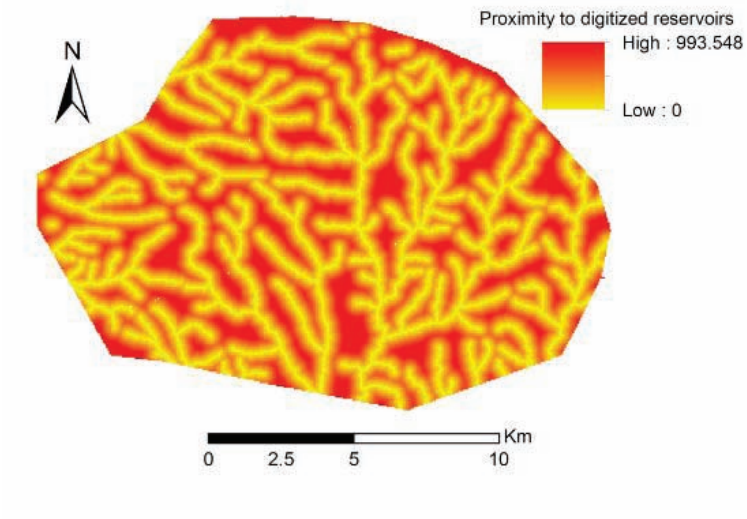


Figure 4.4: Proximity to digitized reservoirs

Proximity to classified reservoirs

Similarly to the method discussed in the section above, spatial distance surface (Figure 4.5) map, showing distances of each point or pixel to the nearest reservoir that describes all water reservoirs extracted from RapidEye image was created and overlaid with point map of communities to derive the nearest distances to the reservoirs.

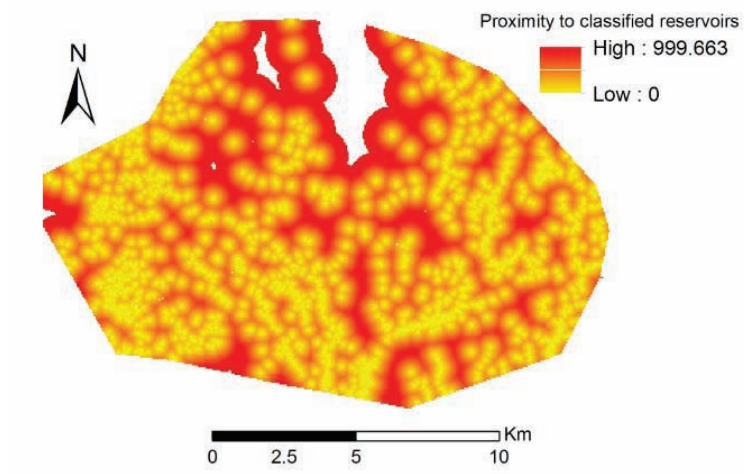


Figure 4.5: Proximity to classified reservoirs

4.3.2 Mapping and geovisualization

Before measuring and modelling spatial dependence, maps of cholera prevalence and risks were generated. Typically, we have counts of the incidence of cholera by spatial unit, associated with counts of populations at risk. The task is then to try to establish whether any spatial units seem to be characterized by higher or lower counts of cases than might have been expected in general terms[9].

The following maps were generated;

- Choropleth maps; these maps use different color and pattern combinations to depict different values of cholera incidence associated with each community.

- Proportional symbol maps; maps showing cholera incidence was generated in ArcMap 10, refuse dumps represented by dots and river segments layers were added to this map to visualize and characterize their distribution spatially.

- Choynowski's probability map[17]; based on the `choynosky ()` function, providing the probability values required for each community.

- Four probability maps; based on the `probmap ()` function in R package as described by Bailey and Gatrell[9]. The function returns a data frame of rates for counts in populations at risk with crude rates, expected counts of cases, relative risks, and Poisson probabilities.

4.4 SPATIAL ANALYSIS

Spatial analysis was carried out in two principle steps. Firstly, autocorrelation analysis was carried out to determine the extent of correlation among neighboring communities and secondly the four spatial covariates described in section 4.3 used as inputs in spatial autoregressive modeling.

4.4.1 Autocorrelation analysis

The following steps were carried out to determine the extent of spatial autocorrelation among neighboring communities:

1. choosing a neighborhood criterion;

2. creating a spatial weight matrix;

3. running statistical test, using weight matrix to examine spatial correlation.

Autocorrelation analysis was carried out in R. Before the analysis, the cholera polygon shapefile was imported into R and a SpatialPolygonsDataFrame created. The R libraries used include spdep[63], maptools[49], rgdal[43] and RColorBrewer[58].

Univariate (Local Indicators of Spatial Autocorrelation) LISA statistics developed by Luc Anselin[6] were invoked by clicking on the Univariate LISA button on the Explore toolbar in Geoda. This brings up a dialog that lets you specify which of the four output options you want to generate. The most relevant of these options are The Significance Map and The Cluster Map, which are unique to the LISA functionality. These two maps were generated using GeoDa.

Neighborhood criterion

The first step in the analysis of spatial autocorrelation is to define the neighborhood structure over the entire area. The neighborhood structure describes the areas that are linked. A first order rook contiguity criterion was used in this study. Rook contiguity uses only common boundaries to define neighbors. The rook neighborhood structure used is illustrated in Figure 4.6.

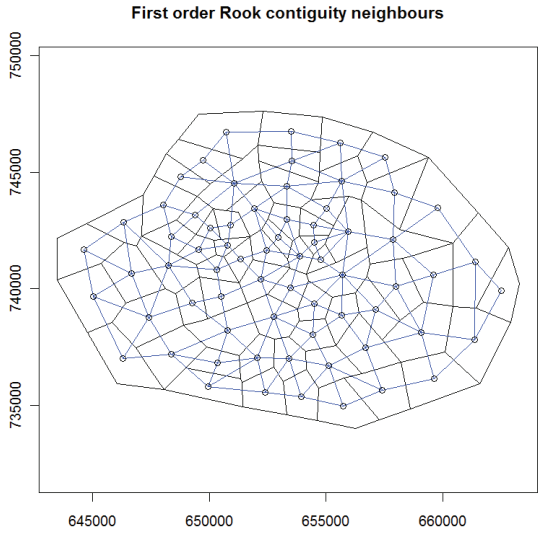


Figure 4.6: Rook neighborhood structure

Creating a spatial weight matrix

A spatial connectivity matrix that contains information of the neighborhood structure of each community was created and weights assigned to the areas that are linked. Row-standardized weights were applied. Row standardization is used to create proportional weights in cases where features have an unequal number of neighbors.

Running statistical test

A significance test against the null hypothesis of no spatial autocorrelation was used to test for the significance of the statistic. Moran's Index under randomization was calculated and a spatial correlogram computed and plotted.

4.4.2 Spatial autoregressive modeling

To attempt to determine the spatial relationship between cholera incidence (the response variable) and the four spatial covariates (explanatory variables) derived in section 4.3.1, i.e, proximity to refuse dumps(ND_RD), density of refuse dumps(Den_RD), proximity to digitized reservoirs(ND_DR) and proximity to classified reservoirs(ND_CR), two set of models were developed;

- *Model A*: relates cholera incidence with proximity to refuse dumps, dumps density and proximity to digitized reservoirs.
- *Model B*: relates cholera incidence with proximity to refuse dumps, dumps density and proximity to classified reservoirs.

The difference between these two models is that model A uses information of reservoirs that were digitized from a topographic map as adopted from Osei et.al[60] whilst model B uses information of reservoirs that was classified from the RapidEye image. The reservoirs are from the same study area but measured separately. This way we can compare their effects on the model.

First, a standard linear regression, i.e. Ordinary least squares (OLS) model as described in chapter 2, section 2.4.1 was fitted for models A and B to determine the linear relationship between the response and the explanatory variables. This was done by running `lm` in R. However the assumption that the error terms in the OLS have a zero mean and are identical and independently distributed are usually violated due to the presence of spatial dependence (autocorrelation). Therefore to include the spatial dependence, the conditional autoregressive model (CAR) as described in chapter 2, section 2.4.1 was used to fit models A and B. This was also implemented in R using the `spautolm` function. The function take into account the neighbors and weights for the autoregression model estimation by Maximum Likelihood arguments.

