SciencePG
Science Publishing Group

Spatial Statistical Analysis in the Classification of Some Seasonal Diseases

Samson Agboola, Mataimaki Benard Joel

Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

Email address:
abuagboola@gmail.com (S. Agboola), manjoello@yahoo.com (M. B. Joel)

Abstract: This paper made used of the ArcGIS environment where a geo-database was created showing the spatial distribution of the hospital, where each hospitals' coordinate where taken using GPS device from Zaria, Kaduna State of Nigeria. The geo-database contained all the information generated with reference to the geographical location where attributes tools was used to query information about the various hospitals in the study area. The average temperature of the each location was taken with the help of Landsat thematic Mapper and compared with the number of patients that visited the hospital with the diseases under study.

Keywords: Spatial, Analysis, ArcGIS, Disease, Mapping, GPS, Patients

1. Introduction

Spatial statistics comprises a set of techniques for describing and modeling spatial data. It extends what the mind and eyes do, instinctively, to assess spatial patterns, distributions, trends, processes and relationships. It helps to achieve better understanding of behavior of geographical phenomena, pinpoints causes of specific geographic patterns and makes decisions with higher level of confidence. Spatial Statistics is the description and analysis of geographic or spatial variation in diseases with respect to demographic, environmental, behavioral, socioeconomic, genetic and infectious risk factors (Elliott and Wartenberg, 2004). The occurrence of 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 (Pfeiffer 2008). Proximity to environmental risk factors is therefore important. Thus, knowledge of the spatial variations of diseases and characterizing its spatial structure is essential for the epidemiologist to understand better the population's interactions with its environment (Frank 2010). Spatial Statistics dates back to the 1800s, when maps of disease rate in different countries began to emerge to characterize the spread and possible cause of the occurrences of diseases such as yellow fever and cholera (SD Walter, 2000). Recent advances in technology now allow not only disease mapping but also the application of spatial statistical methods (Wakefield, *et al.* 2001), (Kulldorff 1997) and Geographical information system (GIS) for further details, consult (Ali, *et al*. 2002) and (Osei, *et al*. 2010).

Bailey and Gatrell (1995) reviewed a number of spatial techniques which they divided into four categories depending upon the type of data for which they are designed; Point pattern data, spatially continuous data, area data and interaction data. According to (Bailey and Gatrell, 1995), Point Pattern statistics are used to analyze the spatial distribution of features which can be modeled in discrete point. Spatial continuous data sometimes termed random field data or more simply geo-statistical data is used to analyze spatially continuous data. In area data which our focus of this research will be, we frequently need to analyze attribute data which refer to polygon (i.e. areas). Methods developed for this purpose include: spatial moving averages, kernel estimation, spatial autocorrelation, spatial correlation and regression. When dealing with rates based on small numbers, there is a risk of extreme in areas with a low population (the small number problem). One response to this which Bailey and Gatrell (1995) did is to map Poison probabilities. In spatial interactions, most methods developed for interactions are based upon the gravity model.

(Tobler 1979), considered the field of spatial statistics

based on the non-independence of observation that is the assumptions that nearby units are in some way associated. This association according (Tobler, 1979) is because of a spatial spillover effect, such as the obvious economic relationship between the city and suburb. (Kandala, *et al.* 2007), in their work on risk factors for Childhood Morbidity in Nigeria used spatial analysis to indicate if there is a decline in childhood vaccination coverage between 1999 and 2003 and to investigate the impact of geographical factors and other important risk factors on diarrhea, cough and fever using geo-addictive Bayesian semiparametic models. In his findings, it was discovered that a higher prevalence of childhood diarrhea, cough, and fever is observed in the northern and eastern states, while lower disease prevalence is observed in the western and southern states. (Toprak and Erdogan 2008), in their work on the distribution of typhoid fever in Turkey, used spatial analysis to explore regional clustering of typhoid fever in Turkey. Their findings showed that spatial analyses and statistic, significantly contribute to the understanding of the epidemiology of diseases. (Kouray, *et al.* 2010), worked on spatial analysis of tuberculosis in an urban West African setting, if there is an evidence of clustering. Their objective was to describe the pattern of tuberculosis (TB) occurrence in Greater Banjul, The Gambia with Geographical Information Systems (GIS) and Spatial Scan Statistics (SaTScan) and to determine whether there is significant TB case clustering. (Basommi 2011) In his work on Malaria epidemiology in Amanse west district in Ghana used spatial analysis in the disease modeling, monitoring, evaluation and providing major intervention for areas at risk. He explored spatial dependency of the malaria risk using Poisson variograms and the risk was used to create surface maps from 2004 to 2009 to identify areas at high risk. He then used Bayesian geostatistical approach to correlate the relationship between the elevation and the disease risk. (Sinkala, *et al*. 2014), worked on Spatial and temporal distributions of outbreaks, assessment of clusters and implications for the control of Foot and mouth disease (FMD) in Zambia. Their objective was to relate the spatio-temporal distribution of FMD cases and what determines their occurrences. A retrospective review of FMD cases in Zambia from 1981 to 2012 was conducted using geographical information systems and the SaTScan software package.

