JHIDC Health Informatics in Developing Countries

www.jhidc.org
Vol.3 • No.2 • 2009

Spatial Predictive Model for Malaria in Nigeria

Adebayo Peter Idowu, Nneoma Okoronkwo and Rotimi E. Adagunodo

Department of Computer Science & Engineering
Obafemi Awolowo University
Nigeria

Abstract

Malaria is one of the diseases that contributed to health problems in Nigeria and this research developed a Geographic Information System forecasting model for prediction of malaria in Nigeria. This research develops a GIS based prediction model for malaria in Nigeria. With the application of GIS in malaria control, Nigeria can move towards certain reduction (if not complete eradication) of malaria infection and deaths through the use of analysis and prediction.

This system will enhance the efficacy of prevention efforts and will substantially reduce costs of prevention and treating malaria in Nigeria. The system will also serve as a good decision support system for public health officers, government office and decision makers.

Keywords
GIS, prediction model, malaria epidemic

Referencing this article

Idowu, A. P., Okoronkwo, N., & Adagunodo, R. E. (2009). Spatial Predictive Model for Malaria in Nigeria [Electronic Version]. *Journal of Health Informatics in Developing Countires, 3(2),* 30-36, from http://www.jhidc.org/index.php/jhidc/issue/view/8

1. Introduction

Malaria is a life-threatening parasitic disease transmitted by female anopheles mosquitoes. Most malaria infections, particularly in sub-Saharan Africa, are caused by Plasmodium falciparum. It is a major cause of death and threatens 2.4 billion people, or about 40% of the world's population living in the world's poorest countries and more than one million deaths are attributable to the disease annually (WHO, 2000).

In areas of Africa with stable malaria transmission, such as Nigeria, infection during pregnancy is estimated to cause as many as 10,000 maternal deaths each year, 8-14% of all low-birth-weight babies and 3-8% of all infant deaths. In Nigeria, malaria prevalence is as high as 80-85% and is the most common cause of outpatient visits to health facilities. With a case mortality rate ranging from 8-12.5% in infants and children, malaria accounts for 30% of child mortality in the country and is consistently recorded as one of the five leading causes of childhood mortality (Akpan , 1996).

In recent years, there has been a growing interest among the ministries of health and other health sector institutions in the use of Geographical Information Systems (GIS) as a tool to strengthen the analytical, management, monitoring and decision-making capacity in public health, as well as a tool for advocacy and communication between technical personnel, policy makers and the general public (Epidemiological Bulletin, 2004). This is as a result of the recognition of the capacity of GIS in managing geographical dimensions, integrating health-related data from various sources, helping to discover and visualize new patterns and geographical relations in data that would otherwise be difficult to identify, and displaying these on maps that constitute a more expressive and visual representation (Issue Focus, 1999). Recognizing the power of this tool (GIS) has led to a growing number of health studies and projects being developed by academic teams and health service professionals that include its use as a tool for analysis and decision making.

Until recently however, the use of GIS in public health generally and more specifically in the control of malaria was largely limited due to two major problems: the prohibitive cost of hardware and the great complexity of GIS software that made it extremely time-consuming as well as costly to extract information relevant to the practical demands of disease prevention and control. With the plummeting prices of hardware and simple new devices now available, applications have been developed to provide greater access to simple, low-cost geographic information and related data management and mapping systems for public health administrators at all levels of the health system.

