Spatial Modeling of Diabetes Cases in Ghana

Oti-Boateng Emmanuel[1], Ngesa Oscar[2], Osei Frank Badu[3]

[1]Pan African University of Basic Science, Technology and Innovation., P. O. Box 62000-00200, Nairobi-Kenya
Email: oti-boateng[at]aims-cameroon.org

[2]Taita-Taveta University college, Taita-Taveta-Kenya
Email: oscanges[a]ttuc.ac.ke

[3]University of Energy and Natural Resources, Sunyani-Ghana
Email: oseibadu2004[at]yahoo.co.uk

Abstract: *Diabetes is a continual daunting menace in Ghana. Although policy makers keep investing in areas to fight the disease yet the number keeps increasing. This brings out the question of whether monies are channeled to the truly needed districts. In this work, a disease map is plotted showing the variations of the relative risk of the disease in all district of Ghana. This maps are critical and informative to policy makers. This allows them to target policies and use the already meager resources well.*

Keywords: Disease mapping, Bayesian analysis, Markov Chain Monte Carlo(MCMC) Spatial analysis, Deviance Information Criteria (DIC)

1. Introduction

1.1 Introduction

Spatial epidemiology comes into play when the main aim is to identify whether or not there is any form of spatial or geographical variations in our data with respect to some identified disease. The incorporation of spatial elements redirects the way epidemiologist understand diseases especially in relation to the environment [18]. This technique dates back from the time of John Snow, who employed this technique in identifying a bore hole contributing to the spread of cholera in his locality [29].In this field , data comes in many forms, however, our main target will be the lattice data which normally appear as count data. With the use of this new technique data data, epidemiologists are able to identify the distribution of some named disease in the form of disease mapping. To do this, the relative risk of the disease in the geographical locations is modeled and in most cases appears in the form of a regression model with some distributional assumptions.

Normally, in the case of count data, a special type of regression model is used called Generalized Linear Model since the count data does not satisfy normality. The inception of the Geographical Information Systems has helped to change the way scientist look at disease mapping ([29]). The 1792 mapping of yellow fever and the 1854 cholera endemic triggered the idea of disease susceptibility mapping. Yellow fever and cholera were mapped in the year 1792 and 1854 respectively and this fueled the idea of disease susceptibility mapping [19]. Disease maps comes in many forms, but in this case, the interest is in the quantitative display of the population at risk and infection prevalence [4, 25]). This will do by using spatial statistics which involves the use of location data in analysis ([25]). There is an additional advantage of filtering noise from the data [8]. Among those who used this method are, [?] connected the changes in etiological factors (environmental variables) to a diseases. Also [7] hierarchically modeled fasciolosis . [?]) presented a geographical map of malaria and identified some of the important environmental factors

of the disease in Sistan and Baluchistan province, Iran , ([27]) applied disease mapping to infectious diseases when a primary case can result in secondary cases, by direct or vector transmission whiles ([24]) used an alternative model i.e. the Gaussian Component Mixture (GCM) model instead of the proper or improper CAR in disease modeling .

Diabetes is a major public health issues in terms of both morbidity and mortality. Diabetes is currently at the epidemic level with 70 percent of those infected living in low and middle income economies([30]). About 87 million people have diabetes in the world and more than 22 million people in the Africa Region; by 2035 this figure will almost double ([15]). Prevalence in Africa as at 2007 was nearing the 10.4 million mark, World Health Organization ([30])). Ghana is one of the 32 countries of the (International Diabetes Foundation for Africa) IDFAFR region. Ghana has its fair share of diabetes 1 [1 which is a group of metabolic diseases in which a person has high blood sugar, a condition which could be attributed to a situation where either the body does not produce enough insulin or because cells do not respond to the insulin that is produced; but it could be controlled and managed with little injections of insulin.] mellitus at monumental score of 450,000 cases of diabetes in 2014 and cost per person with diabetes stands at 148.8USD([15]).

Several studies have been done on the subject, however, a nationwide comprehensive work is yet to be done considering cases of diabetes. This adds to the uniqueness of the work. The purpose of this study is concerned with carrying out a spatial model of diabetes cases in Ghana which will be applied in mapping, to available data on the non-infectious diseases. In our case we wish to map the variation of diabetes in Ghana based on the assumption that disease risk varies solely on spatial factors. Stage 2 which is the methodology spells out the mathematical methods used and the method of estimation for contributing parameter, whiles stage 3 involves the display of results in the form of tables and maps and then stage 4 includes discussion and conclusion.

