SPATIAL ASSOCIATION BETWEEN MALARIA PANDEMIC AND MORTALITY

B.M. Dansu and O.E. Asiribo***

Department of Statistics, University of Agriculture, Abeokuta, P.M.B. 2240, Ogun State, Nigeria
** Email:* mrsdansuedu@yahoo.com

ABSTRACT

Malaria pandemic (MP) has been linked to a range of serious health problems including premature mortality. The main objective of this research is to quantify uncertainties about impacts of malaria on mortality. A multivariate spatial regression model was developed for estimation of the risk of mortality associated with malaria across Ogun State in Nigeria, West Africa. We characterize different local governments in the data and model the spatial structure of the mortality data in infants and pregnant women. A flexible Bayesian hierarchical model was considered for a space-time series of counts (mortality) by constructing a likelihood-based version of a generalized Poisson regression model that combines methods for point-level misaligned data and change of support regression. A simple two-stage procedure for producing maps of predicted risk is described. Logistic regression modeling was used to determine an approximate risk on a larger scale, and geo-statistical ("Kriging") approaches were used to improve prediction at a local level. The results suggest improvement of risk prediction brought about in the second stage. The advantages and shortcomings of this approach highlight the need for further development of a better analytical methodology.

Keywords: Bayesian hierarchical, Generalized Poisson regression, Geo-statistics, Spatial analysis, Kriging, Climatic factors, Malaria pandemic, Mortality

1 INTRODUCTION

1.1 The burden of malaria in Africa

In areas of stable endemic malaria transmission in sub-Saharan western Africa, it has been estimated that in 1995 about 1 million deaths were directly attributable to malaria infection (Snow et al., 1999). Of these deaths, three-quarters were recorded among children below the age of 5 years. In the same population, it was estimated that about 200 million clinical attacks of malaria occurred in the same year. In areas of unstable or epidemic prone malaria in southern Africa ("fringe area"), about 2000 deaths and 200,000 clinical episodes occurred that were due to malaria and were not prevented despite malarial control measures in these areas. Accordingly, a World Bank report of 1993 noted that malaria accounts for an estimated 35 million disability-adjusted life years (DALYs) per year lost in Africa due to ill- health and premature death (World Bank, 1993).

The discovery of an interactive effect between HIV infection and malaria morbidity (Whitworth et al., 2000; Chandramohan & Greenwood, 1998; Verhoef et al., 1999) exacerbates the potential for devastating health consequences in populations with large numbers of individuals who are co-infected. In resource-poor countries in Africa, malaria prevention and treatment consume a large proportion of the health budget, and because it poses a threat to the indigenous population as well as visitors, it acts as a deterrent to tourism and foreign investment in these countries. Malaria therefore not only affects the health status of Africa's population, but also has far-reaching economic consequences inhibiting economic development (Wernsdorfer & Wernsdorfer, 1988). The impact of malaria on the population and its significance on development in the region was recognized by the Abuja, Nigeria Summit in April 2000 as the first African summit of Heads of Government on malaria control. The communiqué from the meeting calls, among other things, for more research on trends in incidence and prevalence of malaria, epidemic outbreaks and clinical epidemiology (Sachs, 2000). A better understanding of the distribution of malaria has been identified as an important tool in its control (Snow et al., 1996). More accurate maps make it possible for interventions to be mounted that are appropriate to the disease profile, which characterizes particular levels of endemicity. However, for clinical trials and evaluation, new approaches should be located correctly, and for planners

of irrigation and other development schemes to take cognizance of the potential effects of these schemes on malaria transmission intensities.

