Spatio-temporal modelling of malaria incidence for evaluation of public health policy interventions in Ghana, West Africa

<u>S.K. Appiah</u>[1, 2], **U. Mueller**[1] **and J. Cross**[1]

[1]*School of Engineering, Faculty of Computing Health Science, Edith Cowan University, 270 Joondalup Drive, Joondalup, Perth WA 6027, Western Australia*
[2]*Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*
Email: sappiah@our.ecu.edu.au

Abstract: Malaria is a major challenge to both the public health and the socio-economic development of Ghana. Major factors which account for this situation include poor environmental conditions and the lack of prevention services. In spite of the numerous intervention measures, the disease continues to be the most prevalent health problem in the country. The risk assessment reports for Ghana were based on household surveys which provide inadequate data for accurate analysis of incidence cases. This poses a serious threat to planning and management for the health care delivery system in Ghana.

Malaria transmission varies with geographical location and time (or season). Spatio-temporal modelling coupled with adequate data has shown to better define the public burden of the disease, providing risk maps to describe the incidence variation in space and time and also identifying high risk areas for health policy implementation. Geostatistics contributes immensely to the prediction of the random processes distributed in space or time in epidemiological studies.

In this study, we conduct spatial statistical analysis of malaria incidence to produce evidence-based monthly maps of Ghana illustrating the patterns of malaria risk over space and time. This is achieved using monthly morbidity cases reported on the disease from public health facilities at district level and population data over the period 1998-2010 to compute the malaria incidence rates, being the number of reported cases per unit resident population of 10,000. Lognormal ordinary kriging is used to model the spatial and temporal correlations, and then back-transformed to estimate the monthly malaria risk at local level.

The space-time experimental variogram describing the correlations structure is modelled with nested spherical and exponential-cosine functions coupled with nugget effect. The modelled variogram indicate both short and long spatial and temporal dependence of the malaria incidence rates at local level with the temporal component exhibiting an increasing seasonal pattern of period of 12 months. The results also indicate varied spatial distribution of malaria incidence across the country, the highest risk being observed in the northern most and several locations in central and western parts of the country, and lowest in some areas in the north and south along the coast. This statistical-based model approach of malaria epidemiology will be useful for short-term prediction and also provide a basis for resource allocation for the disease's control in the country.

Keywords: *Malaria incidence rates (MIR), spatio-temporal modelling, space-time variogram, space-time ordinary kriging, malaria risk map.*

1. INTRODUCTION

Malaria is one of the most prevalent and devastating health problems. It has wide global distribution, although its risk and disease burden have always been geographically specific probably biased towards Sub-Saharan Africa (SSA) (Okafor & Amzat, 2007; WHO, 2008). Malaria is a vector-borne infectious disease mainly transmitted by a bite of female *Anopheles* mosquitoes. In Ghana the entire population is at risk of the disease with the most vulnerable being children under 5 years of age and pregnant women. Malaria is the leading cause of morbidity in Ghana where 3.1- 3.5 million clinical cases of the disease (38% of all out-patient illness) are reported in public health facilities each year (GHS, 2007; NMCP, 2008). The National Malaria Control Programme (NMCP)'s campaign against malaria in Ghana has been constrained by limited geographical and financial access to basic health services and inefficiency in resource allocations. In spite of its efforts to rapidly scale-up the intervention activities, the number of new incidences reported is alarming. This then calls for a comprehensive review of health service interventions to determine if they are contributing to the aim of a 50% reduction in malaria cases.

The malaria disease is preventable and curable. However, in order to prevent or cure the disease in the context of its epidemics , decision-makers need to be aware of the risk the epidemic in space and time (Thomson & Connor, 2001). Understanding spatial and temporal distributions of a disease is often accomplished by applying statistical methods to surveillance data and generating a map that describes the variations in risk (Osnas et al., 2009). Spatial statistics provides tools to analyse spatially and/or temporally distributed data, capitalising on the correlation between incidence to interpolate and delineate areas with high disease risk. Geostatistics is a powerful spatial technology which contributes immensely to prediction of random process distributed in space or time (Ali et al., 2006), and has increasingly been applied in epidemiological studies facilitating quantification of spatial features of disease's transmission and its interpolation within the environment (Saxena et al., 2009). It is based on the theory of a regionalised variable distributed in space and assumes spatial auto-correlation such that samples closer in space are more alike than those further apart (Isaaks & Srivastava, 1989). Disease mapping using interpolation estimates of disease occurrence from a regional database to a continuous surface has been undertakenusing the geostatistical methodology of kriging. Kleinschmidt et al. (2000) showed that kriging improves the prediction of malaria risk of children at the local level in Mali. Gething et al. (2006) use space-time time kriging to predict malaria cases at health facilities in Kenya where monthly records were missing in order to get reliable estimates of outpatient malaria treatment burdens for accurate quantification for health system planning.