Chapter 5
Results and analysis

5.1 CLASSIFICATION

After classification, the following landcover were mapped; waterbodies/streams (reservoirs), built up areas, urban centers (heavily built up), forest, dense vegetation, grasslands and wetlands. Wetlands areas were characterized by flooded areas and or rivers with emergent vegetation. For this reason, they were merged with watebodies. Forest was also merged with the dense vegetation class resulting in five landcover classes. The landcover map resulting from the classification of the RapidEye image is shown below in Figure 5.1. The rivers layer obtained from Osei et.al[60] was

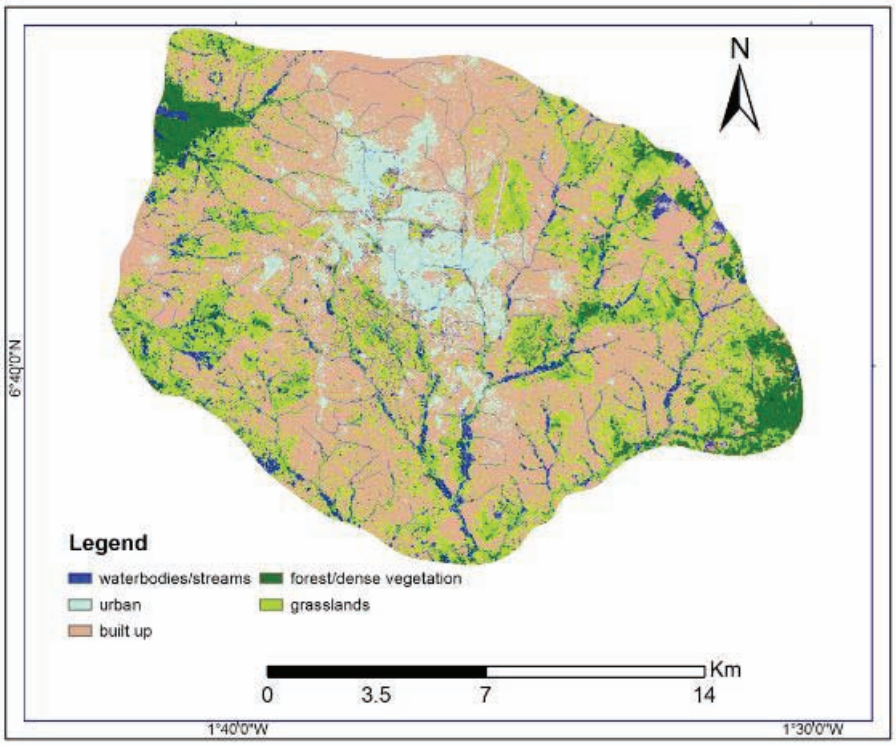


Figure 5.1: Landcover map

used to validate the presence of the rivers on the image. The digitized rivers when overlaid on the classified image lied on the classified streams thereby confirming their presence on the image. Additionally, a random sampling scheme was used to select 85 points (pixels) from the classification output and compared with the reference (true world class). The resulting classification was assessed at 83.53% overall accuracy and an overall Kappa statistics of 0.80. A summary of the accuracy assessment totals are shown in Table 5.1.

The refuse dumps could not be identified on the image. The 124 waste containers placed throughout the city as described in chapter 3(See Figure3.2) are far less than $5m^2$ making them impossible to see in a 5m resolution image. Though the refuse dumps layer adopted from Osei and Duker[59] were overlaid on the image, it could not really be confirmed on the image. This study therefore adopted the refuse dumps data mapped by Osei and Duker[59].

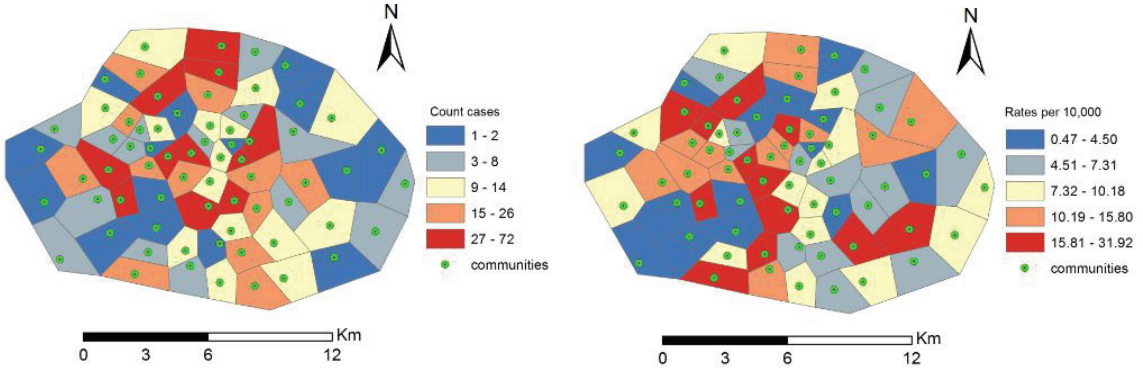
Table 5.1 Accuracy assessment

Overall classification accuracy =83.53%, Kappa statistics = 0.80

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy(%)	Users Accuracy(%)
waterbodies/streams	10	8	8	80.00	100.00
urban	12	9	8	66.67	88.89
built area	23	22	20	86.96	90.91
dense vegetation	7	8	6	85.71	75.00
wetlands	7	10	7	100.00	70.00
forest	12	13	10	83.33	76.92
grasslands	14	15	12	85.71	80.00
Totals	85	85	71		

5.2 MAPPING AND GEOVISUALIZATION

Figure 5.2 show choropleth maps of cholera count cases (5.2a) and raw rates per 10,000 people(5.2b) for each community in the Kumasi Metropolis. Each community is colored according to the category into which its corresponding attribute value falls. Here, communities with red color have the highest counts/rates of cholera incidence; communities with the blue color have the lowest counts/rates. The legend provides us indication of the overall magnitude of cholera incidence and the magnitude of the relative differences in attribute values that correspond to the range of colors used in the map. This map has been able to highlight two extremes; the high areas in red and the low areas in blue as well as highlight the mean areas shown in orange. The community centroids are shown in green dots.



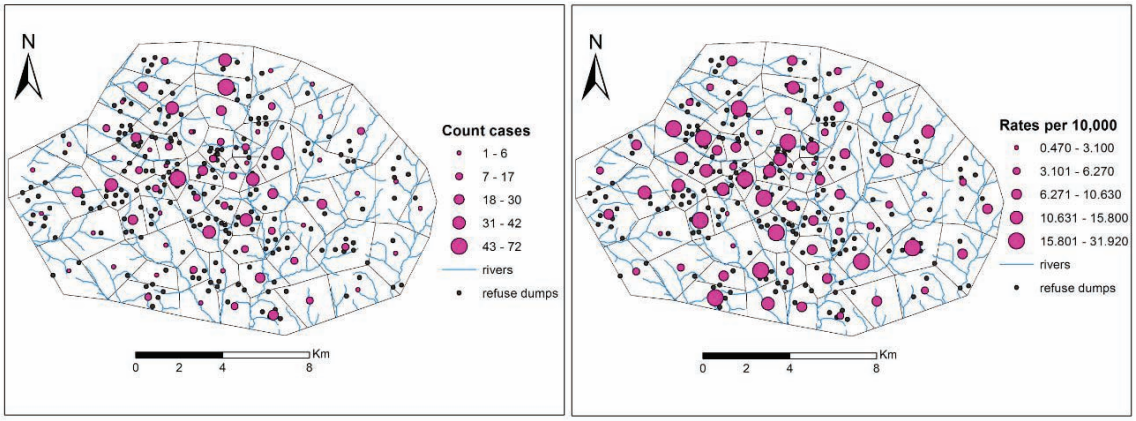
(a) recorded incidence (count cases)

(b) incidence rate per 10,000 people

Figure 5.2: Choropleth maps showing showing cholera prevalence for each community

Figure 5.3 are thematic maps visually depicting how cholera incidence are shown by proportions in each community. In Figure 5.3a, where only the count cases are represented, high cholera cases appear only at few areas in the upper north and at the central parts of Kumasi with minor reported cases at the outer communities. With the rates per 10,000 people, more communities are now showing high risk (Figure 5.3b) than that shown in figure 5.3a. With a proportional symbol map, the symbol size is proportional to the magnitude of the attribute values in each class. In Figure 5.3, the size of each pink circle indicates the relative number of of count cases in (a) and rates per 10,000 people in (b) per community.

Refuse dumps represented by the black dots and river segments in blue were added to this map to visualize and characterize their distribution spatially.



(a) recorded incidence (count cases) (b) incidence rate per 10,000 people

Figure 5.3: Proportional symbol maps showing showing cholera prevalence for each community

The probability map (Figure 5.4) of cholera cases based on Choynowski's[17] approach folds the two tails of the measured probabilities together, so that small values for a chosen α, occur for spatial units (communities) with either unusually high or low rates. For this reason, the high and low communities are plotted separately in Figure 5.4.

The probality map in figure 5.4 shows raw rates (assuming a constant rate across the study area), relative risks, and Poisson probability map values calculated using the standard Poisson cumulative distribution function. This does not fold the tails together, so that communities with lower reported cases than expected, based on population size, have values in the lower tail, and those with higher observed counts than expected have values in the upper tail, as Figure 5.4 shows.

Table 5.2. is a summary statistics of the results returned by running the the `probmap ()` function in R. The function returns a data frame of rates for counts in populations at risk with crude rates, expected counts of cases, relative risks, and Poisson probabilities.