1.2 Transmission of malaria

Malaria is caused by the parasites of genus Plasmodium. The four species of Plasmodium are *P. falciparium*, *P. malariae, P. ovale,* and *P. vivas*. In Africa, the predominant species of the disease-causing parasite is *P. falciparum.* Infection of the human host occurs when a person is bitten by a female Anopheles mosquito that has previously become infected. The parasite, called sporozoite at this stage of its cycle, enters the human body via the saliva of the mosquito that is injected into the blood. The parasites multiply in the liver and re-invade the blood via red blood cells as merozoites. These develop into a stage known as the trophozoite, which is the one visible in blood films, and subsequently divide by the process of schizogony to produce further merozoites, which invade non-infected blood cells. Some of the merozoites develop into new trophozoites, while others develop into male micro or female macrogametocytes. Uninfected Anopheles mosquitoes become infected if they feed on a person with mature gametocytes in their peripheral blood. Within the mosquito, the microgametocytes exflagellate into gametes before fertilizing the macrogametocytes, thereby forming zygotes. The zygote changes into an ookinete and then into an oocyst, which is found in the mid-gut wall of the mosquito. Large numbers of sporozoites are formed within the oocyst. The rate of development of sporozoites in the oocyst is temperature dependent. The sporozoites leave the oocyst to invade the mosquito's salivary glands, from where they can infect another human host when the mosquito takes a blood meal. The incubation period of the parasite in the vector takes 13 days to complete at 24^0 C. for *P.falciparum.* The vector will only become infective if it survives this sporogonic cycle (Gilles & Warrell, 1993).

Malaria as a disease is therefore closely bound to conditions which favour the survival of the anopheles mosquito in the form of habitat and breeding sites and which favour the life cycle of the parasite in term of suitable temperatures. In the absence of any human intervention, these conditions are predominantly determined by climatic and environmental factors.

1.3 Clinical manifestations

Clinically, malaria manifests itself in its mild form as an illness associated with other non-specific symptoms (Bruce-Chwatt, 1980). The first clinical sign will only appear after the incubation period, which varies between nine and fourteen days for falciparum malaria. Clinical diagnosis is usually confirmed by a blood test, involving microscopic evidence of parasites in the blood, or by a rapid diagnostic kit (Craig & Sharp, 1997). However, in endemic countries, infected individuals are often asymptomatic, so that parasitological evidence does not necessarily prove that the symptoms are due to malaria in a particular patient (Bruce-Chwatt 1980; Snow et al., 1997).

Severe life threatening malaria is usually due to P.falciparum malaria. In non-endemic area, cerebral malaria is the sequel that often sets in after the initial general symptoms. In such areas, death due to malaria in both children and adults is usually due to cerebral malaria. In highly endemic areas, severe malaria affects mainly young children and women during pregnancy. In such areas, infants may enjoy a period of inherited immunity of up to 6 months. As this declines, clinical attacks become more severe and often take the form of severe anemia, which is responsible for most deaths due to malaria in these areas. Depending on the intensity of exposure to the parasite, these children develop relative tolerance to malaria infection in their first few years of life. As a result of this, older children and adults usually exhibit mild non life-threatening clinical symptoms, if any.

1.4 Malaria control

In areas of high transmission intensity, the use of insecticide treated bed nets (ITBNs) and materials has become recognized as an effective means of malaria vector control for reducing mortality and severe morbidity in young children and pregnant mothers (Binka, 1997; Abdulla et al., 2001). In an integrated strategy, these would be used in conjunction with rapid and effective algorithms for diagnosis and the availability of efficient and affordable drugs for case management.

In areas of low transmission intensities (particularly in southern Africa), house spraying with residual insecticides (for example, pyrethroids or DDT) has been widely used as an effective means of vector control, coupled with definitive diagnosis and treatment towards parasitological cure (Sharp et al., 2000). More recently, this has been complemented with the use of ITBNs in specific areas (Mnzava et al., 1999).

Malaria parasite control in most parts of Africa, including the malaria 'fringe' areas in the south, has been affected by large-scale parasite resistance to the cheap anti-malarial drugs, such as chloroquine and increasingly to sulphadoxine/pyrimethamine (SP). In Ogun State Nigeria, this has necessitated a recent decision to introduce combination therapy including artemisnin in place of previously used SP.

1.5 Malaria distribution data and measures of transmission intensity

For modeling malaria transmission intensity, the measure of choice is the entomological inoculation rate (EIR), which is the number of infective bites per person per year, because it is a direct measure of exposure to which individuals are subjected. Unfortunately, this is not widely available. Other potential measures would be the vectorial capacity, man-biting rate, and parasite ratio and incidence rates. Irrespective of the merits and demerits of these measures, the only one that is widely available for the whole continent is the parasite ratio or prevalence of infection. This is obtained by a random survey of individuals who are tested for the presence of parasites in their blood. The results of thousands of these surveys taken over time across the length and breadth of malarious areas in Africa have been consolidated in the MARA database (MARA/ARMA Collaboration, 1998). Because of the effects of partial immunity in endemic malaria areas, surveys that include older children and adults do not give a reliable measure of potential infection rates. For this reason, only surveys (or components of surveys) restricted to children less than 10 years of age have been included in analysis for the purpose of malaria distribution modeling. A general problem with such surveys is that they are predominantly located in areas of high transmission intensity, leading to an under-representation of populations living in low transmission environments.