2. Methodology

Count data for validation of our model was retrieved from Ghana Health Service. Data is collected and collated by the district in each region by the District Health Directorates . The corresponding population counts were also retrieved from the Ghana Statistical Service, whose activities are clearly spelt out and defined at the policy level. There was no problem with data in terms missing data or unavailability.

From these collated data, we represent our counts data as vector, $Y = (y_1, y_2, \cdots, N)$. The formula for the expected rate was similar to the one adopted by ([3]), where n_i is the given population of district i . The expected rate was said to represent the population effect for the district in question. Then, Poisson model can be used to represent the count suppose that we want to measure the variation of the disease in Ghana with the aim of identifying the districts with the highest and lowest risk.

Assuming that each individual case is reported, then we present, where n_i is the given population of district i . The expected rate was said to represent the population effect for the district in question. Then, Poisson model can be used to represent the count suppose that we want to measure the variation of the disease in Ghana with the aim of identifying the districts with the highest and lowest risk. Assuming that each individual case is reported, then we present

$$P(Y = y_i) = \frac{\mu_i e^{-\mu_i}}{y_!}, \qquad (1)$$

where y_i can take values from zero and above, and $\mu_i = \theta_i e_i$. In this case, μ_i is the parameter for the Poisson distribution representing the average count per interval and

$$\theta_i = \alpha_0 + \beta\mathbf{X}' + u_i + v_i, \qquad (2)$$

is the relative risk which is expressed in terms of the covariates and the special effects. In our case, we omit the covariates for the assumption that, the relative , θ_i, only varies spatially for all units and these variations are peculiar to each district. It is just okay, if we represent u_i and v_i to be correlated and uncorrelated spatial effects respectively.

Here , d_i and N_i represent the number of neighboring districts respectively k_i, σ_i^2 are unknown and they will have to be determined by choosing appropriate hyper parameters when using Bayesian method. Interestingly, Bayesian method originates from Bayes theorem which is more of a conditional probability where the conditional probability of A given B is the the conditional probability of B given A scaled to one by the conditional

probability of A compared to B [17].

The advantage this method of estimation brings to board is the information it tells us about the the unknown parameter unlike the frequentist. For example, the P-value and confidence interval does not measure what is known about the parameters [7, 12, 17]. With Bayesian inference, we are able to take advantage of the prior i.e. what we know about the parameter when we have no data, $\pi(\beta)$.

To achieve these, the unknown parameters are assigned non-informative prior to depict the minimal information we have about them. We adopt the Gamma distribution to model precision parameters (inverse of variance parameters) with known hyper parameters α_i and δ_i with expectation $\frac{\alpha_i}{\delta_i}$ and variance $\frac{\alpha_i}{\delta_i^2}$. The likelihood then comes in which involves how likely the data Y are given the parameter in question, we denote this by $f(Y | \beta)$. So our posterior, according to Rev. Bayes, is proportional to the product of the prior and the likelihood; written as; $p(\beta | Y) \propto \pi(\beta)f(Y | \beta)$. In our cases, we employ the very know Besag York and Mollie Model for the computation of the prior of the spatial effects. With the fact that, one of the spatial effects always dominating the other,[2] assumed that u and v were independent with the following priors as used in ([20]): They are;

$$p(u_i | u_{-i}) : N\left(\frac{\sum_{j \in N(i)} u_j}{d_i}, d_i^{-1} k_i\right), \qquad (3)$$

whiles,

$$v_i : N(0, \sigma_i^2) \qquad (4)$$

Going by the above, we can compute the posterior of the unknowns , u_i, v_i as,

$$= p(u_i | Y) \propto \pi(u_i)f(Y | u_i) \qquad (5)$$

$$= exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{n}(\theta_i - \psi_i)\right\}.exp\left\{-\frac{1}{2}\sum_{i}\sum_{j \in N(j)}(u_i - u_j)^2\right\}$$
$$(6)$$

That of v_i can be written as;

$$\propto exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{n}(\theta_i - \psi_i)^2\right\}.exp\left\{-\frac{1}{2\sigma_i}\sum_{i=1}^{n}v_i^2\right\}. \qquad (7)$$

The Posterior for the offset, α_0,

$$p(\alpha_0 \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \alpha_0, \sigma^2) p(\beta \mid \sigma^2) p(\sigma^2) \quad (8)$$

The posterior for σ^2 is,

$$p(\tau \mid å) = \omega(\tau \mid \alpha, \delta) \omega(\tau), \quad (9)$$

$$\propto (\tau)^{-(\alpha+1)} exp\left(-\frac{\delta}{\tau}\right), \quad (10)$$

where α and δ are hyper parameters of the gamma distribution represented by $å$.

