A spatial model of malaria risk in The Gambia: predicting the impact of insecticide treated and untreated bednets on malaria infection.

M.C. Thomson*, S.J. Connor*, U. D'Alessandro†, B. Rowlingson‡, P. Diggle‡, M.Cresswell*, & B.M. Greenwood§.

* MALSAT, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK.

† Department of Parasitology, Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium

‡ Medical Statistics Unit, Lancaster University, Lancaster, LA1 4YF, UK

§London School of Hygiene and Tropical Medicine, Keppel Street, London WCIE 7HT, UK.

http://www.liv.ac.uk/lstm/malsat.html

Presented at the 10th Colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, 16-19 November, 1998

Abstract

Recent calls to '*Roll back malaria*' propose a renewed, broad partnership initiative to combat the disease at global, regional, national and local levels. This initiative has developed from the new WHO Global Malaria Control Strategy which recognises malaria as a disease of differing epidemiological types which are determined by a diversity of social, ecological and economic settings.

As part of this strategy we have advocated the use of satellite imagery by control services to provide environmental information for malaria stratification, monitoring and early warning[,]. To achieve this operationally appropriate methodologies must be developed for integrating environmental and epidemiological data into models that can be used by decision-makers for improved resource allocation.

Using methodologies developed by the Famine Early Warning Community and spatial statistics we show a significant association between age related malaria infection in Gambian children and the seasonal amount of environmental 'greenness' as measured using the Normalized Difference Vegetation Index? (NDVI) derived from satellite data. The resulting model is used to predict changes in malaria prevalence rates in children resulting from different bednet control scenarios.

1 Introduction

Improved targeting of resources has been an important driving force behind initiatives to map malaria risk in Africa[4,]. In the absence of reliable, continental wide, epidemiological data environmental information on factors such as rainfall and temperature and vegetation 'greenness' have been used to indicate the length of the malaria transmission season in different parts of the continent[,,,].

The potential for using meteorological satellite data such as the polar orbiting and geostationary satellites operated by the National Ocean and Atmospheric Administration (NOAA) of the USA in the study of malaria transmission dynamics has been reviewed in detail elsewhere[7,]. Using epidemiological and entomological data from The Gambia we have shown that even simple analysis of proxy ecological variables derived from satellite data can indicate variation in environmental factors affecting malaria transmission indices[10]. In particular we have highlighted the potential for using information on the seasonal changes in vegetation growth and senescence as indicators of the length and intensity of the malaria transmission season since these processes are closely allied to rainfall and humidity. Similar work has been undertaken in Kenya where long term mean seasonal changes in vegetation have been shown to be well correlated with the seasonal environmentally based malaria transmission maps produced in the 1950s[8].

Spectral information and derived environmental proxies (such as NDVI) derived from satellite data have been used to predict the distribution and abundance of a range of disease vectors. With rare exception these studies have not addressed the problem of spatial autocorrelation in the data sets. Statistical modeling of the relationship between insect abundance or disease prevalence and environmental data is complicated by spatial relationships, which typically result in positive correlations between observations from spatially close sampling units. These correlations in turn negate the basic assumptions underlying standard linear or generalized linear regression analysis. Failure to allow for spatial correlation typically leads to spuriously small standard errors of regression parameter estimates, and corresponding over-statement of the significance of regression effects.

The Gambian National Impregnated Bednet Programme (NIBP) which was instigated in 1991, had as its objective to treat with insecticide (permethrin) all bednets found in all villages covered by the national Primary Health Care (PHC) programme over a period of 2-3 years. An extensive evaluation of this programme was undertaken and results widely reported elsewhere. Considerable differences in malaria endemicity have been found within The Gambia (i.e. village based parasite prevalence rates varying from 1-89% in children aged 1-4year) and this has been related to ecological differences affecting the vector (species and population density) and the human population (use of bednets).

This study uses data from a cross sectional malaria morbidity survey of children (aged 1-4 year) from 65 villages (Figure 1) from 5 ecologically diverse areas of The Gambia which was carried out at the end of the 1992 rainy season as part of the NIBP evaluation[15]. Each child was given a clinical examination and their age and weight/height was recorded. Details from their health card were taken if a health card was available. Finally, a blood sample was collected by finger prick for thick and thin blood films for determination of malaria parasitaemia. Approximately two thirds of the children came from Primary Health Care villages (PHC), half of which had their bednets treated with permethrin in 1992, a few months before the survey.

