Journal of the Korean Statistical Society (1999), 28: 4, pp 457-468

Partially Observed Data in Spatial Autologistic Models with Applications to Area Prediction in the Plane[†]

Young-Won Kim, Eun-Ha Choi and Sun Y. Hwang[1]

ABSTRACT

Autologistic lattice process is used to model binary spatial data. A conditional probability is derived for the incomplete data where the lattice consists of partially yet systematically observed sites. This result, which is interesting in its own right, is in turn applied to area prediction in the plane.

Keywords: Autologistic Model; Nearest Neighbor; Partially Observed Data; Maximum Pseudo Likelihood; Area Prediction.

1. INTRODUCTION

Let $\{X_{ij}\ ;\ i = 1, \cdots, m, j = 1, \cdots, n\}$ denote binary $(0, 1)$ random variables at the sites labelled (i, j) in a two dimensional lattice. Typically, $X_{ij} = 0$ refers to characteristic of interest-free and $X_{ij} = 1$ stands for "contains the characteristic". The problem of modeling such binary $\{X_{ij}\}$ has received considerable attention in the literature in the context of spatial statistics. The applications of the binary lattice models are commonly encountered in such diverse fields as geography, geology and epidemiology. Refer to Cliff and Ord(1975), Haining(1979), Cressie(1991) and more recently Guyon(1995) among others for the excellent applications of the binary lattice models. See also Besag(1974) for the comprehensive treatments regarding binary models including several auto-models.

We here discuss some topics for the autologistic process which is useful in modeling binary spatial data. Specifically, this paper is mainly concerned with the incomplete data where fundamental difficulty lies in the estimation of parameters involved. This article is organized as follows. Section 2 describes the model and a

[†]The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997(#1997-001-D00074).

[1]Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Korea.

test of independence against clustering. Section 3 is dealing with the interesting situation where the data are partially observed but in a systematic way, which will be specified later. An application of results derived in Section 3 to the area prediction is presented in Section 4 via a simple example.

2. THE MODEL

We consider spatial binary (0, 1) valued process defined on a $m \times n$ rectangular grid in a plane with sites labelled (i, j) where an associated random variable X_{ij} is well defined at each sites. There have been two main approaches for modeling binary spatial process: Joint probability approach and Conditional probability formulation. An excellent review of connections between these two approaches and some applications are given by Ripley(1981). We here mainly concentrate on conditional model approach, precisely, discussing autologistic models.

The big question whether the judicious joint probability of all sites exists when one is modeling through a conditional probability model is affirmatively answered by Besag(1974) appealing to the famous Hammersley-Clifford theorem.

Consider the set of four observations in the "nearest-neighbor" of the site (i, j) as

$$N_{ij} = \{X_{i+1,j},\ \ X_{i-1,j},\ \ X_{i,j+1},\ \ X_{i,j-1}\} \tag{2.1}$$

Denoting the sum of elements in N_{ij} by S_{ij}, viz.,

$$S_{ij} = X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1} \tag{2.2}$$

let

$$P(X_{ij} \mid \textit{all the rest}) = \exp[\beta_0 + \beta_1 S_{ij}]/\{1 + \exp[\beta_0 + \beta_1 S_{ij}]\} \tag{2.3}$$

where $\beta = (\beta_0, \beta_1)$ is a vector of parameters to be estimated.

Notice that the Markov property is assumed in (2.3) with the nearest neighbor N_{ij} and hence this model is a special member of the first order Markov random fields in a plane. The binary model specified by (2.1) to (2.3) is usually called a conditional autologistic model. Pickard(1987) call this Ising model in the plane. Also, Besag(1974) introduce the term "Isotropic first-order scheme" for this model. Pickard(1987) examined in detail this model including various inferential problems and related asymptotics. Also, Refer to Jain(1981) for the diverse applications to such disciplines as geography and geology. Further, as pointed out by Cressie(1991, p.424.), the form (2.3) is a consequence of algebra and is not an assumption for modeling binary spatial data.