Table 5.2 Summary statistics of probability mapping

	Raw(crude) rate	Expected count	Relative risk	Poisson probabilities
minimum	4.699×10^{-5}	1.354	4.953	0.0000
mean	1.022×10^{-3}	13.956	107.691	0.5155
stan. Dev.	6.895×10^{-4}	13.103	72.684	0.3783
maximum	3.192×10^{-3}	53.521	336.429	1.0000

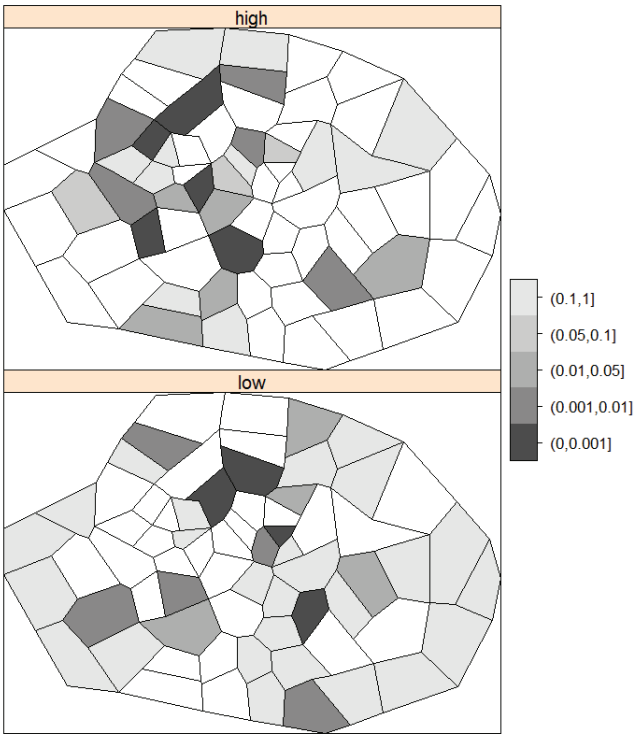
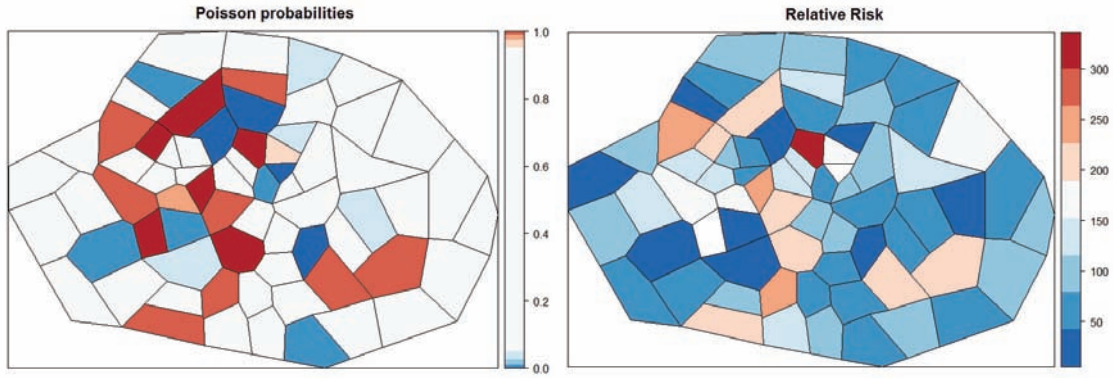


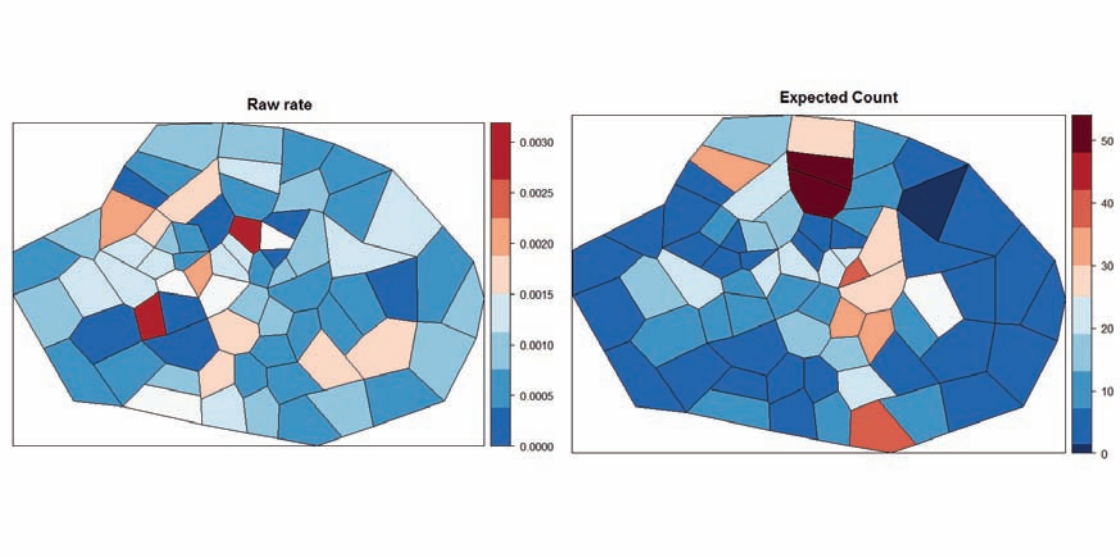
Figure 5.4: Probability map of cholera cases; Choynowski's approach



(a) Poisson probability map (b) Relative risk map

Figure 5.5: Poisson Probability and Relative risk maps of cholera incidence

In Figure 5.5 (a), Poisson probability map values represents the probablility of getting a more "extreme" count than actually observed whilst in Figure 5.5 (b), relative risks shows the ratio of observed and expected counts of cases multiplied by 100.



(a) Raw rates map (b) Expected count map

Figure 5.6: Raw(crude) rates and Expected count maps of cholera incidence

Figure 5.6 (a) is simply the map of crude rates, i.e number of count cases in each community divided by the total population in that community. Figure 5.6 (b) shows the expected counts of cases assuming a global rate.

5.3 SPATIAL ANALYSIS

5.3.1 Autocorrelation analysis

The extent to which neighboring values of cholera cases are correlated was measured using Moran's Index. A significant assessment under randomization procedure was run in R to determine the significance of the computed Moran's Index. There is positive and spatial autocorrelation for cholera incidence in the Kumasi metropolis (Moran's I = 0.138, p = 0.045), as shown in Table 5.3. Table 5.4. shows the standard deviates of Moran's I and a two-sided probability value for 6 lags.

Table 5.3 Moran's Index for spatial autocorrelation of cholera cases

Moran's I	p-value	expectation	variance
0.138	0.045	-0.015	0.006

Table 5.4 Spatial correlogram for CASES

Spatial correlogram for CASES(method: Moran's I)

	estimate	expectation	variance	standard deviate	Pr(I) two sided
1	0.1382	-0.0149	0.0058	2.0083	0.04461 *
2	-0.0454	-0.0149	0.0027	-0.5784	0.56298
3	0.0200	-0.0149	0.0019	0.7918	0.42847
4	-0.0587	-0.0149	0.0018	-1.0310	0.30255
5	-0.0171	-0.0149	0.0019	-0.0494	0.96063
6	-0.1141	-0.0149	0.0031	-1.7757	0.07578 .

Moran's scatter plot and spatial correlogram are also shown in Figure 5.7. The slope of the scatter plot corresponds to the value for Moran's I. It is a measure of global spatial autocorrelation or overall clustering in a dataset. The four quadrants of the scatter plot describe an observation's value in relation to its neighbors; starting with the x-axis, followed by y: High-high, low-low (positive spatial autocorrelation) and high-low, low-high (negative spatial autocorrelation). These quadrants correspond to the clusters and spatial outliers in the LISA maps in Figure 5.8.