Children were classified into three groups: A) children who slept without a net (591, 29%); B) children who slept under an untreated net (886, 43.5%); C) children who slept under an insecticide treated net (562, 27.5%). A total of 728 (35.7%) children where found to be positive for *Plasmodium falciparum* parasites during the morbidity survey.

Seasonal profiles of NDVI (NDVI_S: indicative of the seasonal changes in vegetation greenness throughout the Gambia were obtained from archived satellite data from the NOAA-AVHRR sensor Figures 2a&b. (See Methods). The seasonal profiles from the ten classes used indicate clearly that there is considerable difference in the length and 'greenness' of the vegetation cycle in The Gambia despite its relatively small size and uncomplicated topography.

We used a logistic regression model to estimate the probability of the presence of malarial parasites in each child as a function of NDVI_S, its quadratic

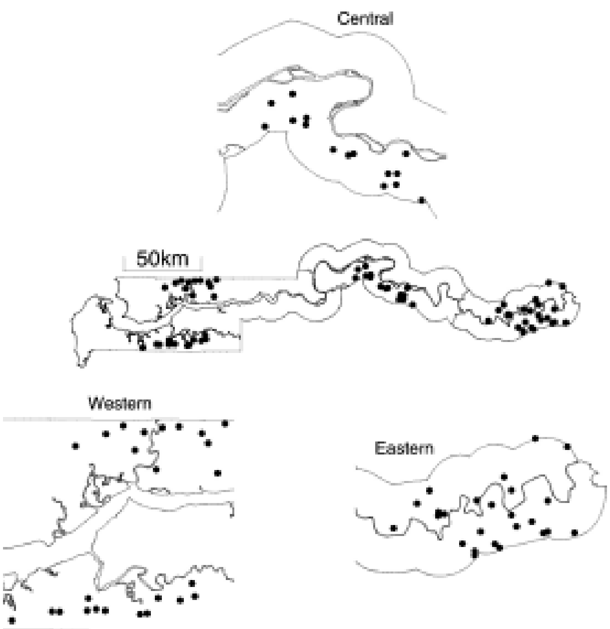

Figure 1 The distribution of the 65 study villages in The Gambia in 5 areas.

function NDVI_S^2, and AGE (in days), and adjusted for the effects of NETS (treated, untreated or absent), PHC (yes or no), (Health) CARD (yes or no) and AREA (a five-level factor) identifying the five ecologically different areas. We then used an adaptation of a method due to Liang and Zeger to make inferences that account for spatial correlation (See Methods).

The results of the non-spatial logistic regression analyses of the presence/absence of malaria in each child (n=2039 from 65 villages) are given in Table 1. AGE (in days) and NDVI_S^2 are both positively correlated with the presence of parasites (nominal p<0.001) whilst NDVI_S, the use of NETS (treated or untreated), living in a PHC village and possessing a health CARD are all negatively correlated with the presence of parasites (nominal p<0.001). As noted above quoted nominal p-values are likely to over-state the true significance of these results.

The results of the logistic regression analyses of the presence/absence of malaria in each child corrected for the spatial inter dependence of the data (spatial model) are also presented in Table 1. The second model has the same estimated regression parameters but, because the standard errors have been corrected for spatial independence, the significance of the co-variates has changed accordingly. In the spatial model the t-value for AGE remains approximately the same as in the non-spatial model indicating (as we would expect) that there is no spatial difference in age structure of 1-4 year olds in The Gambia. The t-value for NDVI_S and NDVI_S^2 changes considerably from -4.408 to –2.139 and 4.507 to 2.169 respectively but both remain significant at the 5% level. This indicates that some but not all of the association observed between presence of parasites in children and length of transmission is due to spatial autocorrelation. The t-values of the variables Untreated Nets, Treated Nets, are also reduced in the