Malaria transmission in Ghana is widely attributed to mass mosquito breeding arising from varied environmental conditions and lack of effective prevention services. Most available health research in Ghana focus solely on biological aspects and characteristics of the people contracting the disease (Obiri-Danso et al., 2003), neglecting the spatial as well as temporal patterns of the incidence. Though these studies are useful, they are unable to identify areas with high prevalence for policy interventions (Osei & Duker, 2008). The annual risk measures reported on Ghana were based on household surveys at limited sentinel sites and health information systems. Such surveys provided inadequate data for the accurate estimation of new incidence cases and the pace and extent at which they occur in various parts of the country. This poses a challenge to the planning and management of health care delivery system. Spatio-temporal modelling coupled with adequate data have shown to better define the public burden of the disease, providing risk maps to describe the incidence variation in space and time and also identifying high risk areas for health policy decisions. We apply geostatistical methodology of space-time ordinary kriging to malaria morbidity cases reported on district health facilities in Ghana with the aim of producing evidence-based malaria risk maps, describing the risk pattern over space and time to assist malaria intervention resource allocation.

2. SPATIO-TEMPORAL MODELLING

Geostatistical kriging techniques are used to produce prediction and error surfaces. These techniques assume some spatial correlations of the natural phenomena can be governed by random processes indicating presence of a certain degree of interdependence between the values of the attribute at different geographic scales. Kriging techniques include ordinary kriging (OK), the most commonly used. The OK interpolation method involves two steps: quantifying the spatial structure of the data and producing a prediction. The former involves fitting

parameter-model to the incidence data (variography) considered to be a random process with a stationary mean, in the case of heterogeneity, and the latter is to predict an unsampled location using the fitted model obtained from the variography, spatial data configuration and the values of the measured sample points around the prediction location are used (Gething et al., 2006).

Let $\{Z(u_\alpha); \alpha = 1, 2, ..., n\}$ be a set of spatial data of an attribute (malaria incidence), Z at n locations across a study area say, $A \subset R^d$, where $u = (U, V)$ is a coordinate vector and denote Z by the associated random function whose values are the random variables $Z : A \rightarrow \{Z(u) : u \in A\}$. To predict values of Z at m unsampled locations $\{u_p; p = 1, 2, ..., m\}$, denoted $\hat{Z}(u_p)$, we exploit the spatial correlation between dispersed values $Z(u_\alpha)$ and make these unobserved predictions. This requires estimates of the covariance between values of Z separated by different spatial lags (h), vectors of distance and direction. These estimates are provided by calculating the covariance or more commonly, the semivariance, between data pairs at a series of regular lags and fitting a continuous model to these semivariances. The variogram model $\gamma(h)$ then provides the semivariance values at any given lag for input into the kriging process. To account for spatial and temporal distributions in the data, we follow (Kyriakidis & Journel, 1999) and treat time (t) as an additional coordinate to obtain the space-time models, (1) and (2), where each location is referenced by its spatial (u_α) and temporal (t_i) locations to obtain the data $\{Z(u_\alpha, t_i); \alpha = 1, 2, ..., n; i = 1, 2, ..., T\}$ and variogram $\gamma(h_s, h_t)$, estimated as half the mean squared difference between data separated by a given spatial and temporal lag (h_s, h_t):

$$\hat{\gamma}(h_s, h_t) = (1/2n_B) \sum_B \{Z(u_\alpha, t_i) - Z(u_\alpha + h_s, t_\alpha + h_t)\}^2, \qquad (1)$$