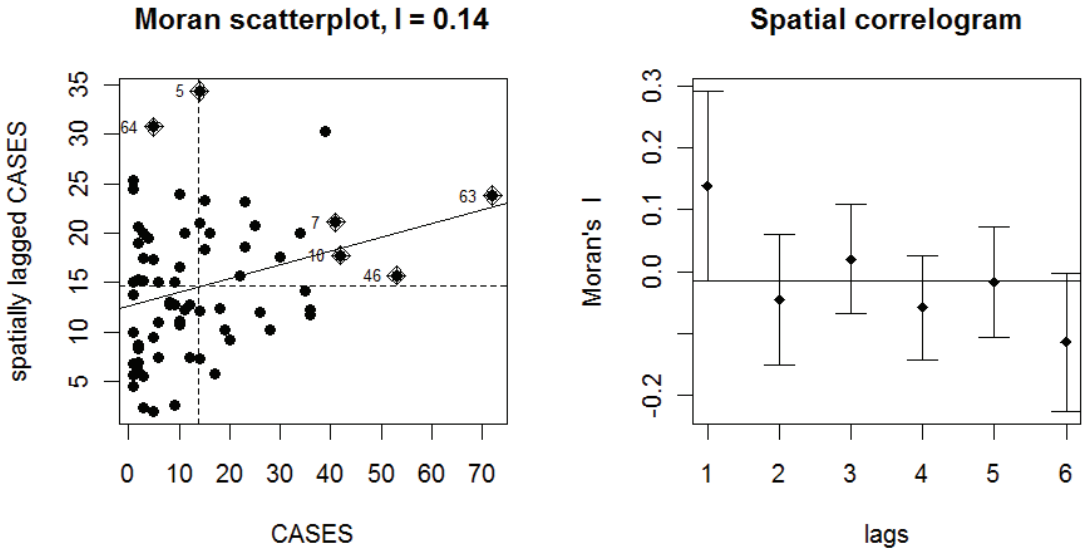
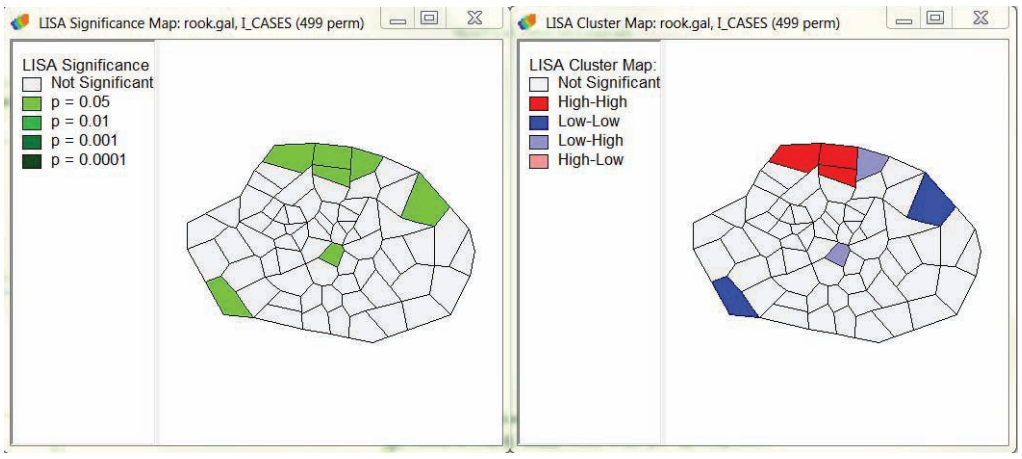


Figure 5.7: Moran scatter and spatial correlogram plots



(a) LISA significance map (b) LISA cluster map

Figure 5.8: LISA significance and cluster maps

The significance map, illustrated in Figure 5.8(a) shows the locations with significant local Moran's I /LISA statistics in different shades of green (the corresponding p values are given in the

legend). Not much significance can be seen from the significance map as only 7 communities has a LISA statistics of p = 0.05. The LISA cluster map, shown in Figure 5.8(b) provides essentially the same information as the significance map but with the significant locations color coded by type of spatial autocorrelation. The four codes are shown in the legend. There is no high-low location in the study area map. These four categories correspond to the four quadrants in the Moran's scatter plot shown in figure 5.7.

5.3.2 Spatial autoregressive modeling

A summary statistics of the four variables used in modelling are shown on Table 5.5. Cholera incidence (count) for the period of study ranged between 1 and 76 recorded cases per community (mean = 13.96 and standard deviation = 14.19).

Table 5.5 Summary statistics of variables used in spatial modeling

variable	minimum	mean	maximum	stan Dev
cases	1.00	13.96	76	14.19
proximity to refuse dumps(ND_RD)(m)	12.28	413.63	1033.6	223.35
density of dumps(Den_RD)(dumps per square km)	0.00	1.83	6.29	1.39
proximity to digitized reservoirs(ND_DR) (m)	56.72	293.72	700.51	137.92
proximity to classified reservoirs(ND_CR) (m)	18.75	309.15	1641.43	239.38

The results of the OLS regression models for Model A(where cholera reservoirs are the digitized rivers adopted from Osei et.al[60] are shown in Table 5.6 and that of Model B(where cholera reservoirs are extracted from RapidEye image) are shown in Table 5.7. Proximity to refuse dumps(ND_RD) in both Models show significant correlation with cholera incidence. The preliminary analysis shows that Density of dumps(Den_RD) is not a significant contributor in both Models(p-Value > 0.6). Also proximity to digitized reservoirs(ND_DR) in Model A is insignificant whilst the classified reservoirs(ND_CR) in Model B shows significant correlation with cholera. However using ND_CR in model B has shown a significant improvement of Model A. (p-Value of Model B < p-Value of Model A).

The results of the CAR models for Models A(Table 5.8) and B(Table 5.9) produce similar results as the OLS models in Tables 5.6 and 5.7. Both OLS and CAR models show that Den_RD is not a significant contributor of cholera incidence(p-Value=0.83 in Table 5.8 and p-Value=0.42 in Table 5.9). For this reason, Models A and B were updated to exclude the Den_RD variable.

Table 5.6 Results of OLS Regression, Model A

Parameter	Estimate	Std. Error	t-Value	p-Value
$\hat{\beta}_0$ (intercept)	28.3491	7.7320	3.666	0.00050 ***
$\hat{\beta}_1$(ND_RD)	-0.0223	0.0108	-2.065	0.04299 *
$\hat{\beta}_2$(Den_RD)	0.6779	1.7122	0.396	0.69348
$\hat{\beta}_3$(ND_DR)	-0.0218	0.0114	-1.907	0.06098 .
Resid. Stan. error	12.69	on 64DF		
Multiple R^2	0.246			
Adjusted R^2	0.211	p-Value	=	3.973×10^{-4}

After updating the OLS models to exclude the Den_RD variable; Both updated Models A (Table 5.10)and B(Table 5.11) still show significant negative relationship for both ND_RD, ND_DR

Table 5.7 Results of OLS Regression, Model B

Parameter	Estimate	Std. Error	t-Value	p-Value
$\hat{\beta}_0$ (intercept)	28.1687	7.2463	3.887	0.00024 ***
$\hat{\beta}_1$(ND_RD)	-0.0218	0.0104	-2.096	0.04007 *
$\hat{\beta}_2$(Den_RD)	0.2474	1.6652	0.149	0.88232
$\hat{\beta}_3$(ND_CR)	-0.0183	0.0064	-2.832	0.00617 **
Resid. Stan. error	12.30	on 64DF		
Multiple R^2	0.292			
Adjusted R^2	0.259	p-Value	$=$	5.696×10^{-5}

Table 5.8 Results of CAR Regression, Model A

Parameter	Estimate	Std. Error	z-Value	p-Value
$\hat{\beta}_0$	24.4031	7.5280	3.2416	0.00118
$\hat{\beta}_1$(ND_RD)	-0.0182	0.0103	-1.7578	0.07878
$\hat{\beta}_2$(Den_RD)	1.3867	1.6667	0.8320	0.40540
$\hat{\beta}_3$(ND_DR)	-0.0194	0.0111	-1.7408	0.08172
LR test value	$= 0.63926$			
$\hat{\lambda} = 0.30089$	$\hat{\sigma^2} = 148.69$	AIC $= 545.82$		

Table 5.9 Results of CAR Regression, Model B

Parameter	Estimate	Std. Error	z-Value	p-Value
$\hat{\beta}_0$	25.7649	7.0933	3.6323	0.00028
$\hat{\beta}_1$(ND_RD)	-0.0191	0.0100	-1.9091	0.05625
$\hat{\beta}_2$(Den_RD)	1.6261	1.6667	0.4214	0.67345
$\hat{\beta}_3$(ND_CR)	0.0062	0.0111	-2.7550	0.00586
LR test value	$= 0.35503$			
$\hat{\lambda} = 0.22493$	$\hat{\sigma^2} = 140.89$	AIC $= 541.83$		

Table 5.10 Results of Updated OLS Regression, Model A

Parameter	Estimate	Std. Error	t-Value	p-Value
$\hat{\beta}_0$ (intercept)	30.9133	4.1957	7.368	$<10^{-6}$ ***
$\hat{\beta}_1$(ND_RD)	-0.0255	0.0070	-3.631	0.00055 ***
$\hat{\beta}_2$(ND_DR)	0.0113	0.0064	-1.916	0.05973 .
Resid. Stan. error	12.61	on 65DF		
Multiple R^2	0.244			
Adjusted R^2	0.221	p-Value	$=$	1.106×10^{-4}

Table 5.11 Results of Updated OLS Regression, Model B

Parameter	Estimate	Std. Error	t-Value	p-Value
$\hat{\beta}_0$ (intercept)	29.1234	3.3286	8.749	$<10^{-6}$ ***
$\hat{\beta}_1$(ND_RD)	-0.0229	0.0069	-3.328	0.00144 **
$\hat{\beta}_2$(ND_CR)	-0.01842	0.0064	-2.877	0.00542 **
Resid. Stan. error	12.21	on 65DF		
Multiple R^2	0.292			
Adjusted R^2	0.270	p-Value	$=$	1.341×10^{-5}

and ND_CR.