We begin by discussing spatial autocorrelations of the model defined by (2.1) through (2.3). If the presence of some "quality" in neighboring sites makes its presence in the site more likely, the phenomenon is said to show spatial positive autocorrelation. It is easily seen from (2.3) that the parameter β_1 indicates "interaction" between X_{ij} and S_{ij} with independence if and only if $\beta_1 = 0$. As β_1 increases from zero, the conditional probability of $X_{ij} = 1$ given S_{ij} becomes more likely and hence positive β_1 measures the tendency of clusters of ones. It would then be interesting to test for independence .vs. clustering of ones.

$$\begin{aligned} H_0 &: \quad \beta_1 = 0 \quad (\textit{independence}) \\ H_1 &: \quad \beta_1 > 0 \quad (\textit{clustering of ones}) \end{aligned} \tag{2.4}$$

A score-type statistic for testing (2.4) is proposed by Hwang and Basawa(1999) in a slightly different context where a unilateral variant of the model is obtained by replacing N_{ij} with $\widetilde{N_{ij}}$ consisting of observations from "lower quadrant neighbor" such that

$$\widetilde{N_{ij}} = \{X_{i-1,j},\ X_{i-1,j-1}\ ,\ X_{i,j-1}\}$$

The resulting model is referred to as *unilateral* conditional autologistic process. Adapting the lines of their work to our model one can obtain:

Theorem 2.1. *Consider the test $\phi(T)$ based on the score statistic T, given by*

$$\phi(T) = I[T \geq z_\alpha]$$

where I stands for the usual indicator function, Z_α is used for the upper α percentile of $N(0,1)$ and

$$T = \sum(X_{ij} - \bar{X})S_{ij}/[\bar{X}(1-\bar{X})\sum(S_{ij} - \bar{S})^2]^{1/2}$$

Here $\bar{X}$ and $\bar{S}$ denote the averages and the summation is taken over all sites in the lattice. Then, as the total sample size mn tends to infinity, $\phi(T)$ is a limiting size α test for the hypothesis (2.4).

Upon retaining H_0 : independence, mn independent Bernoulli random variables with identical success probabilities comprises the lattice. In this case our model is of little use. In the remainder of this paper, it will be assumed that there tends to be clustering of ones, i.e., $\beta_1 > 0$.

3. PARTIALLY OBSERVED DATA

It is often the case in practice that observations of all sites are not available in hand. Examples of such phenomena are typically two-fold: (1) Data containing missing values due to noise or degrading. The missing data here usually occur at random and there is no particular pattern to where there are missing values. One can accomodate missing data by using smoothing and/or interpolation (ref. Cressie(1991), Ch. 6). (2) Data consisting of intermittent observations for the sake of reducing sampling cost, which is the case we are focusing on. For simplicity of discussion, consider the following special pattern of the partially yet systematically observed data. Denoting by $*$ [$\cdot$] the observed [unobserved] sites, respectively, examine the $m \times n$ lattice alternatively "$*$" and "$\cdot$" , as in Figure 3.1.

$$
\begin{array}{cccccccc}
* & \cdot & * & \cdot & * & \cdot & * & \cdot \\
\cdot & * & \cdot & * & \cdot & * & \cdot & * \\
* & \cdot & * & \cdot & * & \cdot & * & \cdot \\
\cdot & * & \cdot & * & \cdot & * & \cdot & *
\end{array}
$$

Figure 3.1 : Partial Observation Scheme
($*$: observed site : m=4, n=8)

For the fully observed data from the (spatial) autologistic process, various methods for estimating $\beta = (\beta_0, \beta_1)$ have been devised: [1] maximum likelihood (ML) estimation (Gidas(1991)); [2] maximum pseudo likelihood (MPL) estimation (Geman and Graffigne(1987)); [3] Coding method (Besag(1974)). It is worth indicating that MPL objective function is much easier to handle than ML function where the likelihood function often contains an intractable and unwieldy normalizing constant (see, for instance, Cressie(1991), p. 459-460). The coding method of Besag can not be directly applicable to the data with incomplete observations.