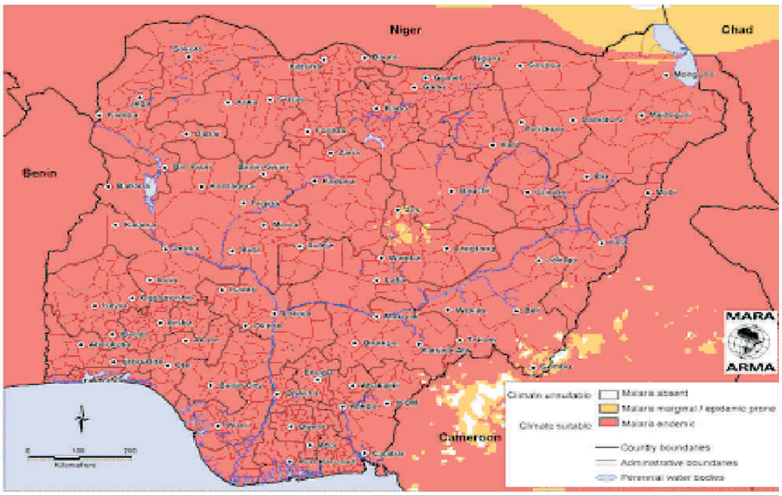
The word malaria comes from the Latin 'mal aria' (bad air) (Alnwick, 2002). In 1880, scientists discovered the real cause of malaria – a one-cell parasite called plasmodium. Later, it was discovered that the parasite enters the human bloodstream through the bite of a female Anopheles mosquito, which needs blood to nurture her eggs. Once the parasite is in to blood and after a period of incubation, the person starts to have fever which is the most typical manifestation of malaria, as well as cough, headache, diarrhoea and muscle pain (Rang et al,1999).

Malaria has been identified as one of the major causes of poverty, sickness and death in Nigeria (DFID, 2008). Available records in Lagos State, Nigeria, alone show that malaria is the most relentless killer disease. Fifteen percent of admissions in Lagos State hospitals and over 50% out-patient cases were involved in malaria treatments. Statistics across Nigeria shows similar trends (Thompson, 2004) and the distribution of malaria in Nigeria is show in figure 1 below. So, there is a need to employ the advantage that GIS offers in disease monitoring so as to develop a GIS that will be use to monitor malaria in Nigeria. Though

Over the years, statistics show that there is an increasing trend of malaria infection and mortality rate in Nigeria. It should however be noted that this data consists only of reported cases and does not take into account incidences of malaria infections or malaria reported deaths that were not reported. Figure 1 shows the distribution of malaria endemic in Nigeria. This research develops a GIS based prediction model for malaria in Nigeria using historical data.

Some organisations such as World Health Organisation (WHO), United State Agency for International Development (USAID), Public Health Agency of Canada (PHAC) had developed some system for health mapping, disease elimination, disease surveillance and as monitoring systems(WHO, 2005; PHAC, 2005) but these system cannot be used in a country like Nigeria to effectively monitor any disease because the system are based on historical data using very few cities. So there is a need for a disease surveillance system. In order to develop a good disease surveillance system that would be able to capture future disease data in Nigeria a GIS data model was proposed (Idowu et al, 2009).

Figure 1
Distribution of Malaria Endemic in Nigeria



Source: Adapted from MARA/ARMA (www.mara.org.za)

With the application of GIS in disease surveillance (especially in malaria control) Nigeria can move towards the certain reduction (if not complete eradication) of malaria infection and deaths through the use of analysis and prediction.

2. Materials and methods

2.1 Study Area and Population

The study area covers Ife central in Osun state, Nigeria, which is one of the biggest city in Osun state. Ife covers an area of approximately 350 square kilometres, lies between latitudes 25°N and 30°N of the equator, and is bounded by the Ife North, Ife south , Ayedaade ,Atakumosa west and Ife east local Governments. A large number of people in this local government are farmers and the population is over 200,0000. Malaria data were collected from Primary Health care Centre Ife Central.

2.2 Clustering

The city was grouped into clusters based on the distance between data points and this research employed Euclidean metric denoted as:

$$d(x,y) = \sqrt{(\sum (x_k - y_k)^2)} \qquad (1)$$

where

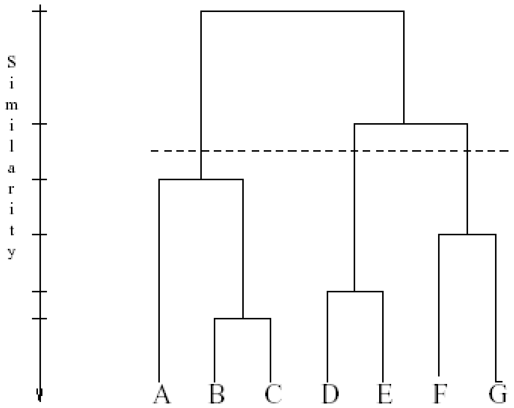
$d(x,y)$ = the Euclidean metric
x_k = the latitude coordinate of the data point on the map
y_k = the longitude coordinate of the data point on the map