Variable	**Regression**	**Non-spatial model 1**			**Spatial model 2**		
	Value	**Std. Error1**	**t-value1**	**Sig1**	**StdError2**	**t-value2**	**Sig2**
(Intercept)	9.063136	2.227954	4.068	**	4.51718	2.006	*
NDVI_S	-0.40171	0.091136	-4.408	**	0.18782	-2.139	*
NDVI_S^2	0.00424	0.000941	4.507	**	0.001955	2.169	*
AGE.DAY	0.000588	0.000118	5.003	**	0.000115	5.1	**
NETS(Untreated)	-0.55205	0.132226	-4.175	**	0.201298	-2.742	**
NETS(Treated)	-0.67628	0.148095	-4.567	**	0.245498	-2.755	**
PHC	-0.43224	0.121212	-3.566	**	0.243331	-1.776	NS
CARD	-0.25624	0.114605	-2.236	*	0.139126	-1.842	NS
AREA 2	-0.82356	0.186853	-4.408	**	0.382706	-2.152	*
AREA 3	-0.65562	0.19583	-3.348	**	0.415301	-1.579	NS
AREA 4	0.14781	0.247191	0.598	NS	0.556225	0.266	NS
AREA 5	0.602226	0.235724	2.555	*	0.508883	1.183	NS

* = p<0.05, ** p<0.001, NS = not significant

Table 1 Logistic regression model of presence or absence of P.falciparum parasitaemia using non-spatial and spatial statistics.

A* -nets n = 591	B* -nets n = 886	C* -nets n = 562	No of children infected with *P.falciparum*	% of children infected with *P.falciparum*	Change ratio
None	None	None	910.1	44.6	1.25
None	Untreated	Untreated	740.8	36.3	1.02
None	Untreated	Treated	728.0	35.7	1.00
None	Treated	Treated	706	34.6	0.97
Untreated	Untreated	Untreated	669.6	32.8	0.92
Treated	Treated	Treated	619.7	30.4	0.85

Italicized = actual data from children classfied into three groups (*) see text.

Table 2 Predicted number of children infected with P.falciparum *under different net usage scenarios.*

spatial model when compared to the non-spatial model but remain highly significant ($p<0.001$). Neither living in a Primary Health Care/non Primary Health Care village or possessing or otherwise a health card remain significant ($p>0.05$) once spatial effects are taken into account.

2 Predicting changes in prevalence rates

Using the model presented in Table 1 we predicted the number of children infected with *P.falciparum* under differing net scenarios (Table 2). These scenarios included: the current situation, if no children slept under a net, if all children slept under a net; if all children slept under a treated net, if current net status prevailed but none were treated.

According to the model, removing nets from those children who currently use them would result in an increase of parasite prevalence from 36% to 45%. Providing net treatment without increasing net coverage would results in a slight decrease in parasite prevalence from 36% to 35%. A more significant drop from 36% to 33% would result if net coverage were increased to include all children even when the nets were not treated. Not surprisingly the best possible result would be obtained if all children slept under a treated net (prevalence rates decrease from 36% to 30%).

The significance of all associations between malaria prevalence and human (AGE, Untreated NETS, Treated NETS, CARD) and village (PHC, AREA) factors and environmental satellite data (NDVI_S and NDVI_S^2) are reduced when taking into account spatial autocorrelation with the exception of child age. However, the associations with bednet use (treated or untreated) and with the satellite data (NDVI_S and NDVI_S^2) are still significant at the 1 and 5% level respectively, even when spatial effects are taken into account. This suggests that these associations have a more direct link and may be determined by a biological mechanism. These results confirm the widely published non-spatial analysis of the effectiveness of insecticide treated bednets in reducing infection with malaria parasites. They also reaffirm the importance of traditional bednets (untreated with insecticide) in reducing children exposure to malaria. Indeed, the study indicates that in The Gambia, increasing the use of normal bednets might be as effective in reducing the rate of infection among children as providing insecticide to treat them. However, it is thought that insecticide-treated bednets are more effective than untreated nets at preventing infections characterized by high-density parasitaemia[14] and this might partly explain their significant impact on childhood mortality. Further studies are required to establish whether a relationship exists also between the outcome of infection (disease and death) and the environmental satellite data analysed in this paper.

According to this spatial analysis, living in a PHC or non-PHC village has no significant effect on the probability of malaria infection. This contradicts earlier non-spatial analyses, which have highlighted the differences in infection rates between children living in PHC and non-PHC villages[14]. According to our analyses such results represent spatial differences in the siting of PHC and non-PHC villages rather than intrinsic qualities associated with the delivery of primary health care.