To employ ML or MPL methods for the incomplete scheme depicted in Figure 3.1, we first need to compute the conditional probability of X_{ij} given the rest sites with observations. Fix the observed site (i, j) and name the (second) nearest eight sites shown in Figure 3.2, where " $\cdot$ " stands for unobserved sites and $(1) = (i+2,\ j)$, $(2) = (i+1,\ j-1)$, $(3) = (i,\ j-2)$, $(4) = (i-1,\ j-1)$, $(5) = (i-2,\ j)$, $(6) = (i-1,\ j+1)$, $(7) = (i,\ j+2)$ and $(8) = (i+1,\ j+1)$. Denote by N_{ij}^{*} the collection of 8 random variables observed at 8-sites.

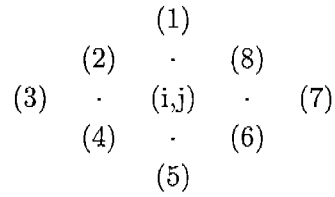

Figure 3.2 : Name of the second nearest sites

It follows from the first order Markov property of the spatial autologistic processes in (2.1) to (2.3) that

$$\begin{aligned} & P(X_{ij} = 1 \mid \textit{rest sites with observations}\ ) \\ = \ & P(X_{ij} = 1 \mid N^*_{ij}) \\ = \ & \Delta_{ij}, \textit{say} \end{aligned} \tag{3.1}$$

Since S_{ij} in (2.3) takes values 0 to 4, write that for each k=0, 1, 2, 3, 4

$$w_k = e^{(\beta_0+\beta_1 k)}/[1 + e^{(\beta_0+\beta_1 k)}] \tag{3.2}$$

Also, let R^*_{ij} be the number of the three sites (6), (7) and (8) in the right hand side in Figure 3.2 that have a value 1, viz.,

$$R^*_{ij} = X_{i-1,j+1} + X_{i,j+2} + X_{i+1,j+1} \tag{3.3}$$

Similarly, coin the term L^*_{ij} for (2),(3) and (4) ; U^*_{ij} for (1),(2) and (8) and reserve the notation D^*_{ij} for the remaining three sites in the down side of Figure 3.2, (4),(5) and(6).

$$D^*_{ij} = X_{i-1,j-1} + X_{i-2,j} + X_{i-1,j+1} \tag{3.4}$$

Introduce the notation a_{1111} and b_{1111} to denote

$$\begin{aligned} a_{1111} \ = \ & \frac{e^{(\beta_0+\beta_1(1+R^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+R^*_{ij}))}} \cdot \frac{e^{(\beta_0+\beta_1(1+L^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+L^*_{ij}))}} \\ & \cdot \frac{e^{(\beta_0+\beta_1(1+U^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+U^*_{ij}))}} \cdot \frac{e^{(\beta_0+\beta_1(1+D^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+D^*_{ij}))}} \end{aligned} \tag{3.5}$$

and

$$b_{1111} = \frac{e^{(\beta_0+\beta_1 R^*_{ij})}}{1+e^{(\beta_0+\beta_1 R^*_{ij})}} \cdot \frac{e^{(\beta_0+\beta_1 L^*_{ij})}}{1+e^{(\beta_0+\beta_1 L^*_{ij})}} \cdot \frac{e^{(\beta_0+\beta_1 U^*_{ij})}}{1+e^{(\beta_0+\beta_1 U^*_{ij})}} \cdot \frac{e^{(\beta_0+\beta_1 D^*_{ij})}}{1+e^{(\beta_0+\beta_1 D^*_{ij})}} \tag{3.6}$$

Note that a_{1111} and b_{1111} differ only in that for a_{1111} one is added in the multiplication factors appearing in front of β_1.