It has been shown that parasite ratios are reasonably well correlated with EIR (Beier et al., 1999). For this reason the parasite ratio is an acceptable proxy for transmission intensity. It needs to be remembered, however, that the parasite ratio is dependent on the age group of children being surveyed and, to some extent, on the season. If the main objective of modeling is to predict malaria risk in broad categories, then the parasite ratio is the most practical measure because of its availability.

Another proxy of transmission intensity that is fairly widely available in southern Africa is parasitologically confirmed disease incidence. Incidence data generally are biased because they may reflect patient access to health services rather than true morbidity, and they are dependent on good denominator data being available at the same level of aggregation as the case data. In the western part of Nigeria, a surveillance system is used that is believed to identify the vast majority of cases because active cases findings supplement the passively reported cases, as part of a malaria control strategy that seeks to identify and treat every infected individual. Reasonably good population data are also available for this area. Incidence data for this population are unique in that they have been recorded over many years. As malaria in the area is seasonal and highly variable over space and time, these data present an unequalled opportunity to investigate the relationship between climatic variability and malaria incidence in a mainly non-immune population and to explore the potential of epidemic prediction using satellite derived meteorological data.

This research therefore used both parasite ratios and malaria incidence data to undertake spatial statistical analysis of malaria distribution. In this research, parasite ratios are used to model the relationship between malaria and climatic factors in order to produce prediction maps of prevalence of infection. The research uses incidence data to analyze spatial and temporal variation in incidence and to investigate relationships between climate and malaria at a small area level by using spatial and spatial-temporal models.

There have been previous projects to map the distribution of malaria in Africa. These have ranged from expert opinion maps (Molineaux, 1988), to suitability maps (Craig et al., 1999), to maps for a single country that have used parasite ratios (Thomson et al., 1999). While this research is not attempting to produce a detailed empirically

derived risk map for the whole continent, it attempts to show approaches using modern statistical methods that are suitable at different levels of scale ranging from different local government areas in Ogun State.

1.6 Spatial statistical modeling and mapping of malaria

There is a wide range of approaches to spatial analysis and modeling in the statistical and geographic information systems (GIS) literature. Many of these approaches have been recently developed in response to an interest in spatial processing and presentation of data and to the opportunities that have been opened up through the collection of small areas data and the development of GIS technology and software. However, the idea of spatial analysis to solve epidemiological problems goes back to the very beginning of epidemiological problems and research (Snow, 1855).

Statistical approaches to spatial analysis have in common the concepts of correlation or non-independence of spatial data. This can be a problem that needs to be taken into account when analyzing such data as the degrees of freedom tend to be exaggerated, or it can be usefully exploited, for example, in stabilizing small counts of cases in small areas by borrowing strength from neighboring areas. Sometimes the mere existence of significant spatial correlation is a statistical result of interest in itself (Walter, 1994). Results of spatial statistical modeling are estimated quantities (parameters) that are intended to quantify the true underlying magnitudes and their uncertainty in a map rather than the mere mapping of recorded data that are subject to sampling error. The role of GIS in such analyses is two fold:

a) To pre-process data, for example, by extracting values, or calculating distance or proximity.
b) To post-process results, for example, by plotting estimated area effect in a map.

The essential core of such spatial analysis is, however, stochastic and uses statistical programs that take account of the random nature of the processes involved. Modeling approaches that are based purely on GIS techniques tend not to deal with the random nature of process explicitly and hence produce point estimates of processed quantities for individual objects or pixels in a map.

In this research, spatial statistical analysis was performed, with GIS employed as a pre- and post-processing tool, but with statistical software used for the main analysis. Two distinctly different approaches to spatial statistical modeling have been followed, without attempting to make direct comparisons between the two. In this research, geostatistical approaches in conjunction with generalized linear mixed models (GLMM) have been followed. Also, hierarchical fully-Bayesian methods using Markov Chain Monte Carlo modeling were used.