Maps of regional disease rates are potential useful tools in examining spatial patterns of disease and for identifying clusters. Bayes and empirical Bayes approaches to this problem have proven useful in smoothing crude maps of disease rates. In recent years, the model of (Waller 1997) has proven to be very popular. This model includes both spatial autocorrelation and spatial heterogeneity effects. The spatial autocorrelation effect attempts to capture "clustering" effects in the data, while the spatial heterogeneity effect attempts to capture spatially unstructured variation. Thus, in addition to maps, study design and simple statistics were important tools in Snow's analysis. Applying statistical methods in a spatial setting raise several challenges. Geographer and statistician

Waldo Tobler summarized a key component affecting any analysis of spatially referenced data through his widely quoted and paraphrased first law of geography: "Everything is related to everything else, but near things are more related than far things" (W. Tobler 1979). This law succinctly defines the statistical notion of (positive) spatial autocorrelation, in which pairs of observations taken nearby are more alike than those taken farther apart. Many diseases of concern to humans, domestic livestock and wildlife species have strong seasonal components to their transmission or biology. Even periodically emerging outbreaks of Ebola hemorrhagic fever appear to have been clustered around dry periods that followed the rainy season in parts of central Africa (Pinzon and Wilcon 2004). Although exact connections between weather and infectious disease patterns are the subject of much debate (Hay and Shanks 2005), detailed information on annual variation in disease processes is becoming important for tracking outbreaks and understanding how pathogens respond to seasonal variability and long-term climate change (Harvell and Mitchell 2002).

(Anamzui-Ya 2012), in his work utilized remote sensing image (RapidEye) to capture potential cholera reservoirs. His findings reveal a high significant association between cholera cases and proximity to classified reservoirs from a Rapid Eye image and refuse dumps than between the proximity to digitized reservoirs from a topographic map and refuse dumps. Assaana also used maps to characterize the distribution of cholera prevalence and risk in the Kumasi metropolis.

In this study, we use the spatial analysis function of ArcGIS to classify some diseases over a period of ten (10) years (2005–2014). Specifically, the aim of this paper is to compare the relationship between the spatial locations and draw spatial maps.

2. Method

2.1. Study Area

The study area is Zaria, a major city in Kaduna State in Northern Nigeria as well as being a Local Government Area. Formerly known as Zazzau, it was one of the original seven Hausa city-states. However, human settlement predates the rise of Zaria as the region, like some of its neighbors, had a history of sedentary Hausa settlement, with institutional but pre-capitalist market exchange and farming. The 2006 Census population showed that Zaria's population was 408,198. Zaria has a total Area of about $300km^2$ (100 sq mi). The old part of the city, known as Birnin Zaria or Zaria-City, was originally surrounded by walls, which now have been mostly removed. The Emir's palace is located in the old city. In the old city and the adjacent Tudun Wada neighbourhood people typically reside in traditional compounds. These two neighborhoods are predominately occupied by the indigenous Hausa. The neighborhoods of Samaru and SabonGari are predominately occupied by Nigerians of southern origin, such as the Igbo. The largest marketplace is in Sabon-Gari. Other more recent

neighborhoods include: Danmagaji/Wusasa, PZ, Kongo, GRA-Zaria, Hanwa, Bassawa, Lowcost Kofan-Gayan and Shikka. Its coordinates is $11^0 04' N \ 7^0 42' E$.

The coordinates of the the five hospitals are as follows:

1. Jamaa - $N 11^0 09' 32.2" E$

2. Salama Hospiatal - $N 11^0 07' 44.2" E$
3. Almadina Hospital - $N 11^0 07' 17.1" E$
4. Gambo sawaba Hospital - $N 11^0 02' 28.8" E$
5. St Luke Hospital - $N 11^0 04' 33.2" E$
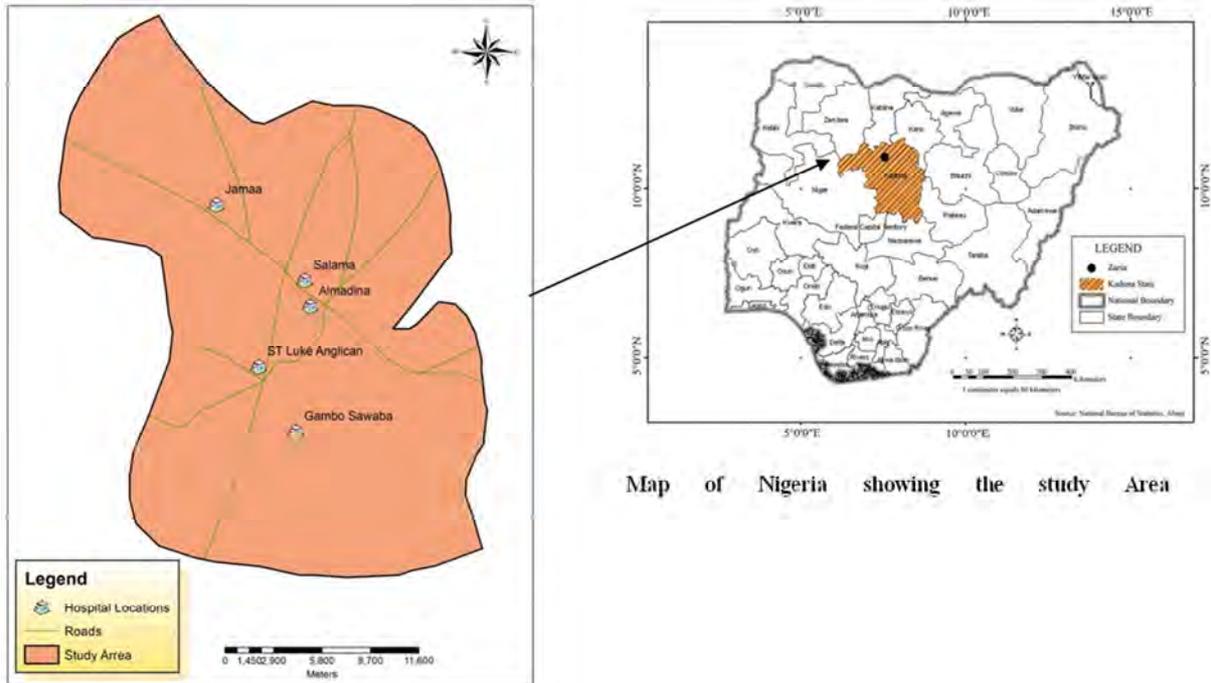
Source: Authors Analysis 2015

Figure 1. Map of the study area showing Spatial Distribution of the Hospitals in the Study Area.
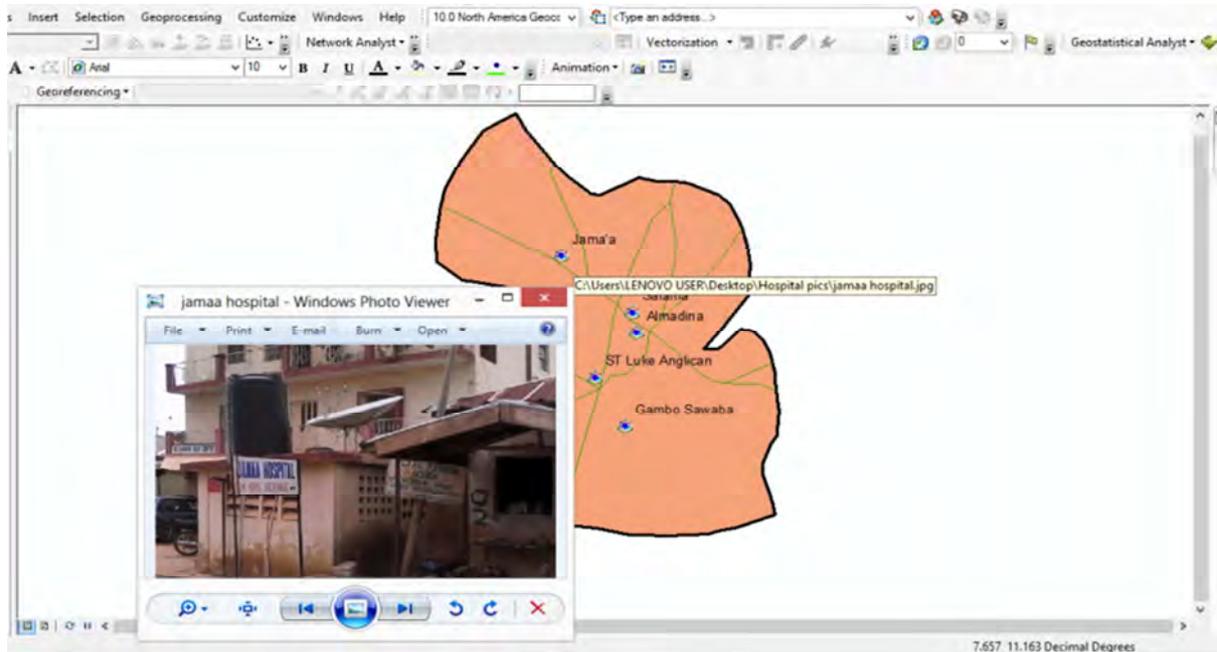
Source: Authors Analysis 2015

Figure 2. Map of the study area depicting the picture location of one of the hospitals (jamaa hospital).