Now the four subscripts "1" in a_{1111} are replaced by "0" with the understanding that "0" is substituting (one minus the corresponding factor) into (3.5). For example, the following notation a_{1101} gives, by adjusting third term in (3.5)

$$a_{1101} = \frac{e^{(\beta_0+\beta_1(1+R^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+R^*_{ij}))}} \cdot \frac{e^{(\beta_0+\beta_1(1+L^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+L^*_{ij}))}} \cdot \left(1-\frac{e^{(\beta_0+\beta_1(1+U^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+U^*_{ij}))}}\right) \cdot \frac{e^{(\beta_0+\beta_1(1+D^*_{ij}))}}{1+e^{(\beta_0+\beta_1(1+D^*_{ij}))}} \tag{3.7}$$

The subscripts in b_{1111} are also handled in analogous way. We are now in a position to compute Δ_{ij} in (3.1).

Theorem 3.1. *For the partially observed scheme outlined in Figures 3.1 and 3.2, the conditional probability of X_{ij} for the observed site (i, j), given the rest sites with observations, call it Δ_{ij}, is given by*

$$\Delta_{ij} = \frac{D_{ij}}{1-C_{ij}+D_{ij}} \tag{3.8}$$

where

$$\begin{aligned} C_{ij} &= \omega_4 a_{1111} + \omega_3(a_{1110}+a_{1101}+a_{1011}+a_{0111}) \\ &\quad + \omega_2(a_{1100}+a_{1010}+a_{1001}+a_{0110}+a_{0101}+a_{0011}) \\ &\quad + \omega_1(a_{1000}+a_{0100}+a_{0010}+a_{0001}) + \omega_0 a_{0000} \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} D_{ij} &= \omega_4 b_{1111} + \omega_3(b_{1110}+b_{1101}+b_{1011}+b_{0111}) \\ &\quad + \omega_2(b_{1100}+b_{1010}+b_{1001}+b_{0110}+b_{0101}+b_{0011}) \\ &\quad + \omega_1(b_{1000}+b_{0100}+b_{0010}+b_{0001}) + \omega_0 b_{0000} \end{aligned} \tag{3.10}$$

Proof: Recall first that the four nearest sites comprise N_{ij} in (2.1) and N_{ij}^* in Figure 3.2 consists of 8 sites for the observed site (i, j). Write that

$$\begin{aligned} \Delta_{ij} &= P(X_{ij}=1|N_{ij}^*) \qquad (3.11) \\ &= \sum P(X_{ij}=1|N_{ij}^*, N_{ij}=(i_1,i_2,i_3,i_4)) \cdot P(N_{ij}=(i_1,i_2,i_3,i_4)|N_{ij}^*) \end{aligned}$$

where i_1, i_2, i_3 and i_4 each takes values 0 or 1 ,and the summation runs over all i_1, i_2, i_3 and i_4 , consisting of sixteen terms. It follows from Markov property that

$$P(X_{ij}=1|N_{ij}^*, N_{ij}) = P(X_{ij}=1|N_{ij}) \qquad (3.12)$$

It can also be written as

$$\begin{aligned} P(N_{ij}|N_{ij}^*) &= \sum_{u=0}^{1} P(N_{ij}|N_{ij}^*, X_{ij}=u) \cdot P(X_{ij}=u|N_{ij}^*) \qquad (3.13) \\ &= P(N_{ij}|N_{ij}^*, X_{ij}=1) \cdot \Delta_{ij} \\ &\quad + P(N_{ij}|N_{ij}^*, X_{ij}=0) \cdot (1-\Delta_{ij}) \end{aligned}$$

Combining (3.11), (3.12) and (3.13), Δ_{ij} can be expressed as a sum of two terms,

$$\Delta_{ij} = A \cdot \Delta_{ij} + B \cdot (1-\Delta_{ij}) \qquad (3.14)$$

with

$$\begin{aligned} A &= \sum P(X_{ij}=1|N_{ij}=(i_1,i_2,i_3,i_4)) \\ &\quad \cdot P(N_{ij}=(i_1,i_2,i_3,i_4)|N_{ij}^*, X_{ij}=1) \end{aligned}$$

and

$$\begin{aligned} B &= \sum P(X_{ij}=1|N_{ij}=(i_1,i_2,i_3,i_4)) \\ &\quad \cdot P(N_{ij}=(i_1,i_2,i_3,i_4)|N_{ij}^*, X_{ij}=0) \end{aligned}$$

where the sum is taken over 16 terms.