Geostatistical or variogram approaches have occasionally been applied to disease mapping (Carrat & Valleron, 1992; Oliver et al., 1992). In these, the method of "ordinary kriging" was used as a means of interpolating disease prevalence or incidence across a map, based on observed values at known grid locations. A variogram is used to model spatial dependence in the observed data. Classical kriging is based on the assumption that the response is a continuous variable, that its underlining value is constant across the map (stationary), and that the covariance between two points is entirely a function of distance between them.

Hierarchical fully-Bayesian methods using MCMC sampling (Gelfand & Smith, 1990) have been widely applied to disease mapping and ecological regression analysis in recent years. In this approach the correlation among neighboring areas is modeled via conditional autoregressive (CAR) priors. Such methods have been developed for data in which the response represents an areal unit as well as for data representing points. However, readily available statistical software using these methods is currently restricted to areas-based spatial data, limiting its application to malaria distribution data, which are generally point referenced, with the exception of the reporting system that is available in South Africa. Virtually all applications of Bayesian disease mapping methods in literature are in the context of rare diseases such as rare cancers in developed countries of Europe and North America. Vector borne diseases in tropical countries differ in that the disease is often not rare and in that the spatial correlation is often much stronger due to the links with climatic and environmental factors. The quality of both disease data and age-sex specific population data is also generally of a lower standard than is the case, for example, with cancer registration data in first world countries. The aim of this research is to estimate malaria prevalence and mortality in Ogun State, Nigeria. A second goal is to estimate malaria prevalence and incidence at map locations or areal units by means of spatial statistical modeling and determine factors associated with spatial and temporal heterogeneity of malaria transmission intensity.

This was done by applying recent methodology in the spatial analysis of correlated disease data and thereby evaluates the potential of this methodology to vector borne disease and other tropical disease data in general. It also attempts to document the time trend of malaria incidence in an area of unstable malaria and to suggest some reasons why malaria incidence has increased so unevenly in this area.

2 THE STUDY

2.1 Data collection and data preparation

Malaria prevalence data were collected from Ministry of Health in Ogun State, Nigeria. Twenty local government area surveys were identified yielding suitable estimates of malaria prevalence. The surveys represent historical data whose screening for inclusion in MARA/AMRA database has been documented elsewhere (MARA/AMRA Collaboration, 1998).

For each survey, the total sample size and number of individuals testing positive was known. The geographical coordinates of each survey were established using paper maps, electronic maps, and global positioning systems. The distribution of surveys across Ogun State was uneven, with higher concentrations of surveys in more densely populated areas and in areas where malaria risk was perceived to be high.

For each of the survey coordinates, long-term climatic averages, normalized difference vegetation index (NDVI) (NDVI Image Bank Africa, 1991), and population density were obtained. A number of published data sets were available for this purpose (Hutchinson et al., 1995; Africa Data Sampler, 1995). The resultant array of variables consisted of: monthly average maximum temperature, monthly average minimum temperature, monthly NDVI, and population density. In addition, the number of months with rainfall in excess of 60mm (regarded as suitable for malaria transmission) was computed for each location. Using GIS, the distance to the nearest water body was also calculated.

All climatic variables were available as long-term averages for each calendar month but not by individual year. The monthly averages of the climatic variables are highly correlated within climatic seasons. The question arises over what period of climatic variables should be sensibly averaged, as the shorter the aggregation period, the stronger the likelihood of a high degree of serial autocorrelation in the values. For the purpose of selecting climatic variables for explaining the variation in malaria prevalence, it was decided to average monthly climatic data over climatic seasons in order to reflect the variation in weather. Temperature and rainfall were averaged over 3 month periods, with the first quarter starting in December to coincide with the beginning of the dry season. The vegetation index NDVI was aggregated over two six-month periods corresponding approximately to the dry season (December to May) and the wet season (June to November) respectively.

2.2 Methods and results

The first stage of this analysis involved ordinary logistic regression analysis to determine the relationship between malaria prevalence and ecological predictors of malaria. In the second stage, we investigated spatial pattern in the residuals of the model and used residual spatial dependence in the data to improve prediction at local level.

2.2.1 Regression analysis

The relationship between malaria parasite prevalence and each individual potential explanatory variable was first investigated by inspection of scatter-plots and by single variable regression analysis. Because parasite prevalence data are binomial fractions, a logistic regression model for grouped (blocked) data was used as standard practice for the analysis of such data (Hosmer & Lemshow, 1989). Predictions of prevalence made from the logistic model always fall within the interval 0 and 1. Larger surveys are implicitly accorded more weight than the smaller ones. The *glm* command in the Statistical Analysis System (SAS 9.1, 2002) was used for the analysis.