2.2. Data and Methodology

This research is reliant on secondary data. Below is the summary of the data from the various hospitals in the study area.

Table 1. Descriptive Statistics of the data collected.

Sex	Age	Occupation	Disease	Date of admission	location
F:1030	Min: 1.00	CHILD: 757	DIARRHEA: 743	25/5/05: 8	ZARIA CITY: 414
M:1132	1st Qu: 5.00	HOUSE WIFE: 412	MEASLES: 255	29/5/05: 8	HANWA: 322
	Median: 19.00	STUDENT: 407	MENINGITIS: 207	24/5/14: 7	SAMARU: 190
	Mean: 22.61	FARMER: 143	PNUEMONIA: 224	28/3/14: 6	SABO: 161
	3rd Qu: 35.00	BUSINESSMAN: 105	TYPHOID: 733	31/10/14: 6	TUDUN WADA: 158
	Max: 90.00	CIVIL SERVANT: 84		1/5/2014: 5	GESKIYA: 145
		(Other): 254		(Other): 2122	(Other): 772

Source: Authors Analysis 2015
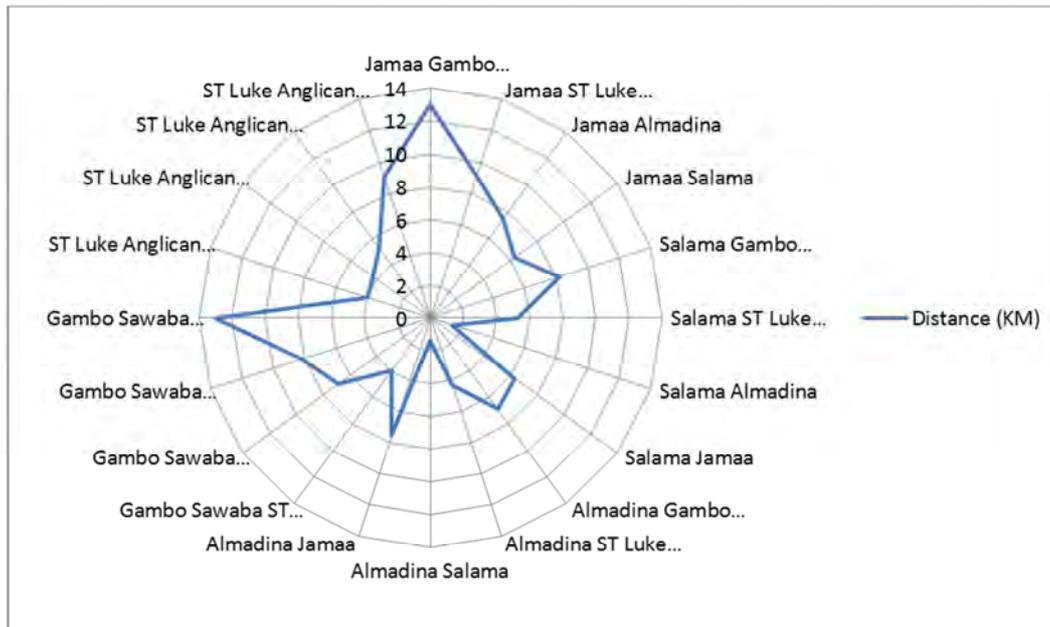
Source: Authors Analysis 2015

Figure 3. Locational Pattern Multi-distance Spatial Cluster Analysis for Hospitals under study.

The figure above shows the picture depicting location of one of the hospitals for the study area, these analysis was conducted in ArcGIS software environment. The spatial locations on the sampling points (hospital) as well as their pictures was collected using GPS device, the sampled points was typed in excel and saved as comma delinated file formats, the points were then added into the ArcGIS environment as waypoints then later added on to the map of the study area as layer, hence the hospital layer was created. Also, the pictures was saved in one folder, the hyperlink tool was latter used to hotlink the pictures to their spatial location on the map. These analysis was conducted in such away that it can be used to assess the geospatial location and pysical structures and to some extent the infrastructures of the various hospitals as a decition making information incase of any need that may arised for better alternatives and also for environmental assessments.