The actual clustering was performed using the hierarchical clustering algorithm which is stated below:

```
j = 0
     for l = 1 to x-1
          for i = 1 to x step 2
               if k(i+1) is not null
                    b[j] = k[i] || k[i+1]
                    j = j + 1
               else
                    b[j] = k[i]
          next i
     next l
```

The basic idea behind hierarchical clustering is that every point is initially in the same cluster. With successive iterations, the closest (according to the defined metric) clusters are merged. This process is repeated until one cluster remains that contain all data points. The resulting diagram from this merging process can be stored, and by moving a virtual line from the top to bottom, the data set can be partitioned into as many clusters as required as shown in Figure 2. The data was thus represented based on clusters in the map rather than individual data points.

Figure 2
Diagram resulting from a hierarchical clustering process.



3. Prediction model

The application used for prediction of malaria occurrence based on extrapolations from past and current malaria patterns. It makes use of variations in geographical, seasonal, weather and socio-economic factors to establish a causal relationship between these factors and malaria occurrences. The prediction model was developed using artificial neural networks. In developing the model, the following steps are taken:

i. Group the malaria data collected into clusters
ii. Determine which clusters represent malaria hot-spots
iii. Develop the model based on the results of (i) and (ii) above

In this step, a mathematical model was developed was used to predict the expected occurrences of malaria based on artificial neural networks.

4. Database schema

The GIS application used a database in which information relating to malaria occurrence was stored. It stores the values of the geographical location (i.e. latitudinal and longitudinal coordinates) of the various data points and clusters on the map. The GIS application developed is a system that needs a large storage capacity due to large volumes of data due to spatial data it will handle. A number of databases met these criteria notable among which are Oracle and Microsoft SQL Server. For this research, the database engine used is the Microsoft SQL Server. A single database is required for the application. It is called the Malaria_Data database and its structure is as shown in Figure 3. Some of the tools used in describing a database are schemas and entity-relation diagrams (ERDs). The database schema is to explain the structure of the database used.

The Malaria_Data database stores all the information required by the application. It consists of three tables which are described in tables 1, table 2 and table 3. The attributes/fields/columns of these tables are explained below:

i. The data table:
This table contains data which represents the attributes of each malaria incidence diagnosed in the hospital. The fields of this table include:

Table 1
Attributes of the *'Data'* Table

Column Name	Data Type	Length	Allow Nulls
id	autoNumber	4	No
sex	Int	4	yes
age	int	4	yes
month	Nvarchar	50	yes
year	Nvarchar	50	yes
address_id	Int	4	yes
xcoord	Float	8	yes
ycoord	Float	8	yes

Table 2
Attributes of the *'Location Points'* Table

Column Name	Data Type	Length	Allow Nulls
id	autoNumber	4	No
location	Nvarchar	50	yes
xcoord	Float	8	yes
ycoord	Float	8	yes
minx	Float	8	yes
maxx	Float	8	yes
miny	Float	8	yes
maxy	Float	8	yes

Table 3
Attributes of the *'Predicted_Data'* Table

Column Name	Data Type	Length	Allow Nulls
id	autoNumber	4	no
sex	Int	4	yes
age	int	4	yes
month	Nvarchar	50	yes
year	Nvarchar	50	yes
address_id	Int	4	yes
xcoord	Float	8	yes
ycoord	Float	8	yes

- id: This is the primary key[1] for the table and thus, acts as the unique identifier for each row. The id column is an auto-number column i.e. its values automatically increment in steps of one as each new record is added to the table.

[1] A primary key is a field chosen by the database designer to uniquely identify each row in a table.

- sex: The sex column, as its name implies, stores the sex of the patient whose record is being stored in the database.

- age: This attribute stores the age of the patient whose record is to be stored.

- month: This represents the month of the year in which a particular malaria occurrence was diagnosed.

- year: This attribute specifies the particular year in which the malaria occurrence was diagnosed.