The results of the updated CAR Model A (Table 5.12) shows very significant correlation between cholera incidence and ND_RD (p-Value $<$ 0.001) and appear to have no significant effect by the ND_DR variable (p-Value $=$ 0.073). However, the updated CAR Model B results (Table 5.13) shows very significant correlation between cholera incidence and both the ND_RD (p-Value $<$ 0.001) and the ND_CR variables. Comparing λ and the Akaike Information Criterion values(AIC), the updated CAR model B has better fit than the updated CAR model A. λ is the spatial correlation or spatial dependence and measures the strength of the relationship. As expected, an inverse relationship between cholera prevalence and proximities to refuse dumps and classified reservoirs was observed.

Table 5.12 Results of Updated CAR Regression, Model A

Parameter	Estimate	Std. Error	z-Value	p-Value
$\hat{\beta}_0$	29.9900	4.2108	7.1220	$<10^{-6}$
$\hat{\beta}_1$(ND_RD)	-0.0249	0.0068	-3.6316	0.00028
$\hat{\beta}_2$(ND_DR)	-0.0200	0.0111	-1.7891	0.07360
LR test value	$=$ 0.34867			
$\hat{\lambda} = 0.2144$	$\hat{\sigma^2} = 150.49$	AIC $=$ 544.27		

Table 5.13 Results of Updated CAR Regression, Model B

Parameter	Estimate	Std. Error	z-Value	p-Value
$\hat{\beta}_0$	28.5056	3.3414	8.5310	$<10^{-6}$
$\hat{\beta}_1$(ND_RD)	-0.0223	0.0067	-3.3115	0.00092
$\hat{\beta}_2$(ND_CR)	-0.0177	0.0062	-2.8319	0.00462
LR test value	$=$ 0.28033			
$\hat{\lambda} = 0.19442$	$\hat{\sigma^2} = 141.31$	AIC $=$ 539.93		

Chapter 6
Discussion

6.1 IMAGE CLASSIFICATION

The overall accuracy or proportions correctly classified of the landuse/land cover was 83.53% with a Kappa statistics of 0.80. The producer's and user's accuracies of the classified water reservoirs class were assessed to be 80% and 100% respectively. The users accuracy is the probability that a certain reference class has indeed actually been labelled as that class, whilst producers accuracy is the probability that a sampled point on the map is indeed that particular class. The overall, users and producers accuracies are important measures of mapping accuracy[1]. The digitized rivers obtained from Osei et.al[60] when overlaid on the classified image lied on the classified reservoirs thereby confirming their presence on the image. The classified reservoirs however contained additional patches of water bodies(ponds/stagnant water) which are however missing from the digitized rivers from Osei et.al[60]. Therefore, the classification of the water reservoirs is a validation of the presence of the digitized rivers from the data of Osei et.al[60] on the image.

6.2 AUTOCORRELATION ANALYSIS

The extent to which neighboring values are correlated was measured using Global Moran's Index. There is positive and spatial autocorrelation for cholera incidence in the Kumasi metropolis (Moran's $I = 0.138$, $p = 0.045$), as shown in Table 5.3. This explains the spatial clustering of high cholera cases recorded in the central and upper north parts of the Metropolis, and low cases around the peripheries(see Figure 5.2(a)). However, further diagnostics reveal a fluctuation between negative and positive Moran's I as shown in Table 5.4 and Figure 5.7. This could explain the co-location of dissimilar values in some few areas, i.e. low values surrounded by high values or vice versa in certain parts of the Metropolis, e.g. the south-west parts as shown in Figure 5.7. Generally the high values occur at the central parts as expected and supports earlier findings by Osei et.al[31]. In their analysis, they observed clustering of high rates around Kumasi Metropolis, with Moran's $I = 0.271$ and $p < 0.01$. The difference between this study and that of Osei et.al[31] is that, they analyzed the entire Ashanti region consisting of 18 districts as spatial units for cholera incidence from a period of 1997 to 2001, whilst this study only focussed on the Kumasi Metropolis(one of the 18 districts) with 68 communities as spatial units. Unlike the districts which have defined boundaries, the communities did not. The boundaries of the communities were estimated from the RapidEye image with the help of Google Earth image. These boundaries which describes the neighborhood structure used in the analysis could also explain for the low p-Value obtained for the Kumasi Metropolis.

The reasons for the clusters at the central parts being that the central parts are the most commercialized areas and therefore, there is always a high daily influx of people from neighboring communities. This high daily influx puts strains on existing sanitation systems, hence increasing the risk of cholera transmission. Also due to overcrowding and high cost of housing, the migrants end up settling in slummy areas where environmental sanitation is poor[31].

6.3 SPATIAL AUTOREGRESSIVE MODELING

In the results of the autoregressive models; updated CAR Model A (Table 5.12) shows very significant correlation between cholera incidence and proximity to refuse dumps (p-Value < 0.001) and appear to have no significant effect by the proximity to digitized reservoirs (p-Value $= 0.073$. i.e. $p > 0.05$). However, the updated CAR Model A(CAR A) results (Table 5.13) shows very significant correlation between cholera incidence with both proximity to refuse dumps (p-Value $<$ 0.001) and proximity to classified reservoirs(p-Value < 0.01). This suggests a very high significant inverse relationship between cholera incidence with the classified reservoirs as expected. The AIC criteria are smaller for the updated CAR Model B(CAR B) indicating a better fit. The λ values;($\lambda = 0.2144$ for updated CAR Model A and $\lambda = 0.19442$ for updated CAR Model B) measures the strength of the autocorrelation(spatial dependence).

These findings confirms the research of Osei et.al[59, 60] which concluded that proximity to refuse dumps and surface water influences the risk of cholera in the Kumasi Metropolis. The reason being that, during outbreaks, runoff from open spaced dumps as a result of rains or flooding may serve as major pathways for the distribution of the bacteria. Infected human excreta washed away from these dump sites run into nearby wells, streams and surface water bodies thereby contaminating the water which can lead to cholera infection when used.

Comparing the results of CAR A to that of CAR B, we see how using information extracted from a remote sensing image(RapidEye) affects the results and conclusions. The classified reservoirs however contained additional patches of water bodies(ponds/stagnant water) which are however missing from the digitized rivers from Osei et.al[60]. This reveals the importance of using satellite imagery to detect environmental factors (cholera reservoirs). Satellite images are available in various formats and provide high spatial and temporal coverage of the earth's surface. The combined use of remote sensing, GIS and spatial statistics provides a strong tool for assessing how land cover variables relate to cholera incidence.

6.4 REMOTE SENSING, GIS AND SPATIAL STATISTICS IN HEALTH STUDIES

Satellite images can greatly enhance mapping of the environmental factors associated with disease risk. With increasing higher spatial and spectral resolutions, more frequent coverage, lower price, and increased availability of images offered by a wide range of new sensors[10], health researchers should be able to extract many more environmental variables associated with disease transmissions. These will provide new opportunities to extend the uses of remote sensing technology beyond a few vector and water borne diseases to studies of vegetation and climatic related diseases.

GIS plays a central role for providing a platform that allows us to store, integrate, manipulate, visualize and perform exploratory analysis on data from various diverse sources[1]. It's powerful capabilities allows us to integrate data represented in different formats e.g., raster or image data, vector or point, line and area data, at different spatial resolutions. GIS is an important tool used to show the distribution of diseases within a community as well as highlight high spots where health officials need to take prompt action. Basically a GIS should also be able to answer to "what?, where? and when?" diseases occur.

In analyzing disease data (public health data), several disciplines come into play than just maps and visual interpretations[78]. Medical science may only provide insights to some specific causes of diseases (e.g., biological mechanisms of transmission and identification of infectious agents). Moreover, not everybody exposed to a suspected cause contracts the disease. As such, it is important to use statistical methods to analyze each person's risk or probability of contracting a disease. The main aim is to identify and quantify any exposures, behaviors, and characteristics that may

modify a person's risk. Waller and Gotway[78] identifies four (4) usefulness why statistical and spatial statistical methods should be employed to analyze public health data; (1) to evaluate differences in rates observed from different geographic areas, (2) to separate pattern from noise, (3) identify disease "clusters," and (4) assess the significance of potential exposures. With spatial statistical methods, we are also able to quantify uncertainty in our estimates, predictions, and maps and provide the foundations for statistical inference with spatial public health data[78].

The general important message is that in addressing problems in health management for infectious diseases, there exist a number of promising tools available for use. A combination of remote sensing, geographical information systems (GIS) and spatial analysis provide important tools that should be exploited in the fight against diseases.