The table 2 and figure 3 shows the analysis for the locational Pattern of the various hospitals in the study area. The analysis was carried out using multi-distance spatial cluster analysis tool in ArcGIS environment and the table and figure showed the multiple locational and clustered distance for the hospitals in the Study area.

The Figure 4 below shows some of the Geo-database of one

of the hospitals in the study area generated from the attributes tables of the Personal geodatabase.

Table 2. Multiple-Distances from to the Hospitals.

From	To	Distance(KM)
Jamaa	Gambo Sawaba	13
Jamaa	ST Luke	9.1
Jamaa	Almadina	7.5
Jamaa	Salama	6.3
Salama	Gambo Sawaba	8.2
Salama	ST Luke	5.3
Salama	Almadina	1.4
Salama	Jamaa	6.3
Almadina	Gambo Sawaba	6.9
Almadina	ST Luke	4.3
Almadina	Salama	1.4
Almadina	Jamaa	7.5
Gambo Sawaba	ST Luke	4
Gambo Sawaba	Almadina	6.9
Gambo Sawaba	Salama	8.2
Gambo Sawaba	Jamaa	13
ST Luke	Gambo Sawaba	4
ST Luke	Almadina	4.3
ST Luke	Salama	5.3
ST Luke	Jamaa	9.1

Source: Authors Analysis 2015

Source: Authors Analysis 2015

Figure 4. Some Geodatabase for one of the hospitals (Gambo Sawaba).

From the figure above, all the results shown were obtained from the personal geodatabase created for the study area. Selection by attributes tool was used to query information on the geodatabase and all the information generated was reference to the geographical location of the hospitals in the study area. These analysis was carried out in order to get information about the various hospitals in the study area without getting to the area to source information, though information is still subject to updates as the case may be. These is one of the advantage of geodatabase over other database (the ability to link attributes information to their geographic location).

The coordinates and pictures of the hospital collected from the field were used. The coordinates used for locational base to hyperlink the pictures captured on the field. The physical natures of the hospital are presented in figures below. The figure below shows attributes of the geo-database and the various hospitals in the study area linked with their physical feature (picture).

Source: Authors Analysis 2015

Figure 5. Geodatabase with its Physical Feature of one of the hospitals (St Lukes).

The figure below show the various queries made on the geo-database so as to see the information about the hospitals in the study area.
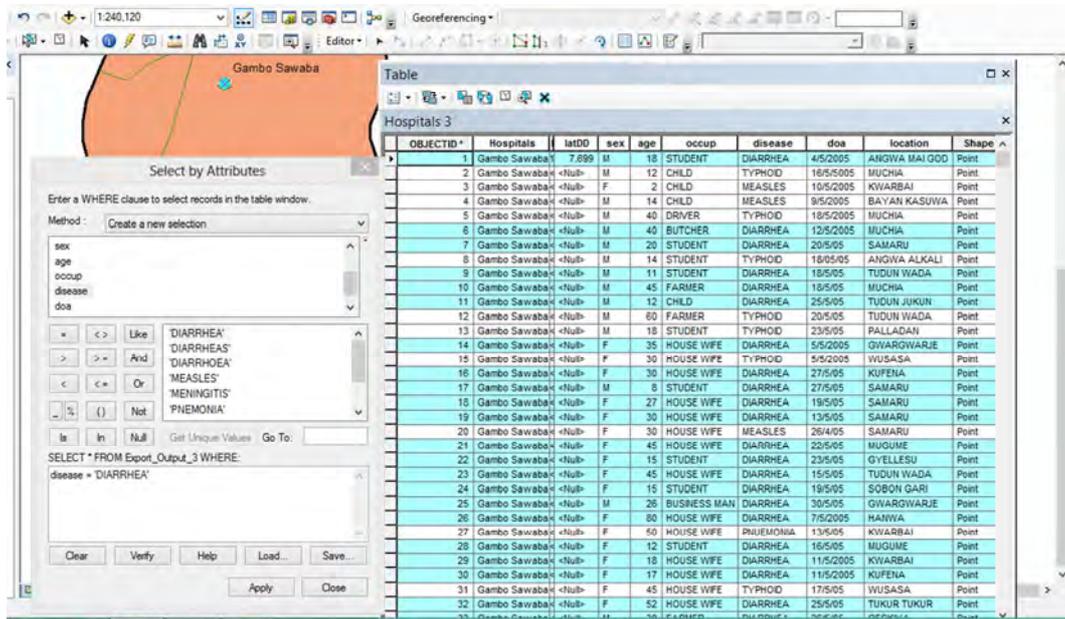
Source: Authors Analysis 2015