where $B = \{[(u_\alpha, t_i), (u_\beta, t_s)] | |u_\alpha - t_\beta| = h_s; |t_s - t_i| = h_t\}$ and n_B is the number of elements in set B. The most commonly used ordinary kriging predictor assumes a constant but unknown mean. In space-time system, we have space-time ordinary kriging (STOK) which predicts the value of Z at (u, t) as a linear combination of the $n(u, t)$ data in space and time:

$$\hat{Z}_{stot}(u_p, t_p) = \sum_{i=1}^{T} \sum_{\alpha=1}^{n} w_{\alpha i} Z(u_\alpha, t_i), \text{ such that } \sum_{i=1}^{T} \sum_{\alpha=1}^{n} w_{\alpha i} = 1, \qquad (2)$$

where the weights $w_{i\alpha}$ are estimated by minimising the error variance, $\sigma_{stok}^2(u, t) = Var[(\hat{Z}(u, t) - Z(u, t)]^2$ subject to the condition $\sum_{i=1}^{T} \sum_{\alpha=1}^{n} w_{\alpha i} = 1$. However, if the observed data $\{Z(u_\alpha, t_i); \alpha = 1, 2, ..., n; i = 1, 2, ..., T\}$ follow the lognormal distribution, then the lognormal ordinary kriging (LOK) serves an alternative to OK to offer a better estimation (Journel & Huijbregts, 1978). Equation (2) is therefore modified for the interpolation of $Z(u, t)$ using the space-time lognormal ordinary kriging (STLOK) (Journel, 1980):

$$\hat{Z}_{stlok}(u_p, t_p) = \exp[\hat{Y}_{stlok}(u_p, t_p) + 0.5\sigma_{stlok}^2 - \lambda], \qquad (3)$$

where $Y(u_\alpha, t_i)$ is the log-transform of $Z(u_\alpha, t_i)$, σ_{stlok}^2 is the kriging variance, λ is Langrage multiplier arising from (2), and the term $0.5\sigma_{stok}^2 - \lambda$ is the non-bias term according to Journel (1980).

3. METHODS

3.1 Study Area and Data

Ghana is a tropical country in West Africa, lying within latitudes 5 and 11 degrees north and covering a total land surface area of about 238,500 km^2 with a coastline of 539,537 km. It shares borders with Cote d'Ivoire to the west, Burkina Faso to the north, Togo to the east; and to the south are the Gulf of Guinea and the Atlantic Ocean. The terrain is mostly low plain, with dissected plateau in the south-central area and the elevation ranges up to 883 metres. The country has an estimated population of 23.4 million inhabitants and administratively divided into 10

regions, each also subdivided into decentralised districts to ensure efficient and effective administration at local level. Data on malaria morbidity cases were obtained from the Centre for Health Information and Management (CHIM) of Ghana Health Services (GHS). The data included monthly morbidity cases reported on health facilities at outpatient departments (OPD) of the 138 districts covering the whole country for the period 1998-2010, comprising 132 (or 156) months with several missing-month records. Morbidity case is defined as a visit at an OPD presumably diagnosed of malaria. In Ghana, most outpatient cases are diagnosed on the basis of clinical symptoms and treatment is presumptive, rather than based on laboratory confirmation (Adams et al., 2004). We also collected population data for the same period from Ghana Statistical Service (GSS) to compute the monthly malaria incidence rates (MIR): $\{I_{it}; i = 1, 2, ..., n; t = 1, 2, ..., T\}$, as defined in Section 3.2.