- xcoord: The 'xcoord' column stores the value of the longitudinal position of the patients residence to enable the data to be plotted on the map.

- ycoord: The 'ycoord' column stores the value of the latitudinal position of the patients residence to enable the data to be plotted on the map.

- address_id: This field acts as a foreign key for the data table and links it to the location_points table. In this way, the locations of malaria occurrences can be stored as numeric rather than string values thus saving memory. Also, the joining of the data and location_points table enables the application to compute latitudinal and longitudinal values for each record input into the database.

ii. The location_points table
This table contains data which represents the attributes of each community within the study area. The fields of this table include:

- id: This column is similar to the id column of the data table. It represents the primary key for this table and is also an auto-number column.

- location: The location field represents the actual name of a community within the study area.

- xcoord: The 'xcoord' column stores the value of the longitudinal position of the center point for the particular community whose record is to be stored in the database.

- ycoord: The 'ycoord' column stores the value of the latitudinal position of the center point for the particular community whose record is to be stored in the database.

The minx, maxx, miny and maxy columns are used to store values which represent the boundary points (in terms of longitude and latitude) for each community in the study area. These values are necessary for the computation of the latitude and longitude values for each instance of malaria data to be stored in the 'data' table.

iii. The predicted data table
The columns in this table are similar to those in the 'data' table. However, the records here are not actual malaria occurrences; rather, there are the results of predictions from the developed prediction model.

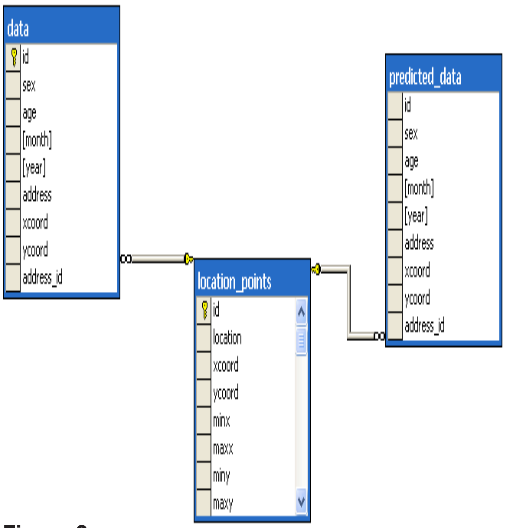
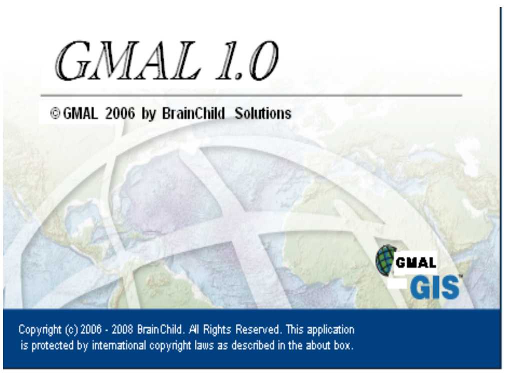


Figure 3
Diagrammatic Representation of The Database Schema

5. Results

5.1 User interface
The GMAL (GIS application for the control and attenuation of malaria) is the final result of the research. The user interface enables the user to interact with the system. In other words, it is the users' point of interaction with the system. The application makes use of a number of dialogs and windows to enhance this interaction. One of them is the main Graphical User Interface (GUI) which is shown in figure 4 and it welcomes the user to the system. To enable a user perform any function other than viewing already stored data on the map, the application requires the user to log into the system with a predefined user name and password. The login window is as shown in Figure 5 by entering an accepted user name and password, the application allows the user to perform some other function restricted by the permissions granted to that user by the system administrator. This ensures that no user is able to view or modify data which he/she has no authority over. The Create Cluster form allows a user to divide data for a particular month into clusters (i.e. it enables the user to group already stored data) in other to determine the hotspots of malaria occurrences in various months and figure 6 shows the Create Cluster form.