Figure 6. Hostipals with Diarroeah Diseases.
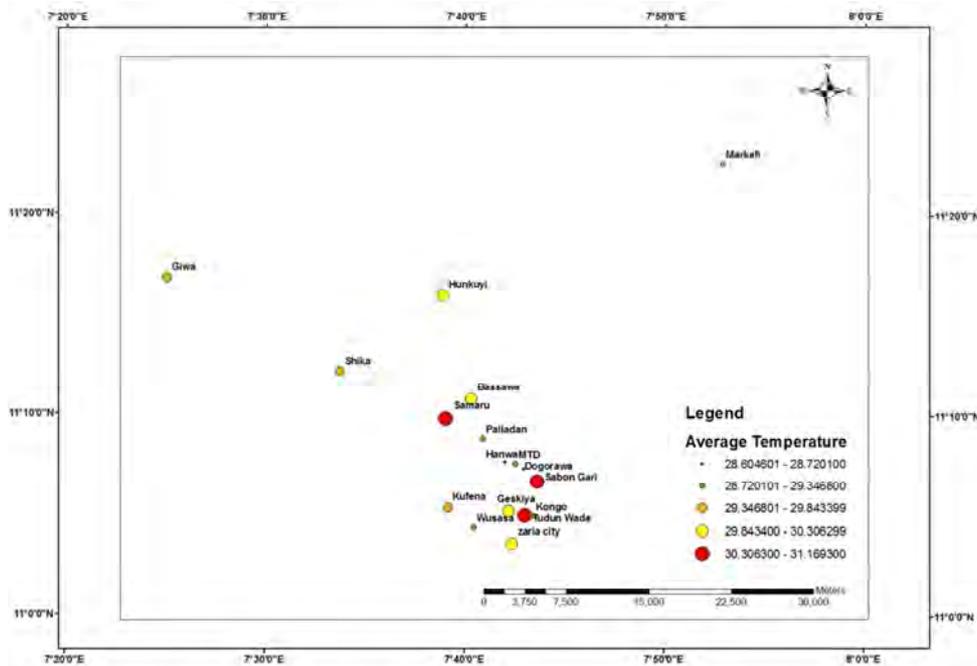
Source: Authors Analysis 2015

Figure 7. Map showing the average temperatures of the patients' location in the study area.

The selection were made using query language such as select where patient had various diseases such as diarrhea, typhoid, measles, pneumonia, meningitis, age of the patients, their locations, sex, occupation (occup) and the date of admitted (doa) to the hospitals for diagnosis and the beryl green colour line on the map. The geo-database table in figure above showing the hospitals that had the diseases and their information. This query can help the ministry of health authorities, managers and decisions maker in terms of development in order to assess the various types of diseases treated by the hospitals and the geo-database gave information on the patients.

The average temperatures of each location was also taken, these was done with the aid of the Landsat Thematic Mapper (TM). The Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors acquire temperature data and store this information as a digital number (DN) with a range between 0 and 255. It is possible to convert these DNs to degrees Kelvin using a two-step process. The first step is to convert the DNs to radiance values using the bias and gain values which are specific to our study area. An optional second step would apply an atmospheric correction using appropriate local values for several parameters, resulting in more accurate surface temperatures. The final step converts radiance data from step one, or the optional step two, into degrees Kelvin, thereafter into degree Celsius manually. The essence of obtaining the temperatures is to correlate the various diseases and their locations where they are dominant.

It was discovered that the temperatures in Samaru, Sabon Gari and Kongo are averagely high which is between 30.3063300 – 31.169300 degrees. Places such as Dogarawa and Hanwa was also discovered to have a relatively low average temperature which is between 28.604601 – 28.720100.

3. Results

Correlation Coefficient between the Temperature and Diseases under study.

The average temperature for each of the locations where correlated with the number of patients that resides in these locations.

Table 4. Correlation Between Temperature and Diseases.

		Temp	Diarrhea	Pnuemonia	Typhoid	Meningitis	Measles
Temp	Pearson Correlation	1	.482	.676	.515	.461	.795
	Sig. (2-tailed)		.472	.772	.660	.537	.753
	N	17	17	17	17	17	17
Diarrhea	Pearson Correlation	.482	1	.913	.897	.969	.839
	Sig. (2-tailed)	.472		.000	.000	.000	.000
	N	17	17	17	17	17	17
Pneumonia	Pearson Correlation	.676	.913	1	.896	.893	.802
	Sig. (2-tailed)	.772	.000		.000	.000	.000
	N	17	17	17	17	17	17
Typhoid	Pearson Correlation	.515	.897	.896	1	.841	.949
	Sig. (2-tailed)	.660	.000	.000		.000	.000
	N	17	17	17	17	17	17
Meningitis	Pearson Correlation	.461	.969	.893	.841	1	.761
	Sig. (2-tailed)	.537	.000	.000	.000		.000
	N	17	17	17	17	17	17
Measles	Pearson Correlation	.795	.839	.802	.949	.761	1
	Sig. (2-tailed)	.753	.000	.000	.000	.000	
	N	17	17	17	17	17	17

Table 4 showed the result of the correlation between temperature and the diseases. The correlation values between temperature and diseases are all positive meaning that there is a positive correlation between the Temperature of each location and the Diseases in the study area of this research.