6.5 LIMITATIONS OF THE STUDY

The limitations in this research include the following;

- The communities do not have defined boundaries. The boundaries of the communities were estimated from the RapidEye image with the help of Google Earth image. This may invariably affects the results of spatial autocorrelation since autocorrelation varies with the neighborhood structures defined. As such the spatial autocorrelation results should be interpreted with caution due to different shape and sizes of the communities.
- The cholera data used is only for a single year(2005) outbreak. Additional data from several outbreaks could enhance the analysis.
- The cholera data are count data aggregated at community levels. This does not contain spatial information about the exact locations of affected individuals. The assumption made was that the population within a community has equal risk.

Chapter 7
Conclusions and recommendations

7.1 CONCLUSIONS

This study utilized remote sensing image(RapidEye) to capture potential cholera reservoirs. Two statistical models were developed to determine the spatial dependency of cholera incidence on; (1) proximity to digitized reservoirs from a topographic map and refuse dumps and (2) proximity to classified reservoirs from a RapidEye image and refuse dumps. The findings reveal a high significant association between cholera cases and proximity to classified reservoirs from a RapidEye image and refuse dumps than between the proximity to digitized reservoirs from a topographic map and refuse dumps. Maps were produced to characterize the distribution of cholera prevalence and risk in the Kumasi metropolis. The specific objectives and research questions in this study are addressed below.

Objective 1 - To map out potential cholera causing factors in the study area from a RapidEye image.

- Which environmental factors relevant to cholera modeling and mapping can be extracted from a RapidEye image?
  - Water reservoirs, consisting of rivers, streams, ponds, lakes and wetlands.
- How can these environmental factors be extracted and what is their quality?
  - A pixel based approach (supervised classification) using the maximum likelihood algorithm. The classification was done using the ERDAS IMAGINE 2011 software.
  - The overall accuracy or proportions correctly classified of the landuse/land cover was 83.53% with a Kappa statistics of 0.80. The producer's and user's accuracies of the classified water reservoirs class were assessed to be 80% and 100% respectively.

Objective 2 - Visualize the relations between cholera incidence, water bodies and refuse dumps using a GIS

- How can the derived remote sensing variables be combined with field data to produce maps of environmental factors?
  - By Integrating all datasets on a common platform in a GIS environment(ArcMap 10 software was used) with their attributes and applying mapping tools (See Figures 5.1 and 5.3).
- How can maps of cholera risks be visualized in a GIS?
  - By computing the risks using spatial analyst tools and visualizing in GIS (See Figure 5.2). Probabilities and risks maps can also be also be computed and plotted in R using the `probmap ()` function (See Figures 5.4;5.5 and 5.6).

Objective 3 - Determine the spatial relationship between cholera incidence and potential cholera reservoirs and refuse dumps using spatial statistics

- Which models are most effective for modeling the effects of environmental risk factors on cholera?
  - Spatial autoregressive models such as the Conditional autoregressive model applied in this study.
  - There is positive and spatial autocorrelation for cholera incidence in the Kumasi metropolis (Moran's $I = 0.138$, $p = 0.045$), as shown in Table 5.3.
  - There are very significant and inverse relationship between cholera incidence and proximities to both refuse dumps (p-Value < 0.001) and reservoirs(p-Value < 0.01).

7.2 RECOMMENDATIONS

Remote sensing data should be explored as a tool for monitoring the epidemiology and control of infectious diseases. Research is needed to determine the appropriate level of geographic detail for specific diseases.

In this study, refuse dumps could not be identified from the remote sensing image, because the refuse dump bins were too small to be identify from a 5m resolution image. Further studies should use a higher resolution image to map the refuse dumps, this will enhance the analysis. Other land cover variables should also be explored to see whether if it has an effect on the disease.

The data used in this study are based on group of individuals at the community level. The inclusion of detailed house hold data or a series of outbreaks over a certain time will be useful to make inference on relatively smaller groups of individuals and/or spatio-temporal analysis.

Appendix A

R codes used

```
#########################################################
#Importing cholera data shapefile  to R
#########################################################
library(maptools)
library(sp)
library(RColorBrewer)
library(spdep)
library(rgdal)
getinfo.shape("D:/Spatial_Analysis/chodata.shp")
cho<-readShapePoly ("D:/Spatial_Analysis/chodata.shp")
class(cho)
attach(cho)

#########################################################
# create rook contiguity neighbors
#########################################################
 cho_nbr<-poly2nb(cho, queen=FALSE) # rook neighborhood
 cho_nbr

#########################################################
# plot of rook neighborhood structure
#########################################################
 coords<-coordinates(cho)
 plot(cho, border="black",axes=TRUE)
 plot(cho_nbr, coords, add=T, col="blue")
 title(main="First order Rook contiguity neighbours")

#########################################################
# neighbors list to weight list
#########################################################
 cho_nbr_w<- nb2listw(cho_nbr)  #Row standardized weights W
 cho_nbr_w                      #Row standardized weights W

 cho_nbr_wb<- nb2listw(cho_nbr, style="B") #Binary weights B
 cho_nbr_wb

#########################################################
 # Moran's I test under normality
 #########################################################
```

```
moran.test(cho$CASES,cho_nbr_w,randomisation=FALSE,
alternative="two.sided")  # W
moran.test(cho$CASES,cho_nbr_wb,randomisation=FALSE,
alternative="two.sided") # B

#########################################################
 # Moran's I test under randomisation
 #########################################################
moran.test(cho$CASES, listw=cho_nbr_w, alternative="two.sided") # W
moran.test(cho$CASES, listw=cho_nbr_wb, alternative="two.sided")# B

#########################################################
 # 999 Monte-Carlo simulation of Moran's I
#########################################################
 morpermCASES <- moran.mc(cho$CASES,cho_nbr_w,999)        # W
 morpermCASES <- moran.mc(cho$CASES,cho_nbr_wb,999)       # B
 morpermCASES
 morpermCASES$res



#########################################################
#Moran scatter and spatial correlogram plots
#########################################################
 par(mfrow=c(1,2))
 moran.plot(cho$CASES,cho_nbr_w, spChk=NULL, labels=NULL,
 xlab="CASES", ylab="spatially lagged CASES", quiet=NULL,
 pch=19,main="Moran scatterplot,
 I = 0.14, p=0.04")
 plot(spco, main="Spatial correlogram")
#########################################################
# PROBABILITY MAPS

#choynowski's approach
#########################################################
ch <- choynowski(cho$CASES, cho$POP)
cho$ch_pmap_low <- ifelse(ch$type, ch$pmap, NA)
cho$ch_pmap_high <- ifelse(!ch$type, ch$pmap, NA)
prbs <- c(0,.001,.01,.05,.1,1)
cho$high = cut(cho$ch_pmap_high, prbs)
cho$low = cut(cho$ch_pmap_low,prbs )
spplot(cho, c("low", "high"),
col.regions=grey.colors(5))

# Poisson probabilities
#################################################
pmap <- probmap(cho$CASES, cho$POP)
cho$pmap <- pmap$pmap
```

```
brks <- c(0,0.001,0.01,0.025,0.05,0.95,0.975,0.99,0.999,1)
library(RColorBrewer)
spplot(cho, "pmap", at=brks,
col.regions=grey.colors(5))
spplot(cho, "pmap", at=brks,
main="Poisson probabilities",
col.regions=rev(brewer.pal(9, "RdBu")))

#Relative Risk
############################################################
relRisk <- probmap(cho$CASES, cho$POP)
cho$relRisk <- relRisk$relRisk
brks <- c(5,42,79, 115,152,189,226,263,300,336)
library(RColorBrewer)
#spplot(cho, "pmap", at=brks,
col.regions=grey.colors(5))
spplot(cho, "relRisk", at=brks,
main="Relative Risk" ,
axes=TRUE, col.regions=rev(brewer.pal
(9, "RdBu")))
hist(cho$relRisk, main="")

 # Map of Raw(crude rates)
############################################################
raw <- probmap(cho$CASES, cho$POP)
cho$raw <- raw$raw
brks <- c(0,0.000375,0.00075,0.001125,0.0015, 0.001875,
0.00225,0.002625,0.003192)
library(RColorBrewer)
#spplot(cho, "pmap", at=brks, col.regions
=grey.colors(5))
spplot(cho, "raw", at=brks,
main="Raw rate" , axes=TRUE, col.regions
=rev(brewer.pal(9, "RdBu")))
hist(cho$raw, main="")

# Map of Expected Count
############################################################
expCount <- probmap(cho$CASES, cho$POP)
cho$expCount <- expCount$expCount
brks <- c(0,1.4,7.2,13,19,25,30,36,42,48,54)
library(RColorBrewer)
#spplot(cho, "pmap", at=brks,
col.regions=grey.colors(5))
spplot(cho, "expCount", at=brks,
main="Expected Count" ,
 axes=TRUE, col.regions=rev(brewer.pal
 (11, "RdBu")))
```

```
hist(cho$expCount, main="")
############################################################
# OLS Models

OLSmodela <- lm(CASES ~ ND_RD + Den_RD + ND_DR, data=cho)
OLSmodelb <- lm(CASES ~ ND_RD + Den_RD + ND_CR, data=cho)
summary(OLSmodela)
summary(OLSmodelb)
############################################################
OLSmodelA<- lm(CASES ~ ND_RD  + ND_DR, data=cho)
OLSmodelB <- lm(CASES ~ ND_RD  + ND_CR, data=cho)
summary(OLSmodelA)
summary(OLSmodelB)
############################################################
##CAR Models
car1 <- spautolm(CASES ~ ND_RD + Den_RD + ND_DR, data=cho,
listw=cho_nbr_w, family="CAR", method="full", verbose=TRUE)
car2 <- spautolm(CASES ~ ND_RD + Den_RD + ND_CR, data=cho,
listw=cho_nbr_w, family="CAR", method="full", verbose=TRUE)
CARmodelA <- spautolm(CASES ~ ND_RD  + ND_DR, data=cho,
listw=cho_nbr_w, family="CAR", method="full", verbose=TRUE)
CARmodelB <- spautolm(CASES ~ ND_RD  + ND_CR, data=cho,
listw=cho_nbr_w, family="CAR", method="full", verbose=TRUE)
summary(car1)
summary(car2)
summary(CARmodelA)
summary(CARmodelB)
############################################################
```