The model also suggests that environmental satellite data (NDVI) which indicates the length of the vegetation growing season may be useful for predicting the levels of malaria endemicity in children once behavioral factors (i.e. bednet usage) are taken into account. It provides empirical support for the use of satellite data to provide first level stratification of regions or countries for malaria control activities[3].

The association of the NDVI with a suitable environment for the survival of infective mosquitoes could explain our finding. NDVI has been shown to be highly correlated with saturation deficit in The Gambia[10] and the limited data available (unpublished MCT) suggests that the entomological innoculation rate of Gambian vector populations increases in association with NDVI and decreases rapidly when vegetation begins to senesce at the end of the rainy season.

In conclusion, this study demonstrates the importance of spatial effects in the inferences that can be drawn from epidemiological models created from multiple regression analysis. It confirms the effectiveness of insecticide-treated bednets in reducing parasite prevalence but suggests that, in the Gambian situation, the use of traditional untreated nets plays also a significant role in reducing malaria exposure in children. The model indicates that earlier studies suggesting that living in a primary health care village provides protection from malaria infection is the result of spatial autocorrelation. The model also suggests that there may be an *a priori* relationship between NDVI and malaria infection. We suggest that this relationship is likely to be based on the fact that high NDVI values are indicative of a moist environment supporting both mosquitos breeding and adult survival.

3 Methods

3.1 Study area

The Gambia is a small country situated on the west coast of Africa. It extends eastwards from the Atlantic Ocean on either side of the River Gambia. In brief the country consists of flat, woodland savannah with swamps bordering the river, which is saline to a distance of approximately 150Km from the coast. The climate is typical Sahelian with a short rainy season which lasts from June/July to October followed by a long dry season covering the remaining months. Minimum and maximum mean monthly temperatures range between approximately 26∞C and 33∞C with highest temperatures in the east of the country. There are therefore no temperature restrictions on malaria transmission in The Gambia but transmission is restricted largely to the rainy when temporary breeding sites and an environment suitable for adult mosquito survival is created.

3.2 Satellite data

Data collected from Channels 1 and 2 of the NOAA-AVHRR[1] sensor during 1992, 1993 and 1995 have been processed to produce a global archive of NDVI images at a temporal resolution of one dekad (10 days) and a spatial resolution of approximately 1.1km^2. These data are held by the NASA PATHFINDER 1km project and can be accessed over the internet[3,4].

Thirty six dekadal images from May 1992-April 1993 of The Gambia were extracted from the PATHFINDER active 1.1km AVHRR archive. The images, in Goodes Homolosine projection, were processed using the image analysis and display software packages IDA/WINDISP[5]. One image was incomplete and it was therefore re-created using the mean of the previous and subsequent images. The images were then entered into a spatial statistical analysis programme, ADDAPIX[6], that has been designed by the Famine Early Warning Community for monitoring seasonal and inter-annual vegetation growth using satellite imagery. ADDAPIX uses the IDA (also developed by FEWS) image format as the basic input file and submits the series of images to a Principle Component Analysis. This is followed by a non-hierarchical clustering procedure with the aim of grouping pixels (image squares) that have a similar spatial and temporal pattern. After a period of experiment 10 classes were chosen as a number large enough to indicate the ecological variation within the Gambia but small enough for a number of classes to contain sufficient villages for comparison of malaria endemicity (Figures 2a&b).

3.3 Spatial analysis

A number of children were excluded from the original data set of 2276 children because either their age was not known, their net usage was not known or no vegetation information was obtained from their village environment due to persistent cloud. After these exclusions 2039 children (89% of the original sample) from 65 villages were included in the analysis.