From Figure 7, it was discovered that Measles was most prominent in Samaru and Zaria City. Also this disease was less prevalent in Giwa, Shika, Hunkuyi, Markafi, Bassawa, MTD, Kufena and Wusasa.

4. Conclussion

The spatial distribution of the five hospitals was conducted in ArcGIS environment. The co-ordinates of the hospitals as well as their pictures were taken using GPS device. These analysis was conducted in such away that it can be used to assess the geospatial location and pysical structures. The distances and locational pattern of the various hospitals was carried out using multi-distance spatial cluster analysis tool in ArcGIS environment which showed multiple locational and clustered distance for the hospitals in the Study area.

The temperature of each location was obtained with the help of Landsat thematic Mapper where it was discovered that Samaru, Sabon-Gari and Kongo had high average temperatures of between 30.3063300 – 31.169300 degrees whereas places as Dogarawa and Hanwa had relatively low temperature of 28.604601 – 28.720100 degrees within the study area.

A Personal Geo-database was created for the hospitals showing their attributes (sex of the patients, Occupation, Diseases, Date of Admission (DOA) and location from where each patient resides. Selection by attributes tool was used to query information on the geodatabase and all the information generated was reference to the geographical location of the hospitals in the study area. These analysis was carried out in order to get information about the various hospitals in the study area without getting to the area to source information, though information is still subject to updates.
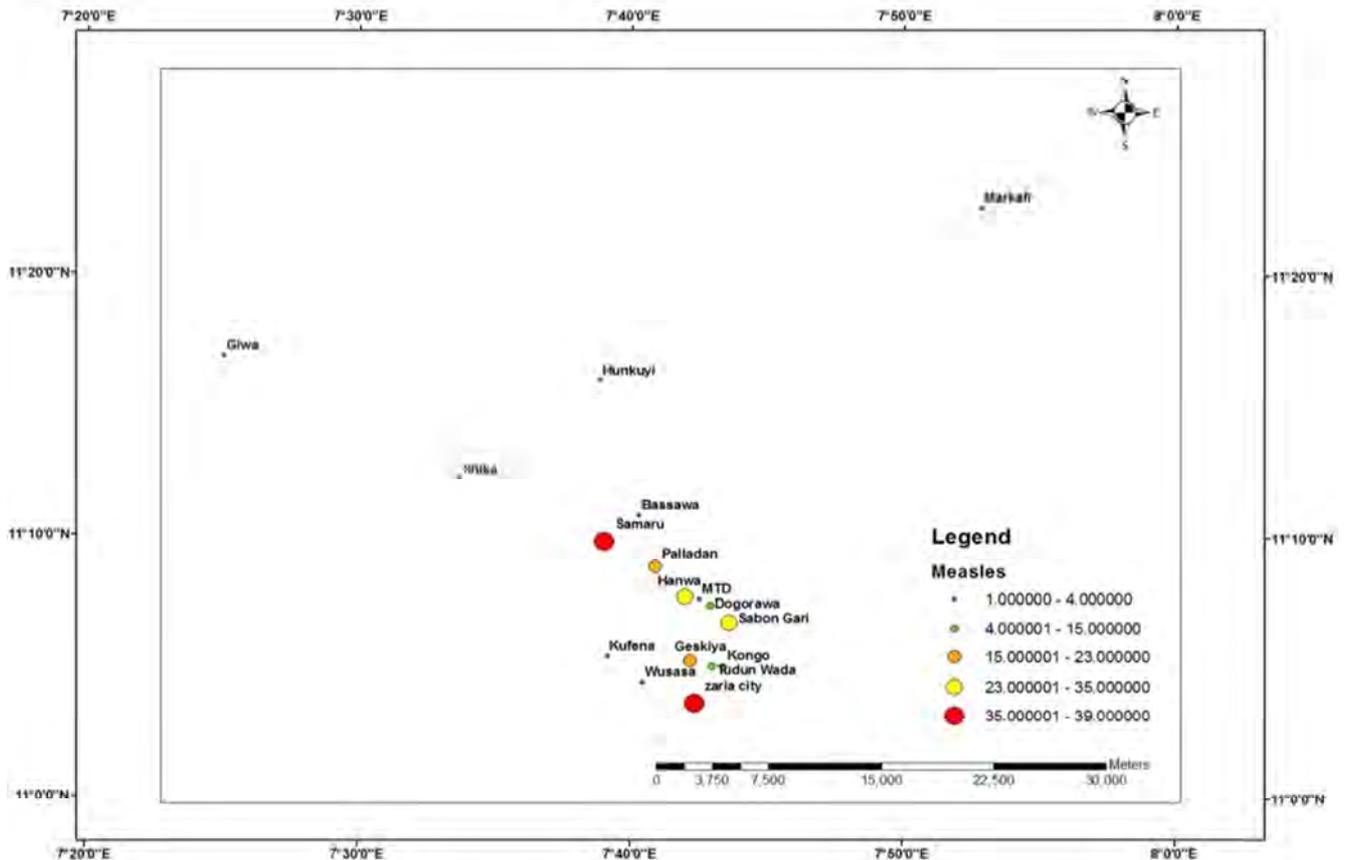
Source: Authors analysis 2015