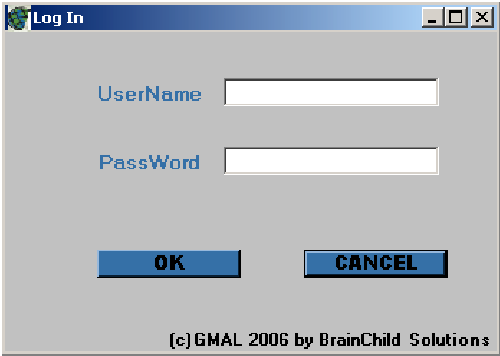


Figure 5
Login Window For GMAL

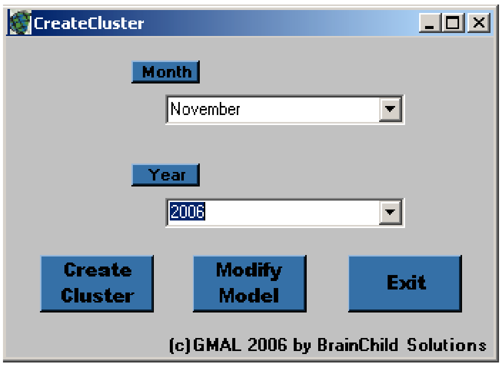


Figure 6
Create Cluster Form

5.2 Prediction model

The prediction model of the application enables it to forecast future occurrences of malaria based on data collected on past and current occurrences. It relies mainly on the data previously supplied by users (which has been stored in the database) and the seasonal climatic changes caused by the slowly varying interactions between the atmosphere to predict a malaria epidemic a month in advance. Although the prediction model can be modified to enable it make predictions on epidemics several months in advance, literature studies show that this advanced warning comes at a price [7]. The earlier the prediction is made, the less accurate it is. However, the gain in lead-time provides governments and non-governmental organizations with the opportunity to plan for a bad season.

The prediction model is developed using an artificial neural network. This network takes as its input, numerically coded representations of suspected causal factors of malaria such as temperature, the presence of water bodies (which are affected by average rainfall), and the population of the area under study. It produces as its output, a numerically coded representation of malaria incidence rates.

The network was then tested by presenting a series of input vectors for which the tester, but not the network, knows the corresponding malaria incidence rate. This process was continued until the robustness of the network was deemed sufficient i.e. the network produced accurate results for the period over which malaria data has been collected. It (the network) was then used as the prediction model.

The network used for this project was developed using a mathematical tool, MATLAB. This tool performs an analysis known as the post-training analysis and then returns three parameters in the form of a regression equation. The first two, m and b, correspond to the slope and the y-intercept of the best linear regression relating targets to network outputs. If we had a perfect fit (outputs exactly equal to targets), the slope would be 1, and the y-intercept would be 0. In this research work, it is obvious that the numbers are very close. The third variable returned by the training analysis is the correlation coefficient (R-value) between the outputs and targets. This is a measure of how well the variation in the output is explained by the targets. If this number is equal to 1, then there is perfect correlation between targets and outputs. In this research work, the number is very close to one (1), which indicates that the ANN model is a good fit. The network outputs are obtained from the equations generated by the analysis.

After the post training analysis, the model was deployed in the real environment. Malaria data (not in the training set) was fed to the neural network and prediction was made based on the ANN model. The dynamic mathematical models generated from the training analysis are as shown in the following equations.

$$M = -0.963T + 0.437 \qquad (2)$$
$$M = 0.946P + 40.00 \qquad (3)$$
$$M = 0.963R + 1.8 \qquad (4)$$

where

- M = predicted malaria occurrence at a specific period (also known as a degree-day[3])
- T = average daily temperature measured in oC
- P = total / estimated population of the area under study
- R = estimated average rainfall over a ten-day period. This is also known as the dekal rainfall rate.

The expected malaria occurrences (based on the given input parameters) as calculated by the equations above are then added to give the total expected malaria occurrence for the next period. It can be seen from the equations above that malaria occurrences have a direct relationship with population and average rainfall i.e. as population and average rainfall increases, malaria occurrences increase too. This can be explained by considering the fact that increase in rainfall ultimately leads to an increase in the number of water bodies in a given area and these serve as breeding sites for the Anopheles mosquito. Temperature, on the other hand, has an inverse relationship with malaria occurrence because an increase in temperature leads to a reduction in the survival rate of the Anopheles mosquito. If the survival rate of the malaria parasite carrier is decreased, malaria occurrence is invariably decreased too.

