Biostatistics (2003), **4**, 1, *pp.* 11–25
Printed in Great Britain

Proper multivariate conditional autoregressive models for spatial data analysis

ALAN E. GELFAND[†]
Department of Statistics, University of Connecticut, Storrs, USA
alan@stat.uconn.edu

PENELOPE VOUNATSOU
Swiss Tropical Institute, Socinstrasse 57, Basle, Switzerland

SUMMARY

In the past decade conditional autoregressive modelling specifications have found considerable application for the analysis of spatial data. Nearly all of this work is done in the univariate case and employs an improper specification. Our contribution here is to move to multivariate conditional autoregressive models and to provide rich, flexible classes which yield proper distributions. Our approach is to introduce spatial autoregression parameters. We first clarify what classes can be developed from the family of Mardia (1988) and contrast with recent work of Kim *et al*. (2000). We then present a novel parametric linear transformation which provides an extension with attractive interpretation. We propose to employ these models as specifications for second-stage spatial effects in hierarchical models. Two applications are discussed; one for the two-dimensional case modelling spatial patterns of child growth, the other for a four-dimensional situation modelling spatial variation in HLA-B allele frequencies. In each case, full Bayesian inference is carried out using Markov chain Monte Carlo simulation.

Keywords: Gene frequencies; Hierarchical model; Markov chain Monte Carlo simulation; Model choice; Nutritional indicators; Spatial regression parameter.

1. INTRODUCTION

Conditional autoregression (CAR) modelling specifications date at least to Besag (1974). However, within the past decade they have found wide application for the analysis of spatial data. In particular, they are naturally employed with areal unit data either through single-stage or hierarchical models. In the former they model the spatial association of the observations directly. For the latter, they are introduced through random effects in the mean structure of the data. In either case they are well suited to model fitting using Gibbs sampling, the former illustrated in Geman and Geman (1984), the latter in numerous papers, e.g. Besag *et al*. (1991) and Clayton and Bernardinelli (1993). Their ease of implementation, particularly with large numbers of areal units and the wide availability of inexpensive, fast computing, has contributed to their dramatically increased usage.

In most of this work the conditional distributions of the CAR variables (which are used in conjunction with the likelihood to implement the Gibbs sampling updating) imply a pairwise difference joint specification, and hence an improper joint posterior distribution. Computationally, this is usually handled

[†]To whom correspondence should be addressed.

by imposing a linear constraint on the variables at each iteration of the Gibbs sampler (as argued in Besag *et al*., 1995). A more serious concern is that, when the precision parameter of the CAR model is unknown, the functional form of the joint distribution of the CAR variables and this parameter is not determined. That is, the normalizing constant of the conditional distribution of the variables given the precision parameter is arbitrary, as a function of this parameter. A remedy is to introduce a so-called spatial autoregression parameter (Cressie, 1993; Sun *et al*., 1998). Suitably constrained, this parameter ensures a proper joint distribution for the resulting CAR model.

All of the foregoing work assumes the CAR variables are univariate. Here, we are concerned with the case of multivariate CAR (MCAR) specifications. Mardia (1988) presents a full theoretical development when the specification consists of the set of multivariate conditional distributions which are the full conditional distributions for each areal unit vector given the vectors at all of the remaining units. He also cites the earlier work. However, these models have received little attention, at least initially, due to computational difficulties. Recent work of Knorr-Held and Rue (2002) employs improper versions of these MCARs in the context of joint disease mapping. Vounatsou and Mueller (2001) uses these improper versions to model child growth. Our contribution here is to examine these models in the context of introducing spatial autoregression parameters, in order to ensure distributional propriety. In particular, a novel parametric linear transformation provides extension with attractive interpretation. We then show how the models may be implemented in hierarchical modelling for a two-dimensional and a four-dimensional application. We note previous work of Kim *et al*. (2000) who, in the two-dimensional case, introduce an alternative specification which they refer to as a two fold CAR model. This model specifies the conditional distribution of each component of the vector at each areal unit given the remaining components for that unit as well as the components for all other units. We clarify when the two specifications will agree. Lastly, recent work of Langford *et al*. (1999) and Leyland *et al*. (2000) create spatial random effects as proximity-based weighted averages of independent normal variates.

The bivariate application is concerned with modelling spatial patterns for bivariate indicators of child growth in the presence of dietary explanatory variables. The two indicators measure height adjusted for age and weight adjusted for age. Association between the indicators for a given individual is certainly expected. But, also spatial association across areal units would be expected resulting from, for example, genetic and environmental factors which are not measured in the dataset.

The four-dimensional application arises in the context of allele frequencies. In particular, focusing on a single gene site or locus, one finds several variant genes called allele types or alleles. The distribution of allele types, so-called allele frequencies, is of interest. The data consist of observed allele frequencies for various spatial locations. Variation in these observed allele frequencies characterizes genetic diversity. Spatial pattern in such genetic variation is anticipated again due to environmental factors and linguistic or epidemiological patterns.

The format of the paper is as follows. In Section 2 we develop the above pair of motivating applications. In Section 3 we present a formal development of the MCAR theory we require for the modelling. In Section 4 we discuss general modelling issues. Sections 5 and 6 present the results of the data analysis.

2. MOTIVATING EXAMPLES

We introduce two examples which motivate the need for MCAR modelling specification.

2.1 *Spatial modelling of child growth*

Child growth is usually monitored using anthropometric indicators such as height adjusted for age (HAZ), weight adjusted for height (WHZ) and weight adjusted for age (WAZ). Independent analysis

of each of these indicators is normally carried out to identify factors influencing growth which may range from genetic and environmental factors (e.g altitude, seasonality) to differences in nutrition and social deprivation (Ulijaseszk *et al*., 1998). Substantial variation in growth is common within as well as between populations. Recently, geographical variation in child growth has been thoroughly investigated for the country of Papua New Guinea (PNG) in Mueller *et al*. (2001). Independent spatial analyses for each of the anthropometric growth indicators identified complex geographical patterns of child growth finding areas where children are taller but skinnier than average, others where they are heavier but shorter, and areas where they are both short and light. These geographical patterns could be linked to differences in diet and subsistence agriculture, leading to the analysis presented here. An alternative analysis built upon socio-economic differences has been developed in Vounatsou and Mueller (2001).