Bibliography

[1] *GI Science and Earth Observation : a process - based approach : also as e-book*. ITC Educational Textbook Series. University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede, 2010.

[2] Paul A. and Manning. Vibrio cholerae, infection and immunity. In *Encyclopedia of Immunology (Second Edition)*, pages 2476 – 2479. Elsevier, Oxford, second edition edition, 1998.

[3] M. Ali, M. Emch, J.P. Donnay, M. Yunus, and RB Sack. Identifying environmental risk factors for endemic cholera: a raster GIS approach. *Health & place*, 8(3):201–210, 2002.

[4] L. Anselin. Local Indicators of Spatial Association (LISA). *Geographical analysis*, 27(2):93–115, 1995.

[5] L. Anselin. An introduction to spatial autocorrelation analysis with GeoDa. *Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois*, 2003.

[6] L. Anselin. GeoDa 0.9 User's Guide. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, IL, 2003.

[7] L. Anselin, I. Syabri, and Y. Kho. Geoda: An introduction to spatial data analysis. *Geographical Analysis*, 38(1):5–22, 2006.

[8] M. Asase, E.K. Yanful, M. Mensah, J. Stanford, and S. Amponsah. Comparison of municipal solid waste management systems in Canada and Ghana: A case study of the cities of London, Ontario, and Kumasi, Ghana. *Waste Management*, 29(10):2779–2786, 2009.

[9] T.C. Bailey and A.C. Gatrell. *Interactive spatial data analysis*. Longman Scientific & Technical Essex, 1995.

[10] L.R. Beck, B.M. Lobitz, and B.L. Wood. Remote sensing and human health: new sensors and new opportunities. *Emerging infectious diseases*, 6(3):217, 2000.

[11] Roger Bivand, with contributions by Micah Altman, Luc Anselin, Renato Assunção, Olaf Berke, Andrew Bernat, Guillaume Blanchet, Eric Blankmeyer, Marilia Carvalho, Bjarke Christensen, Yongwan Chun, Carsten Dormann, Stéphane Dray, Rein Halbersma, Elias Krainski, Pierre Legendre, Nicholas Lewin-Koh, Hongfei Li, Jielai Ma, Giovanni Millo, Werner Mueller, Hisaji Ono, Pedro Peres-Neto, Gianfranco Piras, Markus Reder, Michael Tiefelsdorf, , and Danlin Yu. *spdep: Spatial dependence: weighting schemes, statistics and models*, 2011. R package version 0.5-40.

[12] R.S. Bivand, E.J. Pebesma, V. Gómez-Rubio, and Inc NetLibrary. *Applied spatial data analysis with R*, volume 747248717. Springer New York, 2008.

[13] Nathan S Blow, Robert N Salomon, Kerry Garrity, Isabelle Reveillaud, Alan Kopin, F. Rob Jackson, and Paula I Watnick. Vibrio cholerae infection of drosophila melanogaster mimics the human disease cholera. *PLoS Pathog*, 1:e8, 09 2005.

[14] R.J. Borroto and R. Martinez-Piedra. Geographical patterns of cholera in Mexico, 1991–1996. *International journal of epidemiology*, 29(4):764, 2000.

[15] M. Bradley, R. Shakespeare, A. Ruwende, M. E. J. Woolhouse, E. Mason, and A. Munatsi. Epidemiological features of epidemic cholera (El Tor) in Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 90(4):378–382, 1996.

[16] R. Campbell McIntyre, T. Tira, T. Flood, and P.A. Blake. Modes of transmission of cholera in a newly infected population on an atoll: implications for control measures. *The Lancet*, 313(8111):311–314, 1979.

[17] M. Choynowski. Maps based on probabilities. *Journal of the American Statistical Association*, pages 385–388, 1959.

[18] Deirdre L. Church. Major factors affecting the emergence and re-emergence of infectious diseases. *Clinics in Laboratory Medicine*, 24(3):559–586, 2004.

[19] C. Codeço. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC infectious diseases*, 1(1):1, 2001.

[20] A.E. Collins, ME Lucas, MS Islam, and LE Williams. Socio-economic and environmental origins of cholera epidemics in Mozambique: guidelines for tackling uncertainty in infectious disease prevention and control. *International journal of environmental studies*, 63(5):537–549, 2006.

[21] Rita Colwell. Global climate and infectious disease: The cholera paradigm. *Science*, 274(5295):2025–2031, 1996.

[22] BC Deb, BK Sircar, PG Sengupta, SP De, SK Mondal, DN Gupta, NC Saha, S. Ghosh, U. Mitra, and SC Pal. Studies on interventions to prevent eltor cholera transmission in urban slums. *Bulletin of the World Health Organization*, 64(1):127, 1986.

[23] P. Elliott, JC Wakefield, NG Best, and DJ Briggs. Spatial epidemiology: methods and applications. *Spatial epidemiology*, 1(9):3–15, 2001.

[24] P. Elliott and D. Wartenberg. Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, 112(9):998, 2004.

[25] M. Emch. Diarrheal disease risk in Matlab, Bangladesh. *Social Science & Medicine*, 49(4):519–530, 1999.

[26] O. Felsenfeld. Notes on food, beverages and fomites contaminated with vibrio cholerae. *Bulletin of the World Health Organization*, 33(5):725, 1965.

[27] Manfred M. Fischer, Jinfeng Wang, Manfred M. Fischer, and Jinfeng Wang. Modelling area data. In *Spatial Data Analysis*, SpringerBriefs in Regional Science, pages 31–44. Springer Berlin Heidelberg, 2011.

[28] M.M. Fischer and A. Getis. *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Verlag, 2010.

[29] G. Fleming, M. Merwe, and G. McFerren. Fuzzy expert systems and GIS for cholera health risk prediction in Southern Africa. *Environmental Modelling & Software*, 22(4):442–448, 2007.

[30] A.S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons Inc, 2002.

[31] O. Frank and D. Alfred. Spatial and demographic patterns of cholera in Ashanti region-Ghana. *International Journal of Health Geographics*, 7, 2008.

[32] A.R. Ghosh, H. Koley, D. De, S. Garg, MK Bhattacharya, SK Bhattacharya, B. Manna, G.B. Nair, T. Shimada, T. Takeda, et al. Incidence and toxigenicity of Vibrio cholerae in a freshwater lake during the epidemic of cholera caused by serogroup O139 Bengal in Calcutta, India. *FEMS microbiology ecology*, 14(4):285–291, 1994.

[33] Peng Gong, Bing Xu, and Song Liang. Remote sensing and geographic information systems in the spatial temporal dynamics modeling of infectious diseases. *Science in China Series C: Life Sciences*, 49(6):573–582, 2006.

[34] A. J. Graham, P. M. Atkinson, and F. M. Danson. Spatial analysis for epidemiology. *Acta Tropica*, 91(3):219–225, 2004.

[35] Margaret A. Hamburg. Considerations for infectious disease research and practice. *Technology in Society*, 30(3-4):383–387, 2008.

[36] D.M. Hartley, J.G. Morris Jr, and D.L. Smith. Hyperinfectivity: A critical element in the ability of v. cholerae to cause epidemics? *PLoS Medicine*, 3(1):e7, 2005.

[37] V. Herbreteau, G. Salem, M. Souris, J.P. Hugot, and J.P. Gonzalez. Thirty years of use and improvement of remote sensing, applied to epidemiology: From early promises to lasting frustration. *Health & Place*, 13(2):400–403, 2007.