5.3 Input Module

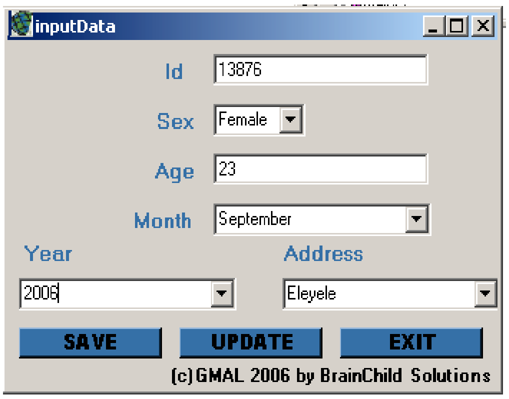
GMAL provides modules for user interaction with the application. These modules are described are discussed below.

Input data form
User inputs to the system are supplied through designed forms. For a user to enter data on malaria

[3] Malaria occurrences are estimated in degree-days rather than calendar-days. Each calendar month is made up of approximately 112 degree-days.

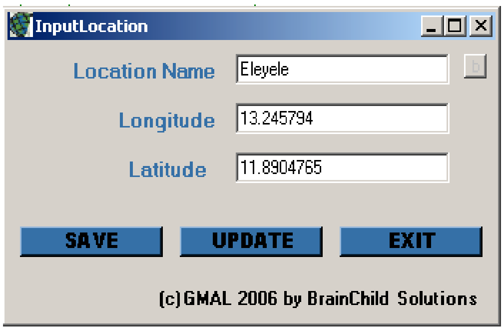
occurrences, he makes use of the Input Data form shown in Figure 7. This form allows the user to enter information for a patient who has been diagnosed to be suffering from a bout of malaria. The user then clicks on the Save button which transfers the information that has been inputted to the database after validating the data. It should be noted that clicking the Save button also generates a latitude and longitude value for the data based on the location selected as the patient's residential address. This allows the data to be plotted on a map. The user can also view, update or delete data that is already stored in the database.

Figure 7
Input Data Form



Input location form
The study area is divided into communities which enable analysis to be carried out on the input data. The names, identification numbers, latitudes and longitudes of these locations need to be stored in the database to enable the application generate proper positions for input data for patients who reside in those communities. It is expected that the user enters information for each new community he wishes to create and this is done through the input location form. Usually, enter data representing information on communities within the study area is a one-time only task i.e. it is done once through the life-time of the application. Figure 8 shows the input location form.

Figure 8
Input Location Form



5.4 Output Design
The output of an application is usually a formatted form of the information obtained after the input data has been processed. GMAL offers a number of output features which are explained in the sections below.

View prevalence form
By default, all users can only view graphically, the spread of malaria through the region for selected months and years. This is made possible within the application through the View Prevalence form shown in Figure 9. Through this form, the user is able to see, on a map, the distribution of malaria cases in the study region. The user can instruct the application to show data for a particular month, a period of months or different months selected at random.

View prediction form
The View Prediction form allows users to view the predicted spread of malaria at a future date based on the output of the prediction model. Like the View Prevalence form, this interface shows the spread of malaria in the study region on a map for the selected month over which the prediction was made.

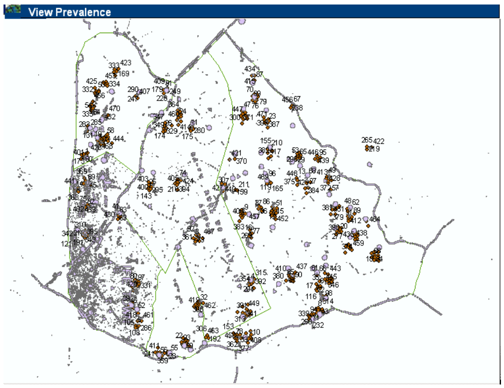


Figure 9
View Prevalence Form

6. System Specifications and Requirements

In order to install the GMAL application on a computer system, some minimum hardware and software specifications must be met and these are discussed below.