3.2   Methodology

Ghana's whole population is at risk of malaria and in view of the high incidence of the disease in each district we chose to model the monthly case counts as incidence rates, obtaining the time series $\{I_{it}\}$ for each district. The incidence rate is defined as the number of reported morbidity cases per unit resident population of 10,000 based on (PSS/GSS, 2009): $I_{it} = (x_{it}/P_{it}) \times 10,000, i = 1, 2, ..., n; t = 1, 2, ..., T$, where x_{it} is the reported malaria morbidity cases at district i for month t, P_t is the mid-year population for month t, $n = 138$ is number of districts observed, and $T = 156$ is the number of consecutive months x_{it} is reported within 1998-2010. The district-month observed MIR data, $\{I(u_\alpha, t_i); \alpha = 1, 2, ..., n; i = 1, 2, ..., T\}$ produced 14,879 sampled locations, representing 79.7% of the entire data locations. Initial exploration of the data showed a highly positively skewed distribution which was log-transformed to reduce some amount of noise (see Figure 1 and Table 2) and also allow the STLOK to be used. The log-transformed data, $Z(u_\alpha, t_i) = In[I(u_\alpha, t_i)]$ were used to obtain a 3-D experimental variogram following (1), with the spatial coordinates u being *easting* (X) and *northing* (Y), and *time* (Z), in months, as the third dimension. Figure 2 gives the global window of the space-time experimental variogram as well as the fitted to it models showing the spatial (N0) and temporal (D-90) dependence of malaria incidence rates at local levels. The experimental variogram $\hat{\gamma}(h_s, h_t)$ was calculated using a tolerance of 50%, lag spacing of 30 km for 15 lags and 80 months for 1-month lag spacing for the spatial and temporal directions respectively. The experimental semivariogram in the spatial domain was then modelled as isotropic with a nugget effect, two spherical functions for the spatial dependence whilst the temporal autocorrelation was fitted with same nugget, a spherical model and exponential-cosine function of period 12. These models are defined by (4), (5) and (6) respectively:

$$\gamma_{nug}(h) = s_0, \textit{ for } h > 0; \textit{ and } 0, \textit{ for } h = 0 \tag{4}$$

$$\gamma_{sph}(h) = s\{(3h/2r) - (h^3/2r^3), \textit{ for } 0 \leq h \leq r; \textit{ and } 1, \textit{ for } h > r \tag{5}$$

$$\gamma_{\exp\cos}(h) = s\{1 - \exp(-3h/r)\cos(bh), 0 \leq h \leq r; \textit{ and } 1, \textit{ for } h > r, \textit{ where } b = 2\pi/\lambda, \tag{6}$$

where s_0 is the nugget, h is the lag distance, s is the sill, r is the range, b is called angular frequency and λ is the period or wavelength of the exponential-cosine function. A grid of $10km \times 10km \times 156$ cells covering the whole Ghana was created and STLOK employed using (2) and (3) to interpolate $\hat{Z}_{stlok}(u,t)$ and then back-transformed respectively, yielding the estimated MIR values, $\hat{I}_{stlok}(u,t)$ at the unsampled locations and monthly malaria risk maps in Ghana. The variography and kriging of the MIR data were all implemented using the geostatistcal package ISATIS 10.

4.   RESULTS

The log-transformed observed MIR data at district-month locations are approximately normally distributed, justifying the use of the lognormal kriging, though its use had often been criticised in the literature (Journel, 1980; Yamamoto, 2008). Figure 1 shows the study area indicating the sampled data locations and histograms of log-transformed data and the kriging estimates. The modelled space-time variogram using the parametric models (4) - (6) was characterised with both short and long ranges. The spatial experimental semivariogram modelled as isotropic exhibits spatial correlations ranging 35–295 km with a maximum sill of 0.275 whilst the temporal

correlations have a sill of 0.235 within period 95 – 1000 months (Figure 2). The temporal component of the space-time variogram exhibited an increasing seasonal pattern of the MIR (with a period of 12 months) until it attains its shortest range of 95 months compared with the spatial correlations which exist at distance of 35 km. The model parameters, as detailed in Table 1, have a nugget effect of 0.035, indicating both spatial and temporal dependence of the malaria incidence at the districts.
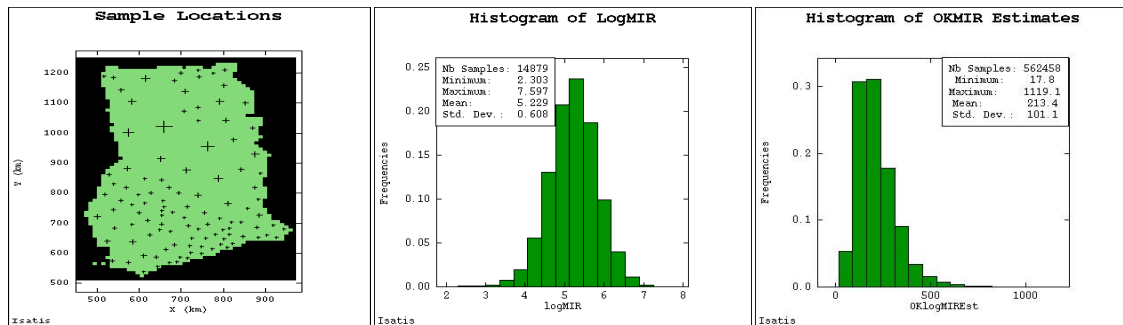