Each of the explanatory variables was adjusted for all of the others using multiple regressions. Non-linearity in the relationship between parasite prevalence and a predictor variable was explored by adding polynomial terms and then grouping the values of continuous variables into categorical ones. Variable selection for the multiple logistic regression model was carried out by a combination of automatic (stepwise) procedures, goodness of fit criteria, and by using judgment in selecting variables that explain malaria prevalence in terms of vector, host, and parasite dynamics of malaria. An additional criterion for selection of the final model was the degree of spatial correlation of the model residuals (see below).

The final multiple logistic regression model contained four significant explanatory variables for the prediction of malaria prevalence. These were distance to water (categorical), average NDVI during the wet season (June to November, also categorical), number of months with more than 60mm rainfall, and average maximum temperature during the quarter March to May. Table 1 summarizes these results.

Table 1. Factors associated with malaria parasite prevalence. Adjusted odds ratios obtained by multiple logistic regressions.

	Unadjusted		**Adjusted**	
Variable	**Odds Ratio**	**95% C.I**	**Odds Ratio**	**95% C.I**
Vegetation index (NDVI) in rainy season (relative to NDVI of 0.50 or less)				
0.50>NDVI<= 0.8	17.16	3.69-57.44	3.43	1.27-11.75
NDVI>0.8	34.70	10.00-109.83	3.95	1.12-17.55
Distance to water (relative to less than 4km)				
between 4 and 40 km	2.45	2.42-2.64	2.55	1.90-3.423
more than 40 km	0.18	0.15-0.21	0.69	0.21-2.31
Average maximum temperature, March to May change per ^{0}C	0.74	0.62-0.87	1.38	1.12-1.62
Length of rainy season (months)				
Change for each month of season Length	1.51	1.49-1.63	1.68	1.32-2.34

The final model explains about 65% of the total variation in malaria if one takes the reduction in deviance as a measure of variation. It must be noted that the final model is 'over dispersed,' i.e. the residual deviance is larger than would be expected for the number of degrees of freedom. This has been taken into account in the model by using a deviance based extra dispersion parameter, which results in inflating the model standard errors of the model parameter by the square root of the dispersion factors (Littell et al., 1996). The inclusion criteria for the variables selected for the final model can therefore be regarded as conservative.