WinBugs 1.4 was used in the computation of the results and analysis ([28]). Winbugs is a statistical program specialized in Bayesian analysis. This time it employs techniques in Markov Chain Monte Carlo (MCMC). MCMC is one of the suitable methods recommended when computation is complex and involves method such as Gibbs sampling and Metropolis-Hastings algorithm ([26]. Bayesian presents many advantages over the frequentists in the sense that confounding problems are always corrected by the data inclusion.

Model estimation was carried out using a Bayesian approach. To be precise, a non informative normal prior was assigned to the offset parameter, α_0 whiles the variance parameters are assigned inverse gamma distributions. The assumption that covariates are not available was considered. WinBugs version 1.4 was used in the implementation ([26]) phase. A single chain of Markov Chain Monte Carlo (MCMC) iterations of 50,000 was run with a burn in period of 10,000 with every tenth sample value retained for analysis. Every tenth sample value considered for arriving at the convergence of the estimates from the remaining 10,000 samples. The decision on convergence was arrived at based on the behavior of our trace plots (as shown in Figure(2)) and autocorrelation plots of the MCMC output ([12]).

3. Data and Results

Annual diabetes cases in Ghana varies across the entire country with most cases occurring in the regional capital and the lowest occurring in the northern part of Ghana especially with newly created districts.falls within the brackets of one (1) to 29,474 with the former being one of the newly created districts; North Tongu in the Volta Region and the later being Tema-Kpone-Akatanmanso district in the Greater Accra region.With maps of this nature, we are able to understand more, the state of diabetes in all the districts of Ghana which will enhance equitable distribution of resources. In this write-up, emphasis were laid on the spatial variations present in the disease rather than covariates. In the case of the correlated spatial effects, u_i, we had the values falling in the domain of

$(-1.5, 1.5)$, but the (95%) credible interval has most of its values as positive signifying a positive relationship between the relative risk and that parameter. In the case of the upper credible interval for the relative risk map, all components of plotted quantity are the same.

In the case of the uncorrelated spatial effect, we had all the values falling within the interval of $(-0.02669, 2.965)$ with most of the 95% credible interval falling within positive, signifying a positive correlation between the the relative risk and the parameter. The case of the offset however was the opposite with negative values in the mean, and the credible intervals telling us of the negative correlation between it and the relative risk.

In Figure (1), it is easily seen that, majority of the districts are at a higher risk with most of them conspicuously present in the southern sector with Bosomtwe (Ashanti Region) standing as the highest followed by Agona East (Central Region) and then Ejuara-Sekyeredumase (Ashanti Region). On the other hand Upper Manya Krobo, Ga East and Savelugu-Nanton are among the lowest risk districts. From Table (0), it can be deduced that v_i is more influential than u_i as it contributes more to the determination of the relative risk. The bigger the value of v_i the more relative risk in the district and vice versa.

4. Discussion

In this write-up, we show the disease variation with respect to every district by way of a disease map. The relative risk is clearly captured based on the assumption that, it varies spatially which was modeled using the known Conditional Autoregressive (CAR) model proposed by ([2]).

WinBugs software was used in the implementation phase where the offset parameters and hyper parameters were assigned non-informatic priors.

The model showed spatial variation for diabetes for all districts in the way of a smooth map and this is an efficient tool for advising policy makers when tackling the disease in the country. Bosomtwe (Ashanti Region) standing as the highest followed by (Central Region) and then (Ashanti Region).

Based on data acquired, those in Bosomtwe (Ashanti), Agona East (Central) and Ejura-Sekyedumase (Ashanti) are at the highest risk of the disease. These results make sense when subjected to the work done by ([5]) who identifies the Central, Brong Ahafo and the Ashanti Regions of Ghana as some of the regions with high cases of the disease. The paper i.e. ([5]), stated that the average risk of these regions are higher than that of the world.

[5, 9] in their works confirmed the unrelenting increase of diabetes in Ghana with [5, 14] connecting the excessive consumption of carbohydrates to the increase in diabetes. These findings make a lot of sense as fufu (from cassava), banku (Maize), yam, rice, etc. form the basis for almost all

foods in the regions [22]. Also, one of the reasons for the high risk could be as a result of urbanization which is in line with the findings of ([1]), where urbanization was accompanied with increase in income and subsequently increase in car ownership and consumption of fatty foods. One of the reasons contributing to the upward surge in the population is the absence of leisure activities as stated by ([5]).