Figure 1. Map of Ghana showing sampled locations (left), histograms of log-transformed MIR (middle) and STLOK estimates of MIR (right) together with their summary statistics.
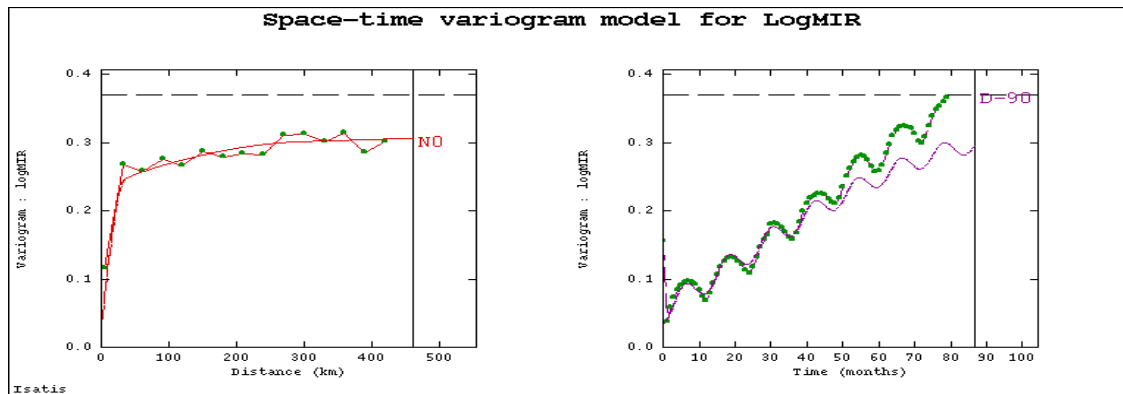

Figure 2. Space-time experimental variogram (red dotted green) fitted with spherical and periodic functions showing spatial (left) and temporal (right) dependence of malaria incidence rates in log-transformed scale.

Table 1. Space-time variogram model parameters for the malaria incidence rates correlations at sampled locations in Ghana.

Structure	Model $\gamma(h)$	Sill s_1	Range r			Period λ	Angular freq
			Easting (km)	Northing (km)	Time (months)	(months)	b
1	Nugget s_0	0.035	-	-	-	-	-
2	Spherical	0.190	35	35	95	-	-
3	Spherical	0.050	295	295	500		-
4	Exp-cosine	0.020	90	90	1000	12	0.524
5	Spherical	0.180	100000	100000	500	-	-

The grid cells created to interpolate health facility-month locations where data were missing generated maps for malaria risk for all the 156 months. The results of malaria incidence in Ghana showed spatial heterogeneity with elevated cases and also increasing over time. Four of these maps for the months: March, June, September and December in 2010 have been selected and shown in Figure 3. High malaria incidence rates are estimated in the northern most and western extending to central parts of the country and lowest in various areas in the north and south, mostly along the coast. Table 2 presents the summary statistics for both the observed samples and predicted values at the unsampled locations. The STLOK predicted a mean of 213.36, variance 10209.26 and skewness 1.318 compared with the observed samples which were more positively skewed (2.024) with a mean of 223.21 and variance 20525.60.
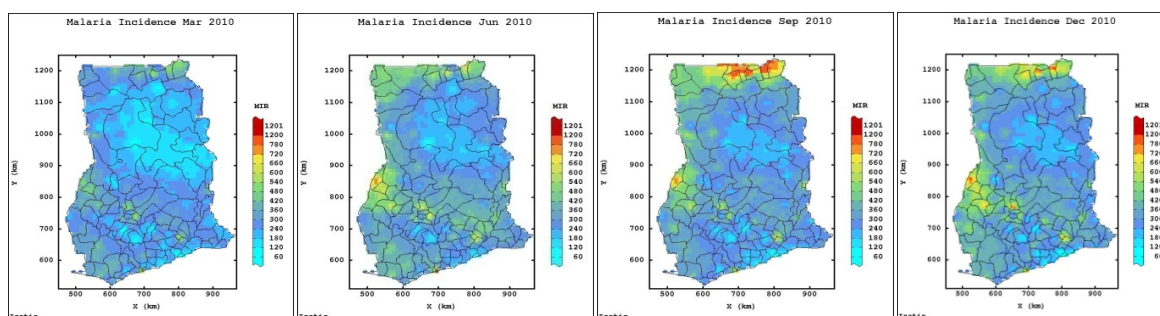
Figure 3. Maps of malaria incidence rates estimated by STLOK for the selected months: March, June, September and December in 2010, indicating high prevalence in the northern and central parts.