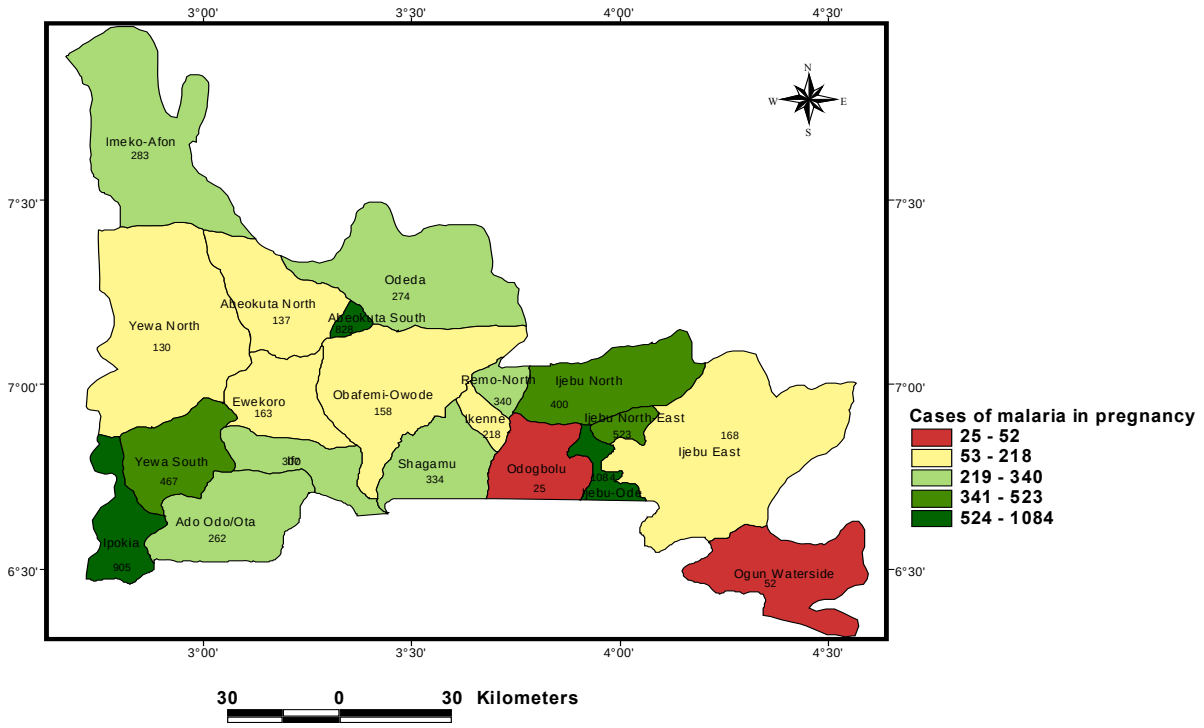

Figure 1. Map of Malaria cases in pregnant women

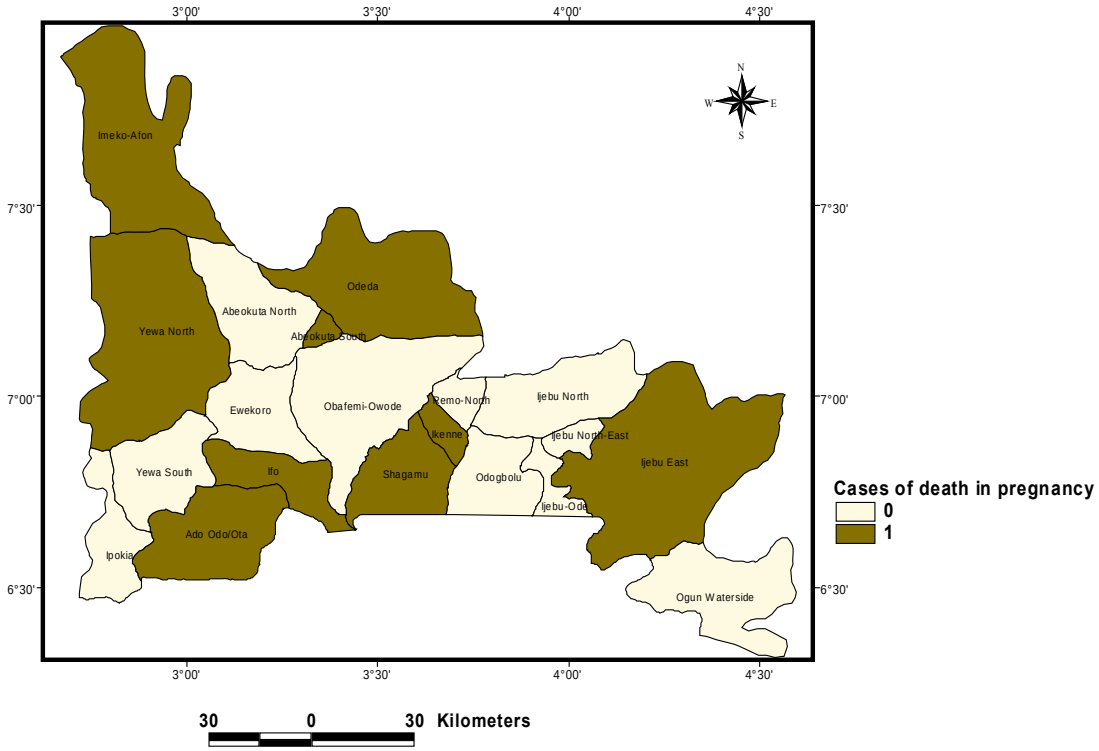

Figure 2. Map of deaths due to Malaria in pregnant women

3 CONCLUSION

This study examined the relationship between malaria prevalence and geographical factors that contribute to malaria spread. It is our view that the model produced in this study is a reasonable representation of malaria risk in Ogun State. The reduction of residual spatial pattern enhances confidence in the fidelity of the model and residual dependence that has been modeled by kriging wherever the density of observed points allow for this. Kriging has been made possible by leveling the map through the regression model, and applying the kriging process to the residuals. The final predictions make sense from the entomological perspective. However, a more systematic approach to this work in future would be a full mixed model with universal kriging to take account of spatial pattern

4 ACKNOWLEDGEMENT

We would like to acknowledge Ogun state Ministry of Health, Nigeria for given us data used in this research work and Geographical Information System (GIS) Department under Research and Development Center (RESDEC) at University of Agriculture, Abeokuta Ogun State Nigeria. Thank you all.