On the other side of the coin are Upper Manya Krobo, Ga East+La-Nkwanta and Savelugu-Nanton, which recorded the lowest risk in ascending order. In the case of the first two, [5] attributed it to the availability of health resources in the districts which is helping to fight the menace. The other districts however is found in the northern part of Ghana. Millet, sorghum, guinea corn and groundnut with small or no protein. Too much protein-contents has been known to trigger diabetes [22]. The last not the least, the people of these regions walk miles to and fro their farms which serves as vigorous exercise for them. [1, 5, 22] has identified

exercising as one of the key therapies against diabetes.

In this work, validation was not based on covariates as the main agenda was to measure the variation of the disease in the country by way districts.

5. Conclusion

In this work, a model was built that has the potential of modeling the relative risk of disease in Ghana. With the above model, disease maps were displayed after a Bayesian method has been used in the parameter estimation. This map is essential for advising policy makers as to which districts to prioritize when distributing mitigating resources. We recommend extending this case to a multiple case where multiple diseases can applied.

Remark 2 This is an example of figure 1.

Table 1: Comparison of Count Models in Ghana

| Results of some random districts			
Model	Relative Risk	u_i	v_i
Upper Manya-Krobo	$0.00292(-0.008464, 0.01482)$	$-1.884(-3.366, -0.5384)$	$-2.993(-4.682, -1.381)$
Ga East + La-Nkwantanang-Madina	$0.0057(-0.006225, 0.01732)$	$-1.677(-2.916, -0.5842)$	$-2.407(-3.635, -0.9156)$
Savelugu-Nanton	$0.005854(-0.00688, 0.1924)$	$-1.706(-3.707, -0.2578)$	$1.275(0.2702, 1.743)$
Ejura-Sekyedumase	$6.369(6.184, 6.561)$	$-0.2397(-1.23, 0.4419)$	$3.218(2.387, 4.255)$
Agona East	$6.814(6.648, 6.986)$	$1.501(1.042, 2.691)$	$1.544(0.2132, 1.8913)$
Bosomtwe	$25.26(24.69, 25.84)$	$1.391(0.8509, 2.543)$	$2.965(1.709, 3.414)$
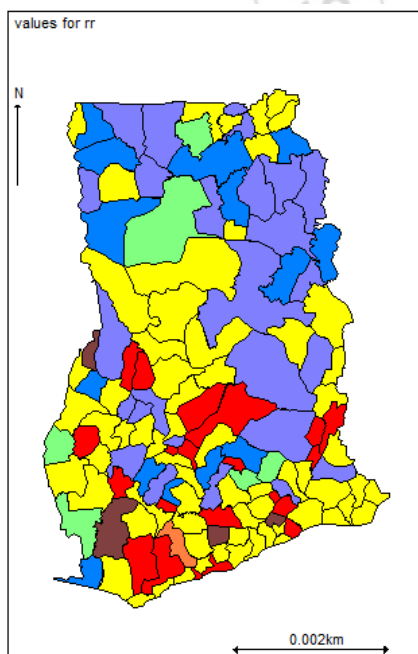
Figure 1: A map of relative risk of diabetes cases in Ghana.
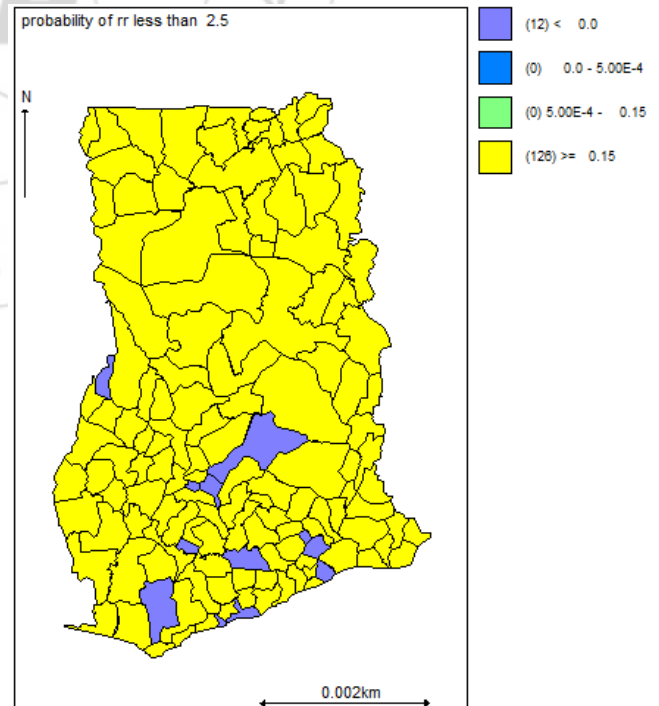