Table 2. Summary statistics for the sampled malaria incidence rates and space-time ordinary log normal kriging (STLOK) estimates

Estimator	Min	Max	Mean	Var	CV	Q1	Median	Q3	Skew
STLOK	17.77	1119.07	213.36	10209.26	0.474	142.18	193.35	263.44	1.318
Sample	10.00	1993.00	223.21	20525.60	0.642	126.00	189.0	281.00	2.024

5.   CONCLUSIONS

Baseline data and reliable monitoring of key malaria indicators are needed to measure whether the goals for morbidity and mortality reduction are achieved. Data from health facilities are potentially useful for monitoring malaria patterns and trends, but have several limitations. The crucial factor affecting the representativeness of hospital data is the extent to which the hospital statistics reflect the burden of disease in the population. Severe malaria cases are only included in the hospital statistics that are also affected by the accessibility of hospital services and by the health care seeking behaviour of the population. These raise concerns of under and over reporting for data captured in the database of CHIM which this study's results are based. Despite these limitations, hospital statistics may be acceptable for monitoring trends if their low sensitivity remains consistent over time (Adams et al., 2004). Spatial maps of malaria morbidity provide useful analytical tools for resource allocation to reduce the disease at the local level. In this study we developed a geostatistical space-time model to describe the correlation structure of malaria incidence at locations in districts and over time with the aim of producing monthly malaria risk maps to predict locations where data were not available and also highlight areas with "hot spots" and at what time.

The results indicated varied spatial and temporal distributions of the disease across the country with elevated cases mostly in the northern most and central parts of the country. Generally, the "hot spots" border with the neighbouring countries, Burkina Faso and Cote d'voire, which equally record high malaria incidence (WHO, 2008). The frequent population movements across the borders put susceptible resident locals at risk of the disease. Baragatti et al. (2009) observe travels outside the city area increases the risk of being exposed to malaria transmission in their spatial analysis of malaria epidemiology in Burkina Faso. Most malaria intervention activities in Ghana are piloted in suspected areas of high record of the disease. These studies include Apawu et al. (2004), Baird et al. (2002) and Binka, et al. (1998), one of which estimates malaria as accounting for over 25% of under-five mortality in Northern Ghana. Efforts to intensify malaria campaigns at these targeted areas will surely reduce malaria burden in the country. The spatial heterogeneity of the observed malaria incidence may also be due to the spatial and temporal distributions of habitats of vector mosquitoes which differ according to the conditions of the local environment. The study focused attention only on spatial and temporal structure of malaria incidence rates without correlating with other possible covariates. A better way of statistically analysing the data might be to correlate the MIR with other possible covariates, taking account of spatio-temporal correlations. However, this effort may be hampered by the lack of co-location of the covariates with the malaria incidence. We seek to undertake this comprehensive approach in a further study. Notwithstanding, we have provided visual description of geographic distribution of malaria morbidity burden and its dynamics with time over the country. This provides useful analytical tools for resource allocation to reduce the disease at the local level in Ghana and also serves as a baseline for further research of smoothed map techniques.