The data for our illustration comes from the 1982–83 PNG National Nutrition Survey (NNS) (Heywood *et al*., 1988). The survey includes anthropometric measures (age, height, weight) of approximately 28 000 children under 5 years of age, as well as dietary, socio-economic factors and demographic data about those children and their families. Dietary data include the type of food that respondents had eaten the previous day. Subsequently, the data were coded to 14 important staples and sources of protein. Each child was assigned to a village and each village was assigned to one of 4566 environmental zones (Resource Mapping Units, RMU) into which PNG has been divided for agriculture planning purposes. A detailed description of the data is given by Smith *et al*. (1993).

The HAZ and WAZ scores which describe the nutritional status of a child were obtained using the LMS method developed by Cole and Green (1992) which yields age-adjusted standard normal deviate Z-scores. The dataset was collected at 537 RMUs. To overcome sparseness and to facilitate computation, we collapsed to 250 spatial units. In the absence of digitized boundaries, Delaunay tessellations were used to create the neighbouring structure in the spatial units. Figure 1 shows a map of the units in the study.

Because of the complex, multidimensional nature of human growth, a bivariate model which considers differences in height and weight jointly might be more appropriate for analysing child growth data in general and to identify geographical patterns of growth in particular. We propose the use of Bayesian hierarchical spatial models and introduce our MCAR specifications to analyse the bivariate pairs of indicators, HAZ and WAZ, of child growth. In Section 5 we detail the modelling. This reveals bivariate spatial random effects at RMU level, calling for an MCAR specification.

2.2 *Geographical mapping of HLA-B gene frequencies*

The study of spatial patterns of gene frequencies provides possible explanations of causes of genetic diversity such as population movements, natural selection or stochastic effects. Genes appear at specified sites, or *loci* along a chromosome. Each locus can be occupied by one of several variant genes called *alleles*. Since chromosomes are paired, each individual will have two alleles at a particular locus.

The Human Leukocyte Antigen (HLA) system is the most powerful tool for the study of human population biology due to the large number of alleles at each locus and the occurrence of population-specific variants. From this system we analyse the class I HLA-B gene located at locus B. Genetic variation in this gene as well as all other genes of the class I group affects regulation of immune responses to many infectious diseases. There have been several studies of genetic diversity in the HLA system (Ting and Morris, 1973; Serjeantson, *et al*., 1992) but limited work has been done in modelling geographical patterns of allele frequencies. Recently, Vounatsou *et al*. (2000) have implemented Bayesian hierarchical models with independent improper univariate CAR priors to investigate geographical mapping of allele and haplotype frequencies for the HLA system in PNG.

The overall data set consists of serological class I HLA typing data on 5645 Papua New Guineans drawn from the general population. The allelic typing was done serologically (Bhatia *et al*., 1989) and, at locus HLA-B, identified the following six alleles: *B13, B27, B39, Bw56, Bw60, Bw62*, recording for

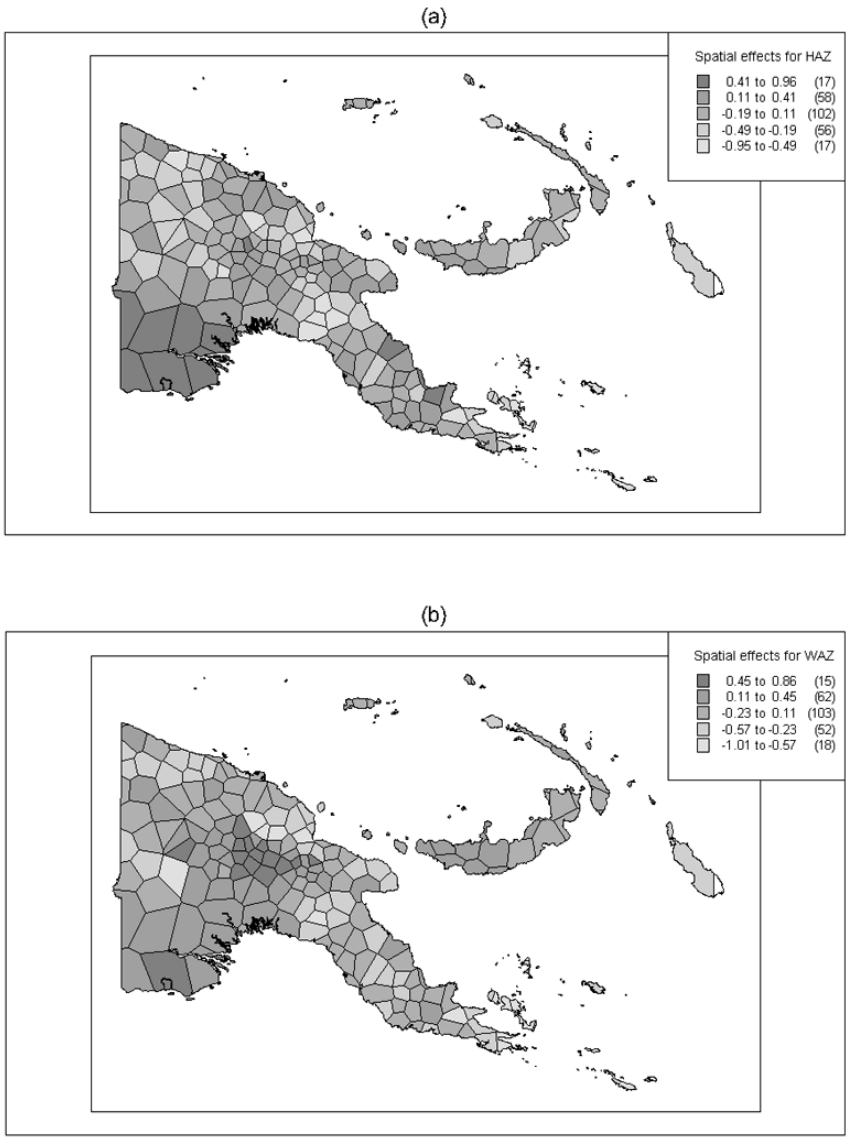

Fig. 1. Greyscale maps of posterior medians of spatial effects, under model III, for nutrition indicators: (a) HAZ, (b) WAZ.

each individual, the pair of alleles found. The dataset was collected at 252 RMUs. We consider the RMUs as the spatial units in our analyses. The sampling intensity in each unit ranged from 1 to 643. In order to increase the power of the analysis we aggregated sparse units with less than five individuals to their closest neighbours. Again, Delaunay tessellations were employed to create a neighbour structure for the resulting 98 spatial units. The units are displayed in Figure 2. Due to the sparseness in the occurrence of allele $B39$ (approximately 2%), we work only with the remaining five. In Section 6 we propose a five-cell

multinomial model for the allelic counts leading to four multiple logit models. A four-dimensional MCAR for the spatial random effects at the RMU level is employed.