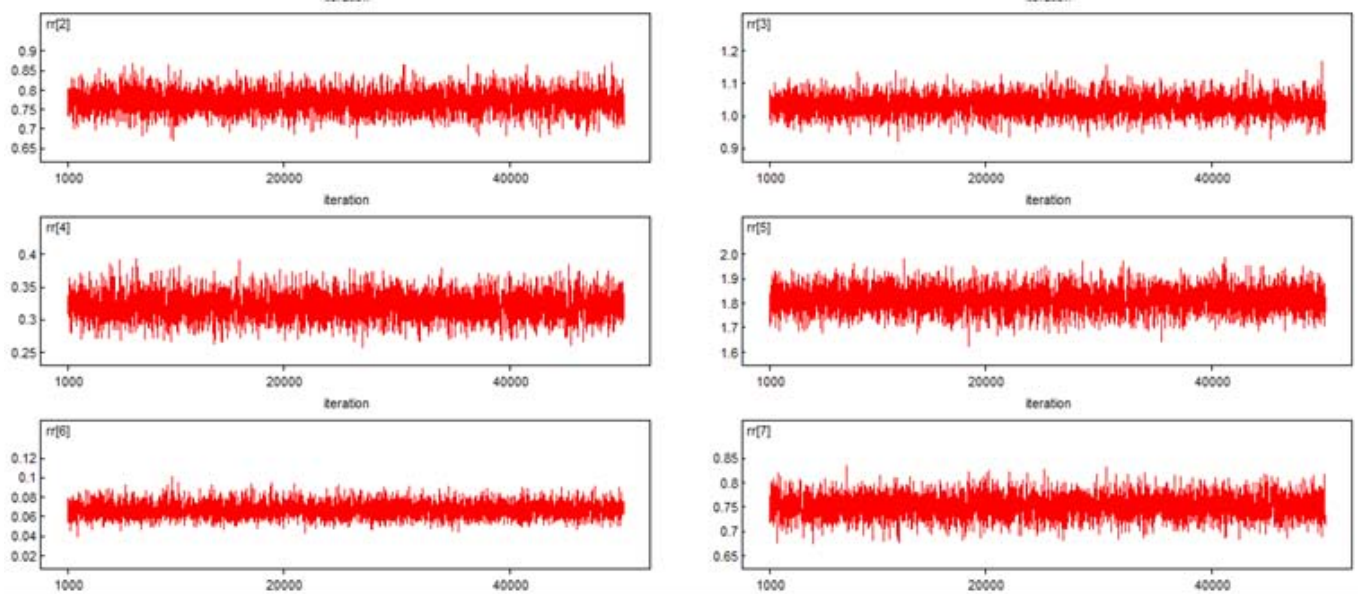**Figure 2:** The 2.5 percent upper credible limits maps

Figure 3: Random selection of trace plot of the Relative Risk

References

[1] Al-Lawati, J. , Jousilahti, P.(2004)., Prevalence and 10-year Secular Trend of Obesity in Oman, Saudi Medical Journal, 25, 346-351.

[2] Besag J and York J and Mollie A, Bayesian image restoration with two applications in spatial statistics (with discussion, Ann Inst Stat Math, 43(1),(1991).

[3] Briggs H(2001) Spatial analysis, concept and challenges. Center for Disease Control (2008)., Preventing Diseases:Investing wisely in health, ht

[4] Clement, EP, "Small Area Estimation with Application to Disease Mapping", International Journal of Probability and Statistics (2014), 15–22.

[5] Cringoli, G and Rinaldi, L and Veneziano, V and Capelli, G and Malone, JB, "A cross-sectional coprological survey of liver flukes in cattle and sheep from an area of the southern Italian Apennines", Veterinary parasitology (2002), 137–143.

[6] Darkwa Sarah,Prevalence of diabetes mellitus and resources available for its management in the Cape Coast Metropolis,Journal of Health and Environmental Sciences,1(1),1-7,(2011).