ACKNOWLEDGEMENTS

REFERENCES

Adams, I., Darko, D., & Accorsi, S. (2004). Malaria: A Burden Explored. *Bulletin of Health Information, 1(1),* 28-34.

Ali, M., Goovaerts, P., Nazia, N., Haq, M. Z., Yunus, M., & Emch, M. (2006). Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. *International Journal of Health Geograghics, 5:45.*

Apawu, M., Owusu-Agyei, S., Dadzie, S., Asoala, V., Anto, F., Koram, K., et al. (2004). Malaria transmission dynamission at a site in northern Ghana proposed for testing malaria vaccines. *Tropical Medicine and International Health, 9*, 164-170.

Baird, J. K., Agyei, S. O., Utz, G. C., Koram, S. L., Barcus, M. J., Jones, T. R., et al. (2002). Seasonal malaria attacks rates in infants and young children in Northern Ghana. *The American Society of Tropical Medicine and Hygiene, 66*(3), 280-286.

Baragatti, M., Fournet, F., Henry, M., Assi, S., Ouedraogo, H., Rogier, C., et al. (2009). Social and environmental malaria risk factors in urban areas of Ouagadougou, Burkina Faso. *Malaria Journal, 8*(13).

Binka, F. N., Indome, F., & Smith, T. (1998). Impact of spatial distribution of permethrin-inpregnated bed nets on child mortality in Northern Ghana. *American Journal of Tropical Medicine and Hygiene, 59*, 80-85.

Gething, P. W., Noor, A. M., Gikandi, P. W., Ogara, E., et al. (2006). Improving imperfect data from health management information systems in Africa using space-time geostatistics. *PLoS Medicine 3*(6), 271.

GHS. (2007). *Annual Report of Ghana Health Service 2007*. Accra: Ghana Health Service (GHS).

Isaaks, E. H., & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press, Inc.

Journel, A. G. (1980). The lognormal approach to predict local distributions of selective mining unit grades. *Mathematical Geology, 12*, 285-303.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining Geostistical*. New York: Academic Press.

Kleinschmidt, I., Bagayoko, M., Clarke, G. P. Y., Craig, M., & Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology, 29*(2), 355-361.

Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical space-time methods: a review. *Mathematical Geology, 31*, 651-684.

NMCP. (2008). Lunch of Ghana Malaria Advocacy Campaign by National Malaria Control Programme (NMCP)/JHU/CCP. *Malaria Watch, 5,* 9.

Obiri-Danso, K., Okore-Hanson, A., & Jones, K. (2003). The microbiological quality of drinking water sold on the streets in Kumasi, Ghana. *Lett. Appl. Microbiology, 37*, 334-335.

Okafor, E. E., & Amzat, J. (2007). Problems of Malaria Menace and Behavioural Intervention for its Management in Sub-Saharan Africa. *Journal of Human and Ecology, 21*(2), 155-162.

Osei, F. B., & Duker, A. A. (2008). Spatial and demographic patterns of Cholera in Ashanti Region-Ghana. *International Journal of Health Geographics, 7.44.*

Osnas, E. E., Heisey, D. M., Rolley, R. E., & Samuel, M. D. (2009). Spatial and temporal patterns of chronic wasting disease: fine-scale mapping of a wildlife epidemic in Wiscosin *Ecol. Appl., 19*(5), 1311-1322.

PSS/GSS. (2009). *Population Projections Release 2009* Accra: Population Statistics Section (PSS), Ghana Statistical Service (GSS).

Saxena, R., Nagpal, B. N., Srivastava, A., Gupta, S. K., & A.P., D. (2009). Application of spatial technology in malaria research and control: some new insights. *Indian Journal of Medical Research, 130,* 125-132.

Thomson, M. C., & Connor, S. J. (2001). A framework for field research in Africa in Malaria Early Warning System (MEWS): Concepts, Indicators and Partners: Roll Back Malaria/World Health Organization.

WHO. (2008). *World Malaria Report 2008*. Geneva, Switzerland: World Health Organisation (WHO).

Yamamoto. (2008, June 25-27). *Assessing the uncertainties for lognormal kriging estimates.* Paper presented at the 8th International symposium on spatial accuracy assessment in natural resources and environmental sciences, Shanghai, P.R. China.