3. Some MCAR theory

For the vector of univariate variables $\phi = (\phi_1, \phi_2, \ldots, \phi_n)$, the zero-centred CAR specification, following Besag (1974), sets

$$\phi_i \mid \phi_j,\ j \neq i, \tau_i^2 \sim N\left(\sum b_{ij}\phi_j, \tau_i^2\right),\ i = 1, \ldots, n \tag{1}$$

Following Brook's (1964) lemma the resulting joint density for ϕ takes the form

$$f(\phi \mid \tau) \propto \exp\left\{-\tfrac{1}{2}\phi^T D_{\tau^2}^{-1}(I - B)\phi\right\} \tag{2}$$

where B is $n \times n$ with $(B)_{ij} = b_{ij}(b_{ii} = 0)$ and D_{τ^2} is an $n \times n$ diagonal matrix with non-zero entries τ_i^2. The requirement that $D_{\tau^2}^{-1}(I - B)$ be symmetric yields the conditions

$$b_{ij}\tau_j^2 = b_{ji}\tau_i^2,\ \forall i, j. \tag{3}$$

(2) is a proper distribution only if $D_{\tau^2}^{-1}(I - B)$ is nonsingular.

In spatial applications a symmetric proximity matrix W is usually created and one sets $b_{ij} = w_{ij}/w_{i+}$ where $w_{i+} = \sum_j w_{ij}$. Then, if $\tau_i^2 = \sigma^2/w_{i+}$, (3) holds. In fact, if we define D_W to be diagonal with entries w_{i+}, we have $D_{\tau^2}^{-1}(I - B) = \frac{1}{\sigma^2}(D_W - W)$. We denote this specification by $CAR(1, \sigma^2)$.

Since $(D_W - W)\mathbf{1} = 0$, in this case (2) is an improper distribution. In particular, there is no meaningful normalizing constant as a function of σ^2. Hence, when σ^2 is unknown, even a proper prior for it does not determine the functional form of the joint distribution for ϕ and σ^2. In practice, this problem is ignored. The ϕ_i are sampled using the full conditional distributions in (1) with a linear constraint imposed. A proper inverse gamma prior is assigned to σ^2 and $f(\phi \mid \sigma^2)$ is assigned a *normalizing* constant proportional to $(\sigma^2)^{(n-1)/2}$ since $n - 1$ is the number of linearly independent ϕ. Then the full conditional distribution for σ^2 becomes an updated inverse gamma.

A suggested repair for the problem (Cressie, 1993; Sun, *et al.*, 1998) is to introduce a parameter ρ into the mean specification in (1), i.e. $E(\phi_i \mid \phi_j,\ , j \neq i, \tau_i^2) = \rho \sum b_{ij}\phi_j,\ i = 1, \ldots, n$. This results in the covariance matrix $D_{\tau^2}^{-1}(I - \rho B)$ which is nonsingular if $\rho \in (\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$ where $\lambda_{\min}$ and $\lambda_{\max}$ are smallest and largest eigenvalues respectively of B. When $B = D_W^{-1}W$ we obtain $\frac{1}{\sigma^2}(D_W - \rho W)$ and so the smallest and largest eigenvalues of $D_W^{-\frac{1}{2}} W D_W^{-\frac{1}{2}}$ are required. It is not difficult to show that in this case $\lambda_{\max} = 1$ and $\lambda_{\min} < 0$. We denote this distribution by $CAR(\rho, \sigma^2)$. ρ can be interpreted as a coefficient which measures spatial association in the sense that $cov(\phi_i, \phi_j \mid \rho, \sigma^2, \phi_l, l \neq i, j) = \frac{\sigma^2 \rho w_{ij}}{\omega_{i+}\omega_{j+} + \rho^2 \omega_{ij}^2}$. Moreover, $\rho = 0$ implies that the ϕ are independent but with individual variances which depend upon the number of neighbours. Hence, independence arises as a limiting case of $CAR(\rho, \sigma^2)$ but not of $CAR(1, \sigma^2)$.

The introduction of $\rho \neq 1$ can be criticized since now $E(\phi_i \mid \phi_j, j \neq i) = \rho \sum \omega_{ij}\phi_j/\omega_{i+}$, i.e. we no longer obtain a mean which is a weighted average of the other ϕ. Also, $\rho = 1$ is analogous to the nonstationary or random walk case in familiar autoregressive time series models and can be advantageous in accommodating more irregular spatial behaviour.

Turning to the MCAR we now let $\phi^T = (\phi_1, \phi_2, \ldots, \phi_n)$ where each ϕ_i is a $p \times 1$. Following Mardia (1988), the zero-centred MCAR sets

$$\phi_i \mid \phi_j,\ , j \neq i, \Sigma_i,\ i = 1, \ldots, n \sim N\left(\sum B_{ij}\phi_j, \Sigma_i\right) \tag{4}$$

Fig. 2. Greyscale maps of posterior medians of spatial effects, under model III, for HLA-B allelic types: (a) $Bw62 : B13$, (b) $B27 : B13$, (c) $Bw56 : B13$, (d) $Bw60 : B13$.

where each B_{ij} is $p \times p$ as is each Σ_i. As in the univariate case, Brook's lemma yields a joint density for ϕ of the form

$$f(\phi \mid \{\Sigma_i\}) \propto \exp\left\{-\tfrac{1}{2}\phi^T \Gamma^{-1}(I - \tilde{B})\phi\right\} \tag{5}$$

where Γ is block diagonal with blocks Σ_i and $\tilde{B}$ is a $np \times np$ with (i, j)th block B_{ij}. Again, symmetry of $\Gamma^{-1}(I - \tilde{B})$ is required.