It then suffices to show that $A = C_{ij}$ and $B = D_{ij}$. To proceed further, it is worth noting that due to the first order Markov property of the model

conditionally on N_{ij}^* and X_{ij}, the four random variables in N_{ij} are mutually independent. This in turn implies that

$$a_{1111} = P(N_{ij} = (1,1,1,1)|N_{ij}^*, X_{ij} = 1)$$

and

$$b_{1111} = P(N_{ij} = (1,1,1,1)|N_{ij}^*, X_{ij} = 0)$$

Fully exploiting the notation obtained by substituting zeros for the subscripts ones in a_{1111} and b_{1111}, and using ω_k in (3.2) for the weights $P(X_{ij} = 1|N_{ij} = (i_1, i_2, i_3, i_4))$ appearing in A and B, it can be deduced that $A = C_{ij}$and $B = D_{ij}$ and hence

$$\Delta_{ij} = C_{ij}\Delta_{ij} + D_{ij}(1 - \Delta_{ij})$$

which essentially concludes the proof. □

4. AREA PREDICTION IN THE PLANE

Consider the $m \times n$ lattice with $\{X_{ij}\}$ each taking values 0 or 1 and define the area for the lattice as the number of sites that have a value 1, viz., $\sum X_{ij}$ where the sum extends over all sites. Denoting O the collection of all sites with observations, the area denoted by U can be written as

$$U = \sum_{O} X_{ij} + \sum_{O^c} X_{ij}$$

Note that X_{ij} is not observable on the site belonging to O^c. It is well known that the MMSE(minimum mean squared error) predictor of a random variable is given by the conditional expectation. Consequently the MMSE predictor of U, denoted by $\widehat{U}$, is the conditional expectation of U given the data in O,

$$\widehat{U} = \sum_{O} X_{ij} + \sum_{O^c} \widehat{X_{ij}} \tag{4.1}$$

where for the site in O^c,

$$\begin{aligned} \widehat{X_{ij}} &= E(X_{ij}|data\ in\ O) \\ &= E(X_{ij}|N_{ij}) : Markov\ property \\ &= P(X_{ij} = 1|N_{ij}) \end{aligned} \tag{4.2}$$

For $\widehat{X_{ij}}$, it is the first step to estimate the model parameter β. To do so, one may proceed to employ ML or MPL methods for β. While ML in asymptotic sense usually provides better estimates than MPL, efficiency of ML over MPL can be traded for tractability and computational ease (see, Cressie(1991)). For MPL the pseudo likelihood function, call it $PL(\beta)$, can be written down in a compact way

$$PL(\beta) = \Pi \Delta_{ij}(\beta)^{x_{ij}} \ (1 - \Delta_{ij}(\beta))^{(1-x_{ij})}$$

where the product is over all sites in O. Let $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)$ be obtained by maximizing objective function $PL(\beta)$ with respect to β. It then follows that for $(i,j) \in O^c$

$$\widehat{X_{ij}} = exp[\widehat{\beta}_0 + \widehat{\beta}_1 S_{ij}]/(1 + exp[\widehat{\beta}_0 + \widehat{\beta}_1 S_{ij}]) \tag{4.3}$$