[38] S.D. Holmberg, D.E. Kay, R.D. Parker, et al. Foodborne transmission of cholera in Micronesian households. *The Lancet*, 323(8372):325–328, 1984.

[39] M. Hugh-Jones. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitology Today*, 5(8):244–251, 1989.

[40] A. Huq, R.B. Sack, A. Nizam, I.M. Longini, G.B. Nair, A. Ali, J.G. Morris Jr, MN Khan, A. Siddique, M. Yunus, et al. Critical factors influencing the occurrence of vibrio cholerae in the environment of Bangladesh. *Applied and Environmental Microbiology*, 71(8):4645, 2005.

[41] E.O. Igbinosa and A.I. Okoh. Emerging vibrio species: an unending threat to public health in developing countries. *Research in microbiology*, 159(7-8):495–506, 2008.

[42] M. S. Islam. The aquatic flora and fauna as reservoirs of vibrio cholerae: a review. *Journal of diarrhoeal diseases research*, 12(2):87–96, 1994.

[43] Timothy H. Keitt, Roger Bivand, Edzer Pebesma, and Barry Rowlingson. *rgdal: Bindings for the Geospatial Data Abstraction Library*, 2011. R package version 0.7-5.

[44] M. Kulldorff. A spatial scan statistic. *Communications in statistics-theory and methods*, 26(6):1481–1496, 1997.

[45] K.M. Kwofie. A spatio-temporal analysis of cholera diffusion in Western Africa. *Economic Geography*, pages 127–135, 1976.

[46] A.B. Lawson and D.G.T. Denison. *Spatial cluster modelling*. CRC press, 2002.

[47] Kelley Lee. The global dimensions of cholera. *Global Change & Human Health*, 2(1):6–17, 2001.

[48] Myron M. Levine, Debasish Saha, A. S. G. Faruque, and Samba O. Sow. *Cholera Infections*, pages 150–156. W.B. Saunders, Edinburgh, 2011.

[49] Nicholas J. Lewin-Koh, Eric Archer Adrian Baddeley Hans-Jórg Bibiko Jonathan Callahan Stéphane Dray David Forrest Michael Friendly Patrick Giraudoux Duncan Golicher Virgilio Gómez Rubio Patrick Hausmann Karl Ove Hufthammer Thomas Jagger Sebastian P. Luque Don MacQueen Andrew Niccolai Tom Short Greg Snow Ben Stabler Roger Bivand, contributions by Edzer J. Pebesma, and Rolf Turner. *maptools: Tools for reading and handling spatial objects*, 2011. R package version 0.8-10.

[50] B. Lobitz, L. Beck, A. Huq, B. Wood, G. Fuchs, ASG Faruque, and R. Colwell. Climate and infectious disease: use of remote sensing for detection of vibrio cholerae by indirect measurement. *Proceedings of the National Academy of Sciences*, 97(4):1438, 2000.

[51] F.J. Luquero, C.N. Banga, D. Remartínez, P.P. Palma, E. Baron, and R.F. Grais. Cholera Epidemic in Guinea-Bissau (2008): The Importance of "Place". *PloS one*, 6(5):e19005, 2011.

[52] S. Mandal, M.D. Mandal, and N.K. Pal. Cholera: A great global concern. *Asian Pacific Journal of Tropical Medicine*, 4(7):573–580, 2011.

[53] A. Maroko, J.A. Maantay, and K. Grady. Using geovisualization and geospatial analysis to explore respiratory disease and environmental health justice in New York city. *Geospatial Analysis of Environmental Health*, pages 39–66, 2011.

[54] J. D. Mayer. *Emerging Diseases: Overview*, pages 321–332. Academic Press, Oxford, 2008.

[55] J. Mendelsohn and T. Dawson. Climate and cholera in KwaZulu-Natal, South Africa: the role of environmental factors and implications for epidemic preparedness. *International journal of hygiene and environmental health*, 211(1-2):156–162, 2008.

[56] I. Mugoya, S. Kariuki, T. Galgalo, C. Njuguna, J. Omollo, J. Njoroge, R. Kalani, C. Nzioka, C. Tetteh, S. Bedno, et al. Rapid spread of Vibrio cholerae O1 throughout Kenya, 2005. *The American journal of tropical medicine and hygiene*, 78(3):527, 2008.

[57] E.J. Nelson, J.B. Harris, J.G. Morris, S.B. Calderwood, and A. Camilli. Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Reviews Microbiology*, 7(10):693–702, 2009.

[58] Erich Neuwirth. *RColorBrewer: ColorBrewer palettes*, 2011. R package version 1.0-5.

[59] F. Osei and A. Duker. Spatial dependency of v. cholera prevalence on open space refuse dumps in Kumasi, Ghana: a spatial statistical modelling. *International Journal of Health Geographics*, 7(1):62, 2008.

[60] F.B. Osei, A.A. Duker, E.W. Augustijn, and A. Stein. Spatial dependency of cholera prevalence on potential cholera reservoirs in an urban area, Kumasi, Ghana. *International Journal of Applied Earth Observation and Geoinformation*, 12(5):331–339, 2010.

[61] Frank B. Osei. *Spatial statistics of epidemic data : the case of cholera epidemiology in Ghana*. PhD thesis, 2010.

[62] Dirk Pfeiffer. *Spatial analysis in epidemiology*. Oxford University Press, GB, 2008.

[63] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[64] RapidEye. RapidEye: Dilivering the world. `http://www.rapideye.net/about/index.htm`, 2011. [Online; accessed 29-November-2011].

[65] RapidEye. Satellite Imagery Product Specifications. `http://www.rapideye.net/upload/RE_Product_Specifications_ENG.pdf`, 2011. [Online; accessed 3-January-2012].

[66] RapidEye. High resolution satellite imagery ortho product (level3A). `http://www.rapideye.net/products/index.htm`, 2012. [Online; accessed 3-January-2012].

[67] Joachim Reidl and Karl E. Klose. Vibrio cholerae and cholera: out of the water and into the host. *FEMS Microbiology Reviews*, 26(2):125–139, 2002.

[68] T. P. Robinson. *Spatial statistics and geographical information systems in epidemiology and public health*, volume Volume 47, pages 81–128. Academic Press, 2000.

[69] David A. Sack, R. Bradley Sack, G. Balakrish Nair, and A. K. Siddique. Cholera. *The Lancet*, 363(9404):223–233, 2004.

[70] H. H. Shugart and Robert E. Shope. *Factors Influencing Geographic Distribution and Incidence of Tropical Infectious Diseases*, pages 13–18. Churchill Livingstone, Philadelphia, 2006.

[71] AK Siddiquei, K. Zamani, AH Baqui, K. Akrami, P. Mutsuddy, A. Eusof, K. Haider, S. Islam, and RB Sacku. Cholera epidemics in Bangladesh: 1985-1991. *Diarrhoeal Diseases Research*, 10(2):79, 1992.

[72] J. Snow. *On the mode of communication of cholera*. John Churchill, 1855.

[73] A. Talavera and E.M. Pérez. Is cholera disease associated with poverty? *The Journal of Infection in Developing Countries*, 3(06):408–411, 2009.

[74] W.R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46:234–240, 1970.

[75] Juliana Useya. Simulating diffusion of cholera in Ghana. Master's thesis, Faculty of Geo-information Science and Earth Observation of the University of Twente, 2011.

[76] Francis A. Waldvogel. Infectious diseases in the 21st century: old challenges and new opportunities. *International Journal of Infectious Diseases*, 8(1):5–12, 2004.

[77] M.M. Wall. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324, 2004.

[78] L.A. Waller and C.A. Gotway. *Applied spatial statistics for public health data*, volume 368. Wiley-Interscience, 2004.

[79] SD Walter. Disease mapping: a historical perspective. *Spatial epidemiology: methods and applications*, pages 223–239, 2000.

[80] WHO. Features: infectious diseases. `http://www.who.int/features/infectious_diseases/en/index.html`, 2011. [Online; accessed 5-August-2011].

[81] WHO. Weekly epidemiological record. http://www.who.int/wer/2010/wer8531.pdf, 2011. [Online; accessed 5-August-2011].

[82] Wikipedia. Cholera. http://en.wikipedia.org/wiki/Cholera#cite_note-Lancet2004-0, 2011. [Online; accessed 9-November-2011].

[83] Wikipedia. Kumasi. http://en.wikipedia.org/wiki/Kumasi, 2012. [Online; accessed 3-January-2012].

[84] Wikipedia. RapidEye. http://en.wikipedia.org/wiki/RapidEye, 2012. [Online; accessed 3-January-2012].

[85] Emma Wikner. Modelling waste to energy systems in Kumasi, Ghana. Master's thesis, Swedish University of Agricultural Sciences(SLU) in Uppsala, 2009.