Hardware requirements
The application requires the following minimum hardware configuration:

- Intel Pentium IV MMX processor with a speed of 1.7GHz
- 3GB of hard disk space
- 256MB of RAM
- CD ROM Drive
- SVGA color monitor
- Standard keyboard and mouse
- Backup device to aid data backup

Software requirements
The minimum software requirements for the system are as listed below:

- Microsoft Windows 2000 Service Pack 2
- Microsoft SQL Server
- Arc Spatial Data Enterprise (ArcSDE)
- ArcEngine Runtime 9.0
- .Net Runtime

7. System features

GMAL includes a number of features and these are discussed below.

Inter-activeness
The application is very interactive. It allows the user to store, retrieve or manipulate data, guiding and correcting him/her through a number of dialogs and windows. It also allows the viewing of data on maps, enabling the user to have a better grip on the spread of malaria through the study area than he would have if the data was displayed in tabular form. The application has a number of error and information display facilities which informs the user when he has entered an invalid data or performed an invalid function.

Help feature
GMAL contains an embedded help facility for the application. The user can click on the "Help" menu to view a number of help topics. The help facility contains information on malaria, Geographical Information Systems (GIS), Ife-Central local government area and on the usage of the application. Users can scroll through and read the help information presented.

Security
The security features of GMAL ensure that users do not have access to restricted data. It allows users to change their login passwords; and allows the system administrator to create users, drop users and grant/revoke privileges to these users. It also assists in the backup and restoring of data. These features are implemented through the security menu.

8. Conclusion

This GMAL would be used for the attenuation and control of malaria in Nigeria and it would not only help stake holders in the health sector make intelligent decisions on how to combat malaria, it simplifies the complex processes involved in using some commercial GIS applications for epidemiology. Driven by menu items, the application developed would allow its users to enter and retrieve data already stored in the database. It would also allow them view this data on maps and perform queries on the database with little or no knowledge of SQL. The system includes a help file which directs users on its use and gives simple instructions. It also includes fine screen display and proper documentation.

Acknowledgments

We acknowledge our references and those that contributed to the development of this system. Thank you all God bless

References

Akpan, H.E.M. 1996, "Home Management of Malaria in Children." *Malaria Society of Nigeria Newsletter* Vol1 No1, pp 8-9.

Alnwick D.M., 2002, *'Malaria, a Parasite Laughing at Cheap Drugs'*. Unspecial No. 605, , pp 48-52

DFID(2008). Rolling back malaria in Nigeria. Available from http://www.dfid.gov.uk/news/files/success_stories/africa/nigeria_malaria.asp [Accessed March 27, 2009]

Epidemiological Bulletin., 2004, "Software Programs for mapping and Spatial Analysis in Epidemiology and Public Health". *Epidemiological Bulletin*, Vol. 25 No. 4, pp1.

Issue Focus, 1999, "Geographical Information Systems in Public Health - Part 1", *Journal for Public Health Management and Practice.* Vol. 4 No. 1,

PA Idowu, Dan Cornford & Lucy Bastin (2009): A Conceptual Data Model for Disease Surveillance, Monitoring and Prediction in Nigeria. Proceeding of Second International Conference on Health Informatics, pp 473-480, January 14-17, 2009, Hotel Gaia Melia Porto, Porto, Portugal.

Public Health Agency of Canada (2005). Disease Surveillance [online]. Available from www.phac-aspc.gc.ca.dsol-smed [Access on 22 Oct 2008].

Rang H.P., Dale M.M. & Ritter J.M. (1999). *'Pharmacology'*, 4th edition. Prentice Hall, London. Pg 726 – 245

Thompson, A., 2004, *'Malaria: When Politics Kills'*, This Day Newspaper, March 23,

World Health Organization.(2000),. "$1 to Roll Back Malaria." *Partnership* Vol1, No 8, , pp 9

World Health Organisation (2005). Public Health mapping and GIS, [Online] Available from http://www.who.int/health_mapping/en [Accessed 22/Oct 2008]