To see how $\widehat{U}$ works, we perform a simulation study by generating 50×30 lattices. Figure 4.1 [4.2] shows a lattice from $\beta = (-1.1, 0.1)[(-2.5, 1.2)]$. MPL method is then applied and in turn yields $\widehat{\beta} = (-1.4, 0.3)[(-2.3, 1.1)]$, respectively. Plugging these estimates into (4.3) $\widehat{U}$ is easily obtained. In the context of image analysis, Switzer(1980) proposes to augment the missing values with the average of the data from the nearest neighbor sites. Adapting the idea to our case, one can devise another predictor $\widetilde{U}$ given by

$$\widetilde{U} = \sum_{O} X_{ij} + \sum_{O^c} \widetilde{X_{ij}} \tag{4.4}$$

where $\widetilde{X_{ij}} = S_{ij}/4$

Both $\widehat{U}$ and $\widetilde{U}$ are readily computable to find that $\widehat{U} = 338.05$ and $\widetilde{U} = 333.45$ for Figure 4.1(true area=361). Also, in the case of Figure 4.2 (true area=418), it is obtained that $\widehat{U} = 417.09$ which is very close to the true value 418, and $\widetilde{U} = 409.95$. In each case, the proposed $\widehat{U}$ appears to get closer to the true value. Table 4.1 gives some computations for $\widehat{U}$ and $\widetilde{U}$ based on 100 independent simulations. For each simulation, 50 by 30 lattice is generated and area prediction errors ($\widehat{U}$ - true area and $\widetilde{U}$ - true area) are calculated. The mean and the standard deviation(s.d.) of these 100 errors are listed in Table 4.1. The first value shown within each cell is the mean over the 100 simulations and s.d. is entered in the square bracket.