Modeling the relationship between environmental data and malarial indices is complicated by the spatial dependence between observations, which invalidates the inferences associated with standard regression calculations. In particular, whilst standard regression modeling applied to spatially dependent data gives satisfactory point estimates of regression parameters,

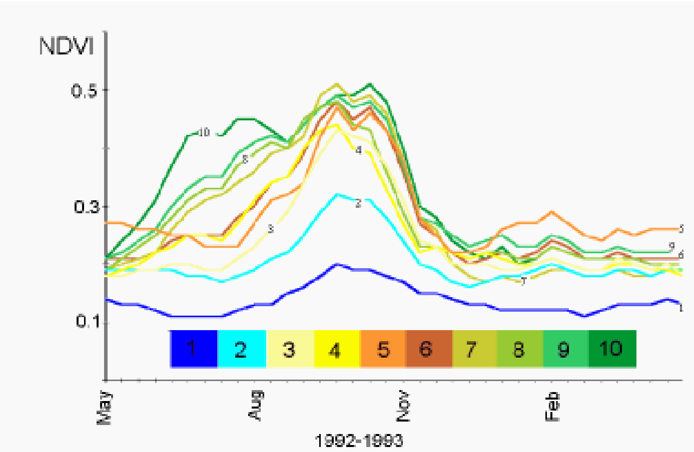

Figure 2a Seasonal profiles of NDVI The NDVI measure (called NDVI_S) used as a proxy for length of the malaria transmission season = the area under the curve for the period 1st dekad of May until 3 dekads prior to the morbidity survey. A lag of one month between NDVI values and malaria cases has been shown in an early study on Gambian data[10]. This value was extracted for each village co-ordinate. A second variable, the quadratic value of NDVI_S (NDVI_S^2) was computed and included in the analysis.

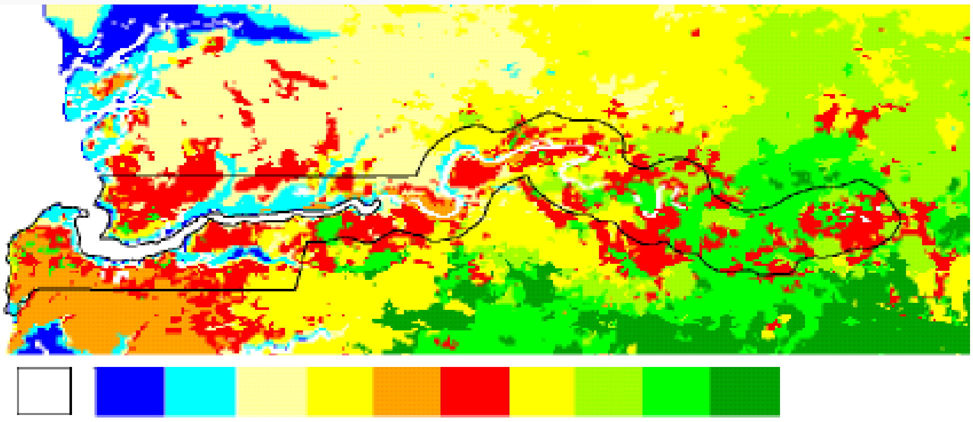

Figure 2b Spatial distribution of the 10 seasonal classes of NDVI. The colours correspond to seasonal profiles in Figure 2a.

the nominal standard errors associated with these parameter estimates are typically too small, and this in turn leads to an exaggerated impression of the significance of regression relationships. We therefore undertook the following analysis:

a) We used a logistic regression model to estimate the probability of the presence of malarial parasites in each child as a.function of the independent variables.

b) We computed the variogram of the standardised residuals from the fitted logistic regression model in order to estimate the spatial dependence in the data. Based on the appearance of the variogram, we assumed that the correlation between a pair of measurements was an exponentially declining function of distance, ie $p(d) = \exp(-ad)$ for some positive value of a.

c) From the fitted exponential correlation model, we adjusted the nominal standard errors of the logistic regression parameter estimates to allow for the effects of spatial dependence, using the method of generalised estimating equations[20]. This allows us to re-assess the nominal significance of particular terms in the regression model, and to simplify the model accordingly.

Having fitted the model, we can use it to compute the estimated probability of infection for any combination of values of the explanatory variables, and so predict the prevalence rates for children in any particular village with a given set of characteristics. We can then repeat this process under any chosen set of circum-stances to quantify the effect on prevalence rates of any change in behavior, e.g. increased use of mos-quito nets.

Acknowledgements

Eric Pfirman for adaptations to WINDISP. Financial support from the Department for International Development, UK, is gratefully acknowledged.

All correspondence should be addressed to M.C. T. (email: mthomson@liv.ac.uk).