A convenient special case sets $B_{ij} = b_{ij} I_{p\times p}$ yielding the symmetry condition $b_{ij}\Sigma_j = b_{ji}\Sigma_i$ analoguous to (3). If as above $b_{ij} = w_{ij}/w_{i+}$ and $\Sigma_i = w_{i+}^{-1}\Sigma$, the symmetry condition is satisfied. Kronecker product notation simplifies the form of $\Gamma^{-1}(I - \tilde{B})$. That is $\tilde{B} = B \otimes I$ with B as in (2) and $\Gamma = D_W^{-1} \otimes \Sigma$ so

$$\Gamma^{-1}(I - \tilde{B}) = (D_W \otimes \Sigma^{-1})(I - B \otimes I) = (D_W - W) \otimes \Sigma^{-1}$$

Again, the singularity of $D_W - W$ implies that $\Gamma^{-1}(I - B)$ is singular. We denote this distribution by $MCAR(1, \Sigma)$. Again, in practice we work with the proper full conditional distributions in (4), imposing p linear constraints.

To consider remedies to the impropriety, analogous to the univariate case, we rewrite (4) in the general form

$$E(\phi_i \mid \phi_j,\ j \neq i, \Gamma) = R_i \sum B_{ij}\phi_j \tag{6}$$

Now $\Gamma^{-1}(I - \tilde{B})$ is revised to $\Gamma^{-1}(I - \tilde{B_R})$ where $\tilde{B}_R$ has (i, j)th block $R_i B_{ij}$. So, in general the symmetry condition becomes $(\Sigma_i^{-1} R_i B_{ij})^T = \Sigma_j^{-1} R_j B_{ji}$, i.e. $\Sigma_j B_{ij}^T R_i^T = R_j B_{ji}\Sigma_i$. (See Mardia, 1988, expression (2.4) in this regard.)

Once again, if $B_{ij} = b_{ij} I_{p\times p}$ and $b_{ij} = w_{ij}/w_{i+}$ the symmetry condition simplifies to

$$w_{j+}\Sigma_j R_i^T = w_{i+} R_j \Sigma_i \tag{7}$$

Finally, if $\Sigma_i = w_{i+}^{-1}\Sigma$, we obtain $\Sigma R_i^T = R_j \Sigma$ which reveals that we must have $R_i = R_j = R$ yielding

$$\Sigma R^T = R\Sigma. \tag{8}$$

For any arbitrary positive definite Σ, a generic solution to (8) is $R = \rho\Sigma^t$. Hence, regardless of t, (8) introduces a total of $\begin{pmatrix} p+1 \\ 2 \end{pmatrix} + 1$ parameters. So, without loss of generality, we can set $t = 0$, hence $R = \rho I$. Calculation as above yields

$$\Sigma_\phi^{-1} = \Gamma^{-1}(I - \tilde{B}_R) = (D_W - \rho W) \otimes \Sigma^{-1}. \tag{9}$$

Hence, under the same restriction to ρ as in the univariate case, a nonsingular covariance matrix results. We denote this model by $MCAR(\rho, \Sigma)$.

If Σ is constrained to be diagonal with elements σ_l^2, R can be diagonal with elements ρ_l yielding the case of p independent CAR specifications. Routine calculation shows that each ρ_l must therefore satisfy the same restrictions as above to ensure a nonsingular covariance matrix. This modelling introduces $2p$ parameters and is conceptually less satisfying than allowing a more general Σ. However, only if all $\rho_l = \rho$ is this a special version of the general Σ case since a diagonal R with a general Σ does not satisfy (8).

Can we allow a general Σ and a general R? If $B_{ij} = b_{ij}\Sigma^{-1}$ with b_{ij} and Σ_i as above we obtain the condition $R_i^T = R_j$ which only requires $R_i = R$ symmetric. This modelling allows $2\begin{pmatrix} p \\ 2 \end{pmatrix} + 2p$

parameters and yields $E(\phi_i \mid \phi_j,\ j \neq i, \Sigma, R) = R\Sigma^{-1}\sum_j b_{ij}\phi_j$. Now $\tilde{B}_R = B \otimes R\Sigma^{-1}$ so $\Gamma^{-1}(I - \tilde{B}_R) = D_W \otimes \Sigma^{-1} - W \otimes \Sigma^{-1}R\Sigma^{-1} = (D_W \otimes R^{-1} - W \otimes \Sigma^{-1})(I \otimes R)(I \otimes \Sigma^{-1})$. It is evident that the conditions for nonsingularity are not simple. Moreover, unlike the earlier case, the choices for R will depend upon the unknown Σ, making model fitting practically intractable. Making R diagonal does not substantially alleviate the problem. As a result we do not pursue this model further.

However, we can provide a generalization of the $MCAR(\rho, \Sigma)$ model which does permit the introduction of a spatial autoregression coefficient for each component of ϕ_i, i.e. a vector $\rho^T = (\rho_1, \rho_2, \ldots, \rho_p)$. First, suppose we rearrange the rows of the $np \times 1$ vector ϕ to block by components rather than by units. That is, we write $\phi' = (\phi_{11}, \phi_{21}, \ldots, \phi_{n1}, \phi_{12}, \ldots, \phi_{n2}, \phi_{1p}, \ldots, \phi_{np})$ so that $\phi' = P\phi$ where P is orthogonal. But also from (9),

$$\Sigma_{\phi'}^{-1} = \Sigma^{-1} \otimes (D_W - \rho W). \tag{10}$$

Let $D_W^{-1/2} W D_W^{-1/2} = Q\Delta Q^T$ where Δ is diagonal with entries λ_i which are the eigenvalues of $D_W^{-1/2} W D_W^{-1/2}$ and Q is orthogonal. Then, if $T_{\rho_j} = D_W - \rho_j W$ it is evident that $T_{\rho_j} = D_W^{1/2} Q \Omega_j Q^T D_W^{1/2}$ where Ω_j is diagonal with $(\Omega_j)_{ii} = 1 - \rho_j\lambda_i$. Also, $T_{\rho_j} = A_j A_j^T$ where $A_j = D_W^{1/2} Q \Omega_j^{1/2} Q^T$. Note that A_j^{-1} exists if $\Omega_j^{-1/2}$ exists. But if ρ_j satisfies our earlier restrictions i.e. $\rho_j \in (\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, then $1 - \rho_j\lambda_i > 0$ for each i so $\Omega_j^{-1/2}$ exists. Next let $G_j = A_1 A_j^{-1}$, $j = 1, \ldots, p$, and let G be block diagonal with blocks $G_1, \ldots, G_p$. G is evidently full rank provided each ρ_j satisfies the foregoing eigenvalue condition. Then straightforward calculation reveals that