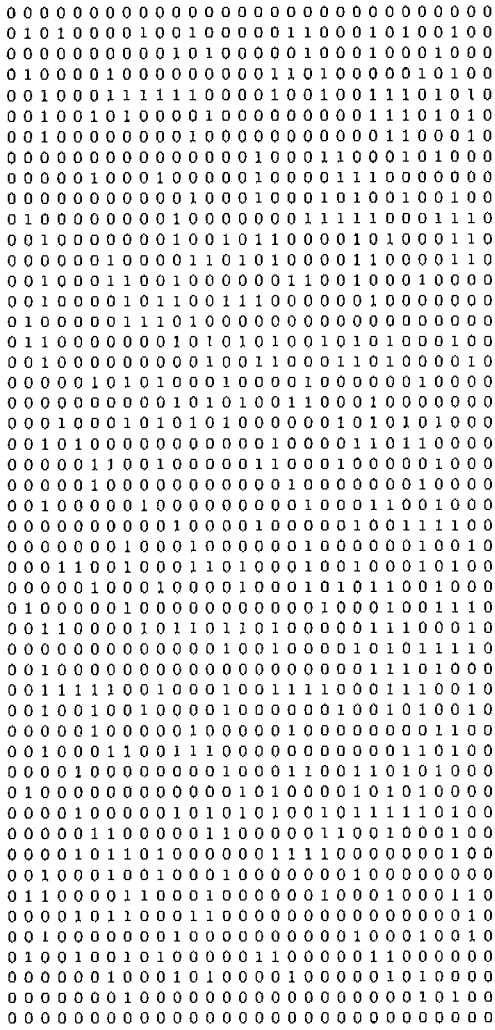

Figure 4.1

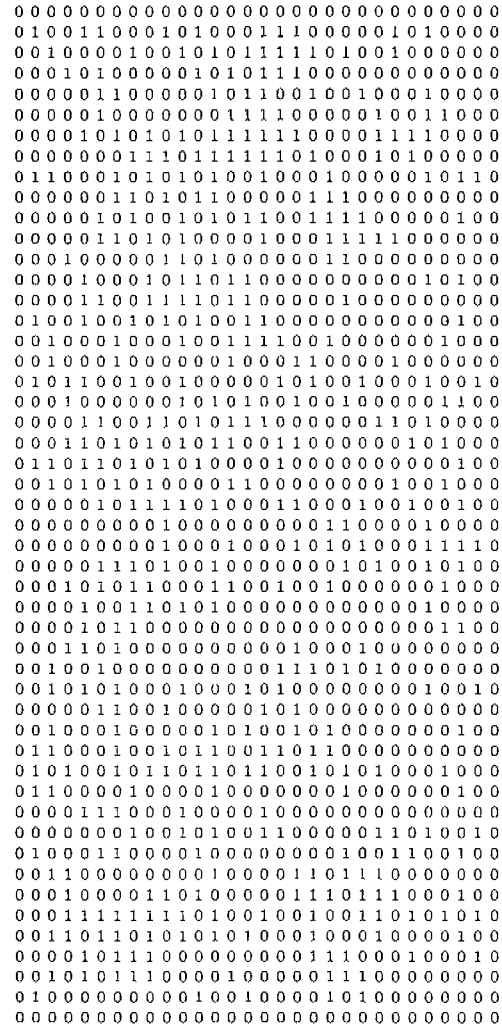

Figure 4.2

Table 4.1: Simulated Mean and Standard Deviation of the area prediction errors

(β_0, β_1)	$For \widehat{U}$	$For \widetilde{U}$
(-1.1, 0.1)	-0.18[12.17]	-0.35[18.36]
(-2.5, 1.2)	0.11[8.33]	0.15[15.19]

As can be seen from the table, $\widehat{U}$ dominates $\widetilde{U}$ in the sense that $\widehat{U}$ is less biased and provides a smaller standard deviation. For the proposed $\widehat{U}$, it seems

that the s.d. decreases as β_1 increases from zero. This makes sense since positive β_1 is associated with the spatial positive autocorrelation, and thus once the four nearest neighbor sites are observed the "uncertainty" of the missing value is clearly getting smaller as β_1 increases. As an alternative to MPL, ML can be used to estimate β in conjunction with a simulation approximation for the likelihood using MCMC method(cf. Geyer(1992)). We will not however go into details on ML method since our goal in this section is to provide a new method for area prediction via spatial autologistic models, rather than to compare various estimation procedures for β.

ACKNOWLEDGEMENT

We thank the referee for the constructive suggestions which lead to a substantial improvement of thc section 4.

REFERENCES

Besag, J. E. (1974). "Spatial interaction and the statistical analysis of lattice systems," *Journal of Royal Statistical Society*, B36, 192-225.

Besag, J. E. (1975). "Statistical analysis of non-lattice data," *The statistician*, 24, 179-195.

Cliff, A. D. and Ord, J. K. (1975). "Model building and analysis of spatial pattern in human geography," *Journal of Royal Statistical Society*, B37, 297-328.

Cressie, N. A. C. (1991). *Statistics for Spatial Data*, Wiley, New York.

Geyer, C. J. (1992), "Practical Markov chain Monte Carlo (with discussion)," *Statistical Science*, 7, 473-511.

Geman, S. and Graffigne, C. (1987). "Markov random field image models and their applications to computer vision," *Proceedings of the International Congress of Mathematicians*(1986, Berkeley, CA), 1496-1517.

Gidas, B. (1991). "Parameter estimation for Gibbs distributions: Fully observed data," in *Markov Random Fields: Theory and Applications*, Academic Press, New York.

Guyon, X. (1995). *Random Fields on a Network: Modelling, Statistics and Applications*, Springer, New York.

Hainig, R. P. (1979). "Statistical tests and process generators for random field models," *Geographical Analysis*, 11, 45-64.

Hwang, S. Y. and Basawa, I. V. (1999). "Inference for a binary lattice Markov process," *Statistics and Probability Letters*, 43, 75-85.

Jain, A. K. (1981). "Advances in mathematical models in image processing," *Proc. IEEE*, 69, 502-528.

Pickard, D. K. (1987). "Inference for discrete Markov fields: The simplest nontrivial case," *Journal of American Statistical Associations*, 82, 90-96.

Ripley, B. D. (1981). *Spatial Statistics*, Wiley, New York.

Switzer, P. (1980). "Extension of discriminant analysis for statistical classification of remotely sensed satellite image," *Journal of the International Association for Mathematical Geology*, 12, 367-376.