References

[1] Advanced Very High Resolution Radiometer sensor onboard the NOAA satellite

[2] Pfirman. E, (1995). Using NDVI for Famine Early Warning. Review Document, Office of Arid Land Studies, University of Arizona, USA

[3] http://edcwww.cr.usgs.gov/landdaac/1KM/comp10d.html

[4] Eidenshink. J, Faundeen. J, (1997). The 1-km avhrr global land data set: first stages in implementation, Internet: http://edcwww.cr.usgs.gov/landdaac/1KM/paper.html

[5] IDA/WINDISP software developed by Eric Pfirman, Office of Arid Land Studies, University of Arizona, USA

[6] ADDAPIX software developed by Silvio Griguolo & Maaimo Mazzanti of the University of Venice, Italy.

[7] Thomson, M.C., Connor, S.J., Milligan, P.J.W., & Flasse, S., (1997)Mapping malaria risk in Africa - what can satellite data contribute? *Parasitology Today*.**8,** *313-318*

[8] Hay, S.I., Snow, R.W., & Rogers, D.H. (1998). From predicting mosquito habitat to malaria seasons using remotely sensede data: practice problems and perspectives. *Parasitology Today* **14**, 306-313

[9] Craig, M.H., R. W. Snow, and D. le Sueur. (in press). A climate-based distribution model of malaria transmission in Africa. *Parasitology Today*

[10] Thomson, M.C. *et al* (1996). The ecology of malaria - as seen from earth observation satellites. *Annals of Tropical Medicine and Parasitology.* **90** 243-264

[11] Hay, S.I., Packer, M.J., Rogers, D.J. (1987). The impact of remote sensing of the study and control of invertebrate intermediate hosts and vectors of disease. *International Journal of Remote Sensing.* **18** 899-930

[12] Kitron, U., Otieno, L.H., hungerford, L.L., Odulaja, A., Brigham, W.U., Okello, O.O., Josely, M., Mohamed-Ahmed, M.M. and Cook, E. (1996). Spatial analysis of the distribution of teste flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *Journal of Animal Ecology* **65** 371-380

[13] McCullagh, P. and Nelder, J.A. (1989). Generalized linear models. (second edition) London: Chapman and Hall

[14] D'Alessandro, U., Olaleye, B., McGuire, W., Thomson, M.C., Langerock, P., Bennet, S. and Greenwood, B.M. (1995). A comparison of the efficacy of insecticide-treated and untreated bednets in preventing malaria in Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **89** 596-598

[15] D'Alessandro, U., Olaleye, B., McGuire, W. Bennet, S.,Langerock, P., Aikins, M.K., Thomson, M.C., Cham, M.K., Cham, B.A. and Greenwood, B.M. (1995). Reduction in mortality and in morbidity from malaria in Gambian children following the introduction of a National Insecticide Impregnated Bednet Programme. *Lancet* **345** 479-483

[16] Thomson, M.C., Connor, S.J., D'Alessandro, U., Bennet, S.,Milligan, P. Aikins, M.K., Langerock, P., Jawara, M. and Greenwood, B.M. (1996). Geographical perspectives on bednet use and malaria transmission in The Gambia. *Social Science and Medicine* **43**(1) 101-112

[17] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13-22

[18] CChoi, H.W., Breman, J.G., Teutsch, S.M., Liu, S., Hightower, A.W., and Sexton, J.D. (1995). The effectiveness of insecticide-impregnated bed nets in reducing cases of clinical malaria: a meta-analysis of published results. *American Journal of Tropical Medicine and Hygiene.* **52** 377-382

[19] Lengeler, C., Armstrong-Schellenberg, J., D'Alessandro, U., Binka, F., and Cattani, J. (1998) Relative versus risk of dying reduction after using

insecticide-treated nets for malaria control in Africa. *Tropical Medicine and International Health* **3** 286-290.

[20] Grigulolo, S. and Mazzanti, M. (1996). Addapix. Pixel by Pixel classification for zoning and monitoring. Technical Report to FAO Rome.

[21] McCullagh, P. and Nelder, J.A. (1989). Generalized linear models. (second edition) London: Chapam and Hall.

[22] Cressie, N.A.CC. (1990). Statisticsw for spatial data. New York, Wiley.