$$G^{-1}(\Sigma^{-1} \otimes T_{\rho_1})(G^{-1})^T = \begin{pmatrix} \Sigma_{11}^{-1} T_{\rho_1} & \Sigma_{12}^{-1} A_1 A_2^T & \ldots & \Sigma_{1p}^{-1} A_1 A_p^T \\ \Sigma_{21}^{-1} A_2 A_1^T & \Sigma_{22}^{-1} T_{\rho_2} & \ldots & \Sigma_{2p}^{-1} A_2 A_p^T \\ \vdots & \vdots & \ldots & \vdots \\ \Sigma_{p1}^{-1} A_p A_1^T & \Sigma_{p2}^{-1} A_p A_2^T & \ldots & \Sigma_{pp}^{-1} T_{\rho_p} \end{pmatrix}. \tag{11}$$

The matrix in (11) is immediately positive definite and can be viewed as the inverse covariance matrix associated with $\Psi' = G\phi'$ where ϕ has the inverse covariance matrix in (10) at $\rho = \rho_1$. Finally, the distribution of $\Psi = (\Psi_1, \ldots, \Psi_n)$ where Ψ_i is $p \times 1$, with $\Psi = P^T\Psi' = P^T G\phi' = P^T G P \phi$ provides a new multivariate CAR specification which we denote by $MCAR(\rho, \Sigma)$. Note that the linear transformation relating Ψ and ϕ is parametric, i.e. it involves the unknown ρ. This class of models has $\binom{p+1}{2} + p$ parameters and reduces to $MCAR(\rho_1, \Sigma)$ when $\rho_j = \rho_1$, $j = 2, \ldots, p$. Note further the interpretation of the diagonal blocks in (11). $\Sigma_{jj}^{-1} T_{\rho_j}$ is the inverse of the conditional covariance matrix of Ψ_j' given $\Psi_l', l = 1. \ldots, p, l \neq j$. Hence, if $\rho_j = 0$, $\Sigma_{jj}^{-1} T_{\rho_j} = \Sigma_{jj}^{-1} D_W$, i.e. $\Psi_{1j}, \Psi_{2j}, \ldots, \Psi_{nj}$ are conditionally independent given all of the other Ψ'. This is analogous to the interpretation of $\rho = 0$ in $CAR(\rho, \sigma^2)$. Such conditional independence is the anticipated conclusion given that we are modelling through the inverse covariance matrix.

As a concluding comment, recent work of Kim *et al.* (2000) develops, for the case $p = 2$, the so-called two fold CAR prior. Rather than specifying the joint distribution through the conditional distributions $\phi_i \mid \phi_j$, $j \neq i$ as we and Mardia (1988) do, they specify the full conditional distributions, $\phi_{il} \mid \phi_{-(il)}$ where $\phi_{-(il)}$ denotes all of the remaining ϕ in ϕ. As well, they introduce ρ to provide a nonsingular inverse joint covariance matrix. Of course, (4) immediately provides $\phi_{il} \mid \phi_{-(il)}$. Moreover, with a proper joint distribution in each case, we can compare the resulting inverse covariance matrices for ϕ. Kim *et al.* specify this distribution blocking the ϕ_{il} by $p(= 2)$ rather than i as we have, obtaining, in our notation,

the inverse covariance matrix

$$\begin{pmatrix} \delta_{11}^{-1}(2D_W + I - \rho_{11}W) & -(\delta_{11}\delta_{12})^{-1/2}(\rho_{10}I + \rho_{13}W) \\ -(\delta_{11}\delta_{12})^{-1/2}(\rho_{10}I + \rho_{13}W) & \delta_{12}^{-1}(2D_W + I - \rho_{12}W) \end{pmatrix}. \tag{12}$$

In (12), δ_{11} and δ_{12} are interpreted as conditional variance parameters for the first and second components of ϕ_i. ρ_{11} and ρ_{12} are autoregressive parameters analogous to our ρ. Finally, ρ_{10} and ρ_{13} are bridging or linking parameters associating ϕ_{il} with $\phi_{i'l}$ and with $\phi_{i'l'}$, $l \neq l'$ respectively. Note that (12) provides a 6 parameter specification.

Working with a general Σ and $R = \rho I$, arranging the inverse covariance matrix in their form, we obtain $\Sigma^{-1} \otimes (D_W - \rho W) =$

$$\begin{pmatrix} \Sigma_{11}^{-1}(D_W - \rho W) & \Sigma_{12}^{-1}(D_W - \rho W) \\ \Sigma_{12}^{-1}(D_W - \rho W) & \Sigma_{22}^{-1}(D_W - \rho W) \end{pmatrix}. \tag{13}$$

Note that (13) provides a four-parameter specification. If w_{i+} is constant across i then (13) is a special case of (12). Choice between these two modelling specification will depend upon the context. Which conditional distribution is more natural to specify? In our applications we prefer to model using Σ, to enable routine implementation of association between the components of the ϕ tempered by ρ. Also, our models are applicable for arbitrary p and fitting (see Section 3) is straightforward, even with increasing dimensions. Presently, the Kim *et al*. approach is limited to $p = 2$.

4. Modelling issues

The MCAR specifications of the previous section are employed in models for spatial random effects arising in a hierarchical model. As is customary, we employ a Bayesian framework to fit and infer with respect to these models. We consider two settings. In the first we have a linear model with continuous data $\boldsymbol{Y}_{ik}, i = 1, \ldots, n,\ k = 1, \ldots, m_i$ where $\boldsymbol{Y}_{ik}$ is a $p \times 1$ vector denoting the kth response at the ith areal unit. The mean of the $\boldsymbol{Y}_{ik}$ is $\boldsymbol{\mu}_{ik}$ where $\mu_{ikj} = (\boldsymbol{X}_{ik})_j \boldsymbol{\beta}^{(j)} + \phi_{ij}$, $j = 1, \ldots, p$. Here $\boldsymbol{X}_{ik}$ is a $p \times s$ matrix with covariates associated with $\boldsymbol{Y}_{ik}$ having jth row $(\boldsymbol{X}_{ik})_j$, $\boldsymbol{\beta}^{(j)}$ is an $s \times 1$ coefficient vector associated with the jth component of the $\boldsymbol{Y}_{ik}$ and ϕ_{ij} is the jth component of the $p \times 1$ vector $\boldsymbol{\phi}_i$. Given $\left\{\boldsymbol{\beta}^{(j)}\right\}$, $\{\boldsymbol{\phi}_i\}$ and V the $\boldsymbol{Y}_{ik}$ are conditionally independent $N(\boldsymbol{\mu}_{ik}, V)$. Adding a prior for $\left\{\boldsymbol{\beta}^{(j)}\right\}$ and V and one of the MCAR models from Section 2 for the $\boldsymbol{\phi}_i$ completes the second stage of the specification. Finally, a hyperprior on the MCAR parameters completes the model.