[7] Draper David, Bayesian Hierarchical Modeling, (2002)

[8] Durr, PA and Tait, N and Lawson, AB, "Bayesian hierarchical modelling to enhance the epidemiological value of abattoir surveys for bovine fasciolosis", Preventive veterinary medicine (2005), 157–172.

[9] Elliot Paul and Wakefield, Jon C and Best, Nicola G and Briggs, DJ and others, "Spatial epidemiology: methods and applications.", Oxford University Press (2000).

[10] Ernest Attuquaye Quaye and Edward O. Amporful and Patricia Akweongo and Moses K. Aikins, PhD,Analysis of the Financial Cost of Diabetes Mellitus in Four Cocoa Clinics of Ghana,7(1),49-53,(2015).

[11] Gelfand Allan E and Smith Adrian F M, Sampling-based Approaches to Calculating Marginal Densities,Journal of the American Statistical Association, 85, 398-409, (1990)

[12] Gelfand and Smith, Sampling-Based Approaches to Calculating Marginal Densities, Springer Series in Statistics, 85(410), 519-550 ,(1990)

[13] Gelman A. and J.B. Carlin and H.S. Stern and D.B. Rubinn, Bayesian data analysis, Chapman and Hall/CRC,(2003).

[14] Huang Yan and Zhang Liqin and Zhang Pusheng, A Framework for Mining Sequential Patterns from Spatio-Temporal Event Data Sets, (2002)

[15] Ina Danquah and George Bedu-Addo and Karl-Johann Terpe and Frank Micah and Yaw A Amoako and Yaw A Awuku and Ekkehart Dietz and Markus van der Giet and Joachim Spranger and Frank P Mockenhaupt,Diabetes mellitus type 2 in urban Ghana: characteristics and associated factors,BioMed Central,210(3),1471-2458(2012).

[16] International Diabetes Foundation of Africa, Diabetes scorecard in Ghana, IDF Press (2014)

[17] Karsten, Ahnert and Mulansky and Mario,Boost. Numeric. Odeint,(2009)

[18] Ken Rice, "A very brief Introduction to Bayesian Statistics", Preventive veterinary medicine (2014).

[19] Lawson, A. B., "Statistical methods in spatial epidemiology", John Wiley and Sons (2013).

[20] Mohammadreza Rajabi and Ali Mansourian and Petter Pilesjo, "Disease susceptibility mapping using spatial modeling techniques" (2015).

[21] Ngesa Oscar,A Flexible Random Effects Distribution in Disease Mapping Models, South African Statistical Journal,48(1),83-93(2014)

[22] Nkurunziza, Hermenegilde and Gebhardt Albrecht and Pilz Jurgen,Bayeshian modelling of the effect of climate on malaria in Burundi,Malaria Journal,9(114), 1475-2875,(2010)

[23] Ofori-Boadu Agnes R.N., The dawn of cooking, BioMed Central, (2003)

[24] Pace R Kelley,Conditional Autoregressions with Doubly Stochastic Weight Matrices,National Science

Foundation,7(1),(1981)

[25] Paula and Lawson Moragaa and Andrew B,Geographical mapping and Bayesian spatial modeling of malaria incidence in Sistan and Baluchistan province, Iran, Computational Statistics and Data Analysis, 56,985-992,(2012)

[26] Rinaldi, Laura and Musella, Vincenzo and Biggeri, Annibale and Cringoli, Giuseppe, "New insights into the application of geographical information systems and remote sensing in veterinary parasitology", Geospatial Health (2006), 33–47.

[27] Spiegelhalter, D. and A. Thomas, N. Best and D. Lunn,Winbugs user manual version1.4,Journal of the Royal Statistical Soceity,series B,(2002).

[28] Sylvain C. and Charras-Garrido Myriam and Abrial David and Yao-Lafourcade Anne-Franoise, Spatiotemporal disease mapping applied to infectious diseases, 26,32-37, (2015)

[29] Unit, M R C Biostatistics,Notes on WinBugs,(2015)

[30] Waller B and Lance A, Spatial Disease Mapping, Encyclopedia of Life Support Systems, UNESCO Press(2004).

[31] WHO, Risk predictive modeling for diabetes and cardiovascular disease,Bulletin of the World Health Organization,51(1),(2014) Zayeria Farid & Salehib Masoud & Pirhosseinic Hasan (2011),Geographical mapping and Bayesian spatial modeling of malaria incidence in Sistan and Baluchistan province, Iran,Asian Pacific Journal of Tropical Medicine, 4(12),985-992.

Acknowledgements