Figure 8. Map showing the average temperature of Measles patients in the study area.

References

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

[2]   Anamzui-Ya, Jerry Asaana. *Spatial Analysis and Mapping of Cholera causing factors in Kumasi, Ghana.* Kumasi, 2012.

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

[4]   Basommi, Laari Prosper. *SPATIAL ANALYSIS OF MALARIA EPIDEMIOLOGY IN THE AMANSE WEST DISTRICT.* Ghana, 2011.

[5]   C. Coll, J. M. Galve, J. M. Sánchez, and V. Caselles, "Validation of Landsat-7/ETM+ Thermal-Band Calibration and Atmospheric Correction With Ground-Based Measurements", IEEE Trans. Geosci. Remote Sens., vol. 48, no. 1, pp. 547–555, Jan. 2010

[6]   Elliott, P, and D Wartenberg. "Spatial epidemiology: current approaches and future challenges."*Environmental health perspectives*, 2004: 9.

[7]   Elliott, P, JC Wakefield, NG Best, and DJ Briggs. "Spatial epidemiology: methods and applications." *Spatial epidemiology*, 2001: 3-15.

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

[9]   Harvell, C. D., and C. E., Ward, J. R., Altizer, S., Dobson, A., Ostfeld, R. S. et al. Mitchell. "Climate warming and disease risks for terrestrial and marine biota." (Science,) 296, no. 2158-2162 (2002).

[10] Hay, S., and G., Stern, D., Snow, R., Randolf, S. & Rogers, D. Shanks. "Climate variability andmalaria epidemics in the highlands of East Africa." (Trends Parasitol.,) 21, no. 52-53 (2005).

[11] Kandala, Ngianga-Bakwin, Chen Ji, Nigel Stallard, Saverio Stranges, and P. Francesco Cappuccio. "Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria." *The American Society of Tropical Medicine and Hygiene*, 2007: 770-778.

[12] Kouray, K, et al. "Spatial analysis of tuberculosis in an Urban West African setting: is there evidence of clustering?" *Tropical Medicine and International Health*, 2010: 664-672.

[13] Kulldorff, M. "A spatial scan statistics." *Communication in statistics-theory and methods*, 1997: 1481-1496.

[14] Osei, F. B, 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*, 2010: 331-339.

[15] Pfeiffer, Dirk. "Spatial analysis in epidemiology." *Oxford University Press*, 2008.

[16] Pinzon, J., and J., Tucker, C., Arthur, R., Jahrling, P. & Formenty, P Wilcon. "Trigger events: enviroclimatic coupling of Ebola hemorrhagic fever outbreaks." (Am. J. Trop. Med Hyg.,) 71, no. 664-674 (2004).

[17] SD Walter. "Disease mapping: a historical perspective. Spatial epidemiology: methods and application." 2000: 223-239.

[18] Sinkala, Yona, Martin Simuuza, B. John Muma, U. Dirk Pfeiffer, J. Christopher Kasanga, and Aaron Mweene. "Foot and mouth disease in Zambia: Spatial and temporal distributions of outbreaks, assessment of clusters and implications for control." *Onderstepoort Journal of Veterinary Research*, 2014: 6.

[19] Tobler, W. R. "Cellular geography In Gales S, Olsson G (eds)." *Philosophy in geography. Dordrecht, Reidel*, 1979: 379-86.

[20] Toprak, D, and S Erdogan. "SPATIAL ANALYSIS OF THE DISTRIBUTION OF TYPHOID FEVER IN TURKEY." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.*, 2008: 8.

[21] Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. E. "Hierarchical spatio-temporal mapping of diseases rates." (Journal of the American Statistical Association) 92, no. 607-617 (1997).