The second setting merely changes the first stage to a multinomial. Here k disappears and $\boldsymbol{Y}_i$ is assumed to follow a multinomial distribution with sample size n_i and with $(p+1) \times 1$ probability vector $\boldsymbol{\pi}_i$. Working on the logit scale, using cell $p+1$ as the baseline, we set $\log(\frac{\pi_{ij}}{\pi_{i,p+1}}) = \boldsymbol{X}_i^T \boldsymbol{\beta}^{(j)} + \phi_{ij}$, $j = 1, 2, \ldots, p$ with $\boldsymbol{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ interpreted as in the previous paragraph. It is clear that other multivariate first stages could be used such as other multivariate exponential family models.

In each setting, for the specification of the joint distribution of the $\boldsymbol{\phi}_i$ we used as model I, $MCAR(1, \Sigma)$; model II, $MCAR(\rho, \Sigma)$; model III, $MCAR(\boldsymbol{\rho}, \Sigma)$. Model fitting is implemented using a Gibbs sampler with Metropolis updates where needed. The full conditionals for the $\boldsymbol{\beta}$ will typically be normal in the first setting and Metropolis or adaptive rejection sampling (Gilks and Wild, 1991) enables updates for the second setting. For models I and II, the full conditionals for the $\boldsymbol{\phi}_i$ will be likelihood-adjusted versions of the conditional distributions which define the MCAR and are updated as a block. For model III, we can work with either the $\boldsymbol{\phi}$ or the $\boldsymbol{\psi}$ parametrization. With a non-Gaussian first stage, it will be awkward to pull the transformed effects out of the likelihood in order to do the updating. However, with a Gaussian first stage, based upon our limited experience, it is more efficient to work on the transformed

scale. Under the Gaussian first stage, the full conditional for V will be seen to follow an inverse Wishart. Σ will also be seen to follow an inverse Wishart. The ρ do not follow standard distributions; discretization provides the most computationally feasible approach to update them, avoiding Metropolis steps.

We have chosen an illustrative prior for the ρ in the ensuing examples following three criteria. First, we insist that $\rho < 1$ to ensure propriety but allow $\rho = 0.99$. Second, we do not allow $\rho < 0$ since this would violate the similarity of spatial neighbours which we seek. Third, since even moderate spatial dependence requires values of ρ near 1 (see the reply to the discussion in Besag *et al.* (1991)) we place prior mass which favours the upper range of ρ. In particular, we put equal mass on the following 31 values: 0, 0.05, 0.1, ... , 0.8, 0.82, 0.84, ... , 0.90, 0.91, 0.92, ... , 0.99.

Finally, model choice arises in selecting among the three MCAR models. Here, the mean vector is not changed in these investigations. Rather interest resides in comparing the spatial explanations. Multivariate versions of the Gelfand and Ghosh (1998) criterion for multivariate Gaussian and multinomial data, respectively, are employed. Details are provided in that paper and are not presented here. Rather, for each model, we supply the values of the goodness-of-fit term (G), the penalty term (P) and the sum D.

5. ANALYSIS OF THE CHILD GROWTH DATA

Recalling the discussion of Section 2.1, it may be helpful to provide explicit expressions, with obvious notation, for the modelling and the resulting association structure. We have, for the jth child in the ith RMU,

$$\boldsymbol{Y}_{ij} = \begin{pmatrix} (HAZ)_{ij} \\ (WAZ)_{ij} \end{pmatrix} = \boldsymbol{X}_{ij}^T \begin{pmatrix} \boldsymbol{\beta}^{(H)} \\ \boldsymbol{\beta}^{(W)} \end{pmatrix} + \begin{pmatrix} \phi_i^{(H)} \\ \phi_i^{(W)} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{(H)} \\ \epsilon_{ij}^{(W)} \end{pmatrix}. \tag{14}$$

In (14), under say model II,

$$\operatorname{cov}((HAZ)_{ij}, (HAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) = \operatorname{cov}(\phi_i^{(H)}, \phi_{i'}^{(H)}) + V_{11}\mathbf{1}_{i=i', j=j'}$$

$$\operatorname{cov}((WAZ)_{ij}, (WAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) = \operatorname{cov}(\phi_i^{(W)}, \phi_{i'}^{(W)}) + V_{22}\mathbf{1}_{i=i', j=j'}$$

$$\operatorname{cov}((HAZ)_{ij}, (WAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) = \operatorname{cov}(\phi_i^{(H)}, \phi_{i'}^{(W)}) + V_{12}\mathbf{1}_{i=i', j=j'}$$

where $\operatorname{cov}(\phi_i^{(H)}, \phi_{i'}^{(H)}) = (D_W - \rho W)_{ii'}\Sigma_{11}$, $\operatorname{cov}(\phi_i^{(W)}, \phi_{i'}^{(W)}) = (D_W - \rho W)_{ii'}\Sigma_{22}$, and $\operatorname{cov}(\phi_i^{(H)}, \phi_{i'}^{(W)}) = (D_W - \rho W)_{ii'}\Sigma_{12}$. The interpretation of the components of Σ and V, particularly Σ_{12} and V_{12}, is now clear.