5 REFERENCES

Abdulla, S., Schellenberg, J., Nathan, R., Mukasa, O., Marchant, T., Smith, T., Tanner, M., & Lengeler, C. (2001) Impact on malaria morbidity of a programme supplying insecticide treated nets in children aged under 2 years in Tanzania: community cross sectional study. *BMJ: 322(7281):*249-50.

Beier, J., Killen, G., & Githure, J. (1999) short report: entomologic inoculation rate and plasmodium falciparum malaria prevalence in Africa. *Am J Trop Med Hyg 51(3):*271-280.

Binka, F. (1997) *Impact and determinants of permethrin impregent bednet on child mortality in Northern Ghana.* PhD dissertation, University of Basel.

Bruce-Chwatt, L. (1980) *Essential Malariology*. Heineman Medical Books Ltd, London.

Carrat, F., & Valleron, A. (1992) Epidemiologic mapping using the "kriging" method: application to an influence-like illness epidemic in France. *America Journal of Epidemiology 135(11):*1293-1300.

Chandramohan, D. & Greenwood, B. (1998) Is there an interaction between human immunodeficiency virus and plasmodium falciparum. *International Journal of Epidemiology 27*:296-301.

Craig, M. & Sharp, B. (1997) *Comparative evaluation of four techniques for the diagnosis of Plasmodium Falciparam infections.*

Craig, M., Snow, R., & le Sueur, D. (1999) A climate-based distribution model of malaria transmission in Africa. *Parasitology. Today 5(3*): 105-111.

Gelfand, A. & Smith, A. (1990). Sampling- based approaches to calculating marginal densities. *Journal of the America Statistical Association 85*:398-409.

Gilles, H. & Warrell, D. (1993). *Bruce-Chwatt's Essential Malariology.* Edward Arnold. London.

Hosmer, D. & Lemshow, S. (1989) *Applied Logistic Regression*. John Wiley and Sons: New York.

Hutchinson, M., Nix, H., McMahon, J., & Ord, K. (1995) *Climate data: a topographic and climate database* (CD-Rom). Centre for Resource and Environmental Studies, The Australian National University, Canberra, ACT 0200, Australia.

Molineaux, L. (1988) *The epidemiology of human malaria as an explanation of its distribution, including some implications for its control*.

Sachs, J. (2000) Economic analysis indicates the burden of malaria is great. In: *The Africa Summit on Roll Back Malaria, Abuja*; page 27-35, WHO/CDS/RBM/2000.17

Sharp, B, (2001) Malaria. *South Africa Health Review 2000*. Health System Trust, Durban.

Snow, R. (1999) a continental risk map for malaria mortality among African children. *Parasitology Today15 (3)*:99-104.

Snow, R. (1996) The need for maps of transmission intensity to guide malaria control in Africa. *Parasitology Today 12*: 455-457.

Thomson, M., Connor, S., D'Alessando, U., Rowlingson, B., Diggle, P., Cresswell, M., & Greenwood, B., (1999) Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the important of spatial correlation in the interpretation of results. *Am J Trop Med Hyg 61(1):*2-8.

Wernsdorfer, G. & Wernsdorfer, W. (1988) Social and economic aspects of malaria and its control. In: Wernsdorfer, W. & McGregor, I (eds.) *Malaria: Principles and Practice of Malariology, Vol.2.* Churchill Livingstone: London.

Whitworth, J. (2000) Effect of HIV-1 and increasing immunosuppression on malaria parasitaemia and clinical episodes in adults in rural Uganda: a cohort study. *Lancet 356*: 1051-56.

White, G. (1974) Anopheles gambiae complex and disease transmission in Africa. *Trans R soc Trop Med Hyg 68*: 278-301.

Verhoeff, F., Brabin, B., Hart, C., Chimsuku, L., Kazembe, P., & Broadhead, R. (1999) Increased prevalence of malaria in HIV-infected pregnant women and its implications for malaria control. *Tropical Medical and International Health 4(1):* 5-12