We adopted noninformative uniform prior specifications on $\boldsymbol{\beta}^{(H)}$ and $\boldsymbol{\beta}^{(W)}$. For Σ and V we use inverse Wishart priors, i.e. $\Sigma^{-1} \sim W(\Omega_1, c_1)$, $V^{-1} \sim W(\Omega_2, c_2)$ where Ω_1, Ω_2 are $p \times p$ matrices and c_1, c_2 are shape parameters. Since we have no prior knowledge regarding the nature or extent of dependence, we choose Ω_1 and Ω_2 diagonal. The data will inform about the dependence a posteriori. Since the $\boldsymbol{Y}_{ij}$ are centred and scaled on each dimension, setting $\Omega_1 = \Omega_2 = I$ seems appropriate. Finally, we set $c_1 = c_2 = 4$ to provide low precision for these priors. We adopted for ρ_1 and ρ_2 the prior discussed in the previous section. Simulation from the full conditional distributions of the $\boldsymbol{\beta}$ and the $\psi_i, i = 1, \ldots, n$ is straightforward as they are standard normal distributions. Similarly, the full conditionals for V^{-1} and Σ^{-1} are Wishart distributions. We implemented the Gibbs sampler with 10 parallel chains.

Table 1. *Model comparison for child growth data*

Model	*G*	*P*	*D*
I	34 300.69	33 013.10	67 313.79
II	34 251.25	33 202.86	67 454.11
III	34 014.46	33 271.97	67 286.43

Table 2. *Posterior summaries of the dietary covariate coefficients, covariance components and autoregression parameters for the child growth data using model III*

	Height (HAZ)			Weight (WAZ)		
Covariate	2.5%	50%	97.5%	2.5%	50%	97.5%
Global mean	−0.35	−0.16	−0.01	−0.48	−0.25	−0.15
Coconut	0.13	0.20	0.29	0.04	0.14	0.24
Sago	−0.16	−0.07	−0.00	−0.07	0.03	0.12
Sweet potato	−0.11	−0.03	0.05	−0.08	0.01	0.12
Taro	−0.09	0.01	0.10	−0.19	−0.09	0.00
Yams	−0.16	−0.04	0.07	−0.19	−0.05	0.08
Rice	0.30	0.40	0.51	0.26	0.38	0.49
Tinned fish	0.00	0.12	0.24	0.04	0.17	0.29
Fresh fish	0.13	0.23	0.32	0.08	0.18	0.28
Vegetables	−0.08	0.08	0.25	0.02	0.19	0.35
V_{11}, V_{22}	0.85	0.87	0.88	0.85	0.87	0.88
V_{12}	0.60	0.61	0.63	—	—	—
Σ_{11},Σ_{22}	0.30	0.37	0.47	0.30	0.39	0.52
Σ_{12}	0.19	0.25	0.35	—	—	—
ρ_1,ρ_2	0.95	0.97	0.97	0.10	0.80	0.97

Table 1 depicts results on model comparison criteria among the three MCAR models. Model III is preferred, offering sufficient improvement in goodness-of-fit to offset the increased complexity penalty. Summaries of the posterior quantities under model III are shown in Table 2. These were obtained from posterior samples with size of 1000 after running a 10-chain Gibbs sampler for about 30 000 iterations with a burn-in of 5000 iterations and thinning of 30 iterations. Among the dietary factors, high consumption of sago and taro are correlated with lighter and shorter children, while high consumption of rice, fresh fish and coconut are associated with both heavier and taller children. Children from villages with high consumption of vegetables or tinned fish are heavier.

The posterior for the correlation associated with Σ, i.e. $\frac{\Sigma_{12}}{\sqrt{(\Sigma_{11}\Sigma_{22})}}$, has mean 0.67 with interval estimates (0.57,0.75) while the posterior for the correlation associated with V, i.e. $\frac{V_{12}}{\sqrt{(V_{11}V_{22})}}$, has mean 0.71 with interval estimate (0.70,0.72). In addition ρ_1 and ρ_2 differ. Hence the use of the MCAR seems justified as is the advantage to distinct ρ in model III. Posterior medians of the spatial effects ϕ_i^H and ϕ_i^W appear in Figures 1(a) and (b). The difference in spatial patterns between Figures 1(a) and (b) demonstrate the value in employing a bivariate CAR specification.

Table 3. *Model comparison for HLA-B allele data*

Model	G	P	D
I	597.11	1328.04	1925.15
II	583.02	1330.87	1913.89
III	565.95	1324.951	1890.46

6. ANALYSIS OF THE ALLELE FREQUENCY DATA

Recalling the discussion of Section 2.2, allele frequencies are assumed to be multinomially distributed. Suppose that at a given locus, alleles $B_1, B_2, \ldots, B_p, B_{p+1}$ have been identified by serological typing. Each individual has a pair of alleles at that locus which can be homozygous (e.g $B_l/B_l, l = 1, \ldots, p+1$) or heterozygous (e.g. $B_l/B_k, l \neq k = 1, \ldots, p+1$). Let $\pi_1, \pi_2, \ldots, \pi_{p+1}$ be the allele frequencies, r_l be the number of times allele l appears in a heterozygote pair and s_l be the number of times that allele l appears in a homozygote pair. Then the number of occurrences of a single allele B_l will be $n_l = 2s_l + r_l$ and the counts $(Y_1, Y_2, \ldots, Y_{p+1})$ will follow a multinomial distribution, $(Y_1, Y_2, \ldots, Y_{p+1}) \sim \text{Multinomial}(\sum_{l=1}^{p+1} Y_l, \pi_1, \ldots, \pi_{p+1})$, under the assumptions of Hardy–Weinberg equilibrium. In fact, for each areal unit i, we observe a vector of multinomial counts, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i,p+1}) \sim \text{Multinomial}(n_i, \pi_{i1}, \ldots, \pi_{i,p+1})$. The dataset also includes several unit level covariates but for the purpose of this analysis we considered only the effect of altitude, a factor with the following categories: $\leqslant 600$ m, 600–1200 m, 1200–1800 m, and > 1800 m.

The multinomial data was modelled by converting to multiple logits, considering allele $B13$ as the baseline category, i.e. $\log(\frac{\pi_{ij}}{\pi_{i,p+1}}) = \mu_j + \sum_{l=1}^{3} \beta_{jl} x_{il} + \phi_{ij}$ where x_{il} indicates whether or not the ith unit is in the lth altitude category. In the MCAR specification, an inverse Wishart prior, $\Sigma^{-1} \sim W(\Omega, c)$ was adopted where Ω is a $p \times p$ matrix and c the shape parameter. We choose $\Omega^{-1} = I_4$ where I_4 is a four-dimensional identity matrix and $c = 6$ to obtain a rather non-informative prior. For any spatial autoregression parameters, we assume the prior from Section 4. Again the Gibbs sampler was implemented using 10 independent chains run in parallel. After convergence, we collect a final sample from the posterior of size 1000 by taking every 30th value of each chain.

Table 3 compares the fit of the three models. The results show that adopting the model III MCAR prior provides both the smallest goodness-of-fit (G) and the smallest penalty term (P) in the model comparison criterion. This result is also supported by estimates of the posterior distribution of the components of the covariance matrix Σ, given in Table 4. The posteriors reveal non-zero covariances. In fact, converting to correlations, posterior medians range from 0.298 to 0.723. The altitude coefficients indicate that higher altitude encourages the incidence of the other alleles relative to allele $B13$. Finally, the posterior medians of the four sets of spatial effects are displayed in Figures 2(a)–(d). Again, the difference in spatial patterns across the maps reveals the benefit of the multivariate modelling.

ACKNOWLEDGEMENTS

The authors are grateful to Michael Alpers and Kuldeep Bhatia for making the data available. The authors also thank Ivo Mueller for useful discussion regarding the child growth data. The work of the first author was supported in part by NSF DMS 9971206. The work of the second author was supported by Swiss National Foundation grant 3200-57165.99.

Table 4. *Posterior summaries of the altitude coefficients, components of* Σ *and autoregression parameters for HLA-B data using model III*

	Allele	$Bw62 : B13$			$B27 : B13$			$Bw56 : B13$			$Bw60 : B13$		
Altitude (m)		2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5	2.5%	50%	97.5
<600		−0.349	−0.08	0.321	−2.713	−1.506	−0.966	−0.221	0.402	0.741	−0.258	−0.007	0.250
600–1200		1.111	1.830	2.474	1.727	2.438	3.732	1.300	1.898	2.405	0.674	1.521	2.094
1200–1800		1.396	1.748	2.172	1.161	1.679	2.356	1.210	1.694	2.022	1.031	1.451	1.740
>1800		1.172	1.902	2.336	0.747	1.428	2.735	1.363	1.721	2.196	0.914	1.352	1.783
Components of Σ	$Bw62 : B13$	2.384	3.680	5.586	0.587	1.519	2.910	0.820	1.616	2.773	0.866	1.737	2.953
Covariances	$B27 : B13$				2.008	3.159	5.047	−0.092	0.680	1.781	0.320	1.035	2.278
	$Bw56 : B13$							1.309	2.101	3.330	0.505	1.156	2.114
	$Bw60 : B13$										1.039	1.728	2.832
$\rho_1, \rho_2, \rho_3, \rho_4$		0.05	0.40	0.65	0.88	0.97	0.99	0.91	0.97	0.99	0.05	0.82	0.98

REFERENCES

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society, Series B* **36**, 192–236.

BESAG, J., YORK, J. and MOLIE, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

BESAG, J., GREEN, P., HIGDON, D. and MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3–66.

BROOK, D. (1964). On the distinction between the conditional probability and joint probability approaches in the specification of the nearest neighbour systems. *Biometrika* **51**, 481–489.

CLAYTON, D. G., BERNARDINELLI, L. and MONTOMOLI, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology* **22**, 1193–1202.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley.

COLE, T. J. and GREEN, P. J. (1992). Smoothing reference centiles; The LMS method and penalized likelihood. *Statistics in Medicine* **11**, 1305–1319.

GELFAND, A. E. and GHOSH, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions of Pattern Analysis and Machine Intelligence* **6**, 721–741.

GILKS, W. R. and WILD, P. (1991). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C* **41**, 337–348.

HEYWOOD, P. F., SINGLETON, N., ROSS, J. (1988). Nutritional status of young children - The 1982/3 National Nutrition Survey. *Papua New Guinea Medical Journal* **31**, 91–101.

KIM, H., SUN, D. and TSUTAKAWA, R. (2000). A bivariate Bayes method for improving the estimates of mortality rates with 2-fold CAR model. Technical report, University of Missouri, Columbia, MO.

KNORR-HELD, L. and RUE, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, to appear.

LANGFORD, I. H., LEYLAND, A. H., RASBASH, J. and GOLDSTEIN, H. (1999). Multilevel modelling of the geographical distributions of diseases. *Journal of the Royal Statistical Society, Series C, Applied Statistics* **48**, 253–268.

LEYLAND, A., LANGFORD, I., RASBASH, J. and GOLDSTEIN, H. (2000). Multivariate spatial models for event data. *Statistics in Medicine* **19**, 2469–2478.

MARDIA, K. V. (1988). Multi-dimensional multivariate gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**, 265–284.

MUELLER, I., VOUNATSOU, P., ALLEN, B. J. and SMITH, T. (2001). Spatial patterns of child growth in Papua New Guinea and their relation to environment, diet, socio-economic status and subsistence activities. *Annals of Human Biology* **28**, 263–280.

SERJEANTSON, S. W., BOARD, P. G., and BHATIA, K. K. (1992). Population genetics in Papua New Guinea: A perspective on human evolution. In RD Attenborough and MP Alpers (eds.), Human Biology in Papua New Guinea: The Small Cosmos, Oxford: Clarendon, pp. 198–233.

SUN, D., TSUTAKAWA, R. K., KIM, H. and HE, Z. (2000). Bayesian Analysis of Mortality Rates with Disease Maps. *Statistics and Medicine* **19**, 2015–2035.

TING, A., and MORRIS, P. J. (1973). The relationship of six ethnic groups (Chinese, Malys, Indians, New Guinea Highlands and Coastal Natives, Australian Caucasians) based on the HLA system. In J. Dausset and J. Colombani (eds.), *Histocompatibility Testing*, Copenhagen: Munksgaard, p.275.

VOUNATSOU, P., MUELLER, I. (2001). Bayesian modelling of spatial patterns of bivariate indicators of child growth. Submitted.

VOUNATSOU, P., SMITH, T. and GELFAND, A. E. (2000). Spatial modelling of multinomial data with latent structure; an application to geographical mapping of human gene and haplotype frequencies. *Biostatistics* **1**, 177–189.

WATERLOW, J. C. (1992). *Protein-energy, Malnutrition.* London: Arnold.

[*Received July 10, 2001*; *revised March 21, 2002*; *accepted for publication April 16, 2002*]