NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Bayesian modelling of spatial compositional data

by

Håkon Tjelmeland and Kjetill Vassmo Lund

PREPRINT
STATISTICS NO. 1/2001

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S1-2001.ps
Håkon Tjelmeland has homepage: http://www.math.ntnu.no/~haakont
E-mail: haakont@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

Bayesian modelling of spatial compositional data

Håkon Tjelmeland and Kjetill Vassmo Lund
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway

Abstract

Compositional data are vectors of proportions, specifying fractions of a whole. Aitchison (1986) defines logistic normal distributions for compositional data by applying a logistic transformation and assuming the transformed data to be multi-normal distributed. In this paper we generalise this idea to spatially varying logistic data and thereby define logistic Gaussian fields. We consider the model in a Bayesian framework and discuss appropriate prior distributions. We consider both complete observations and observations of only subcompositions or individual proportions, and discuss the resulting posterior distributions. In general, the posterior can not be analytically handled, but the Gaussian base of the model allows us to define efficient Markov chain Monte Carlo algorithms. We use the model to analyse a data set of sediments in an arctic lake. These data has also previously been considered, but then without taking the spatial aspect into account.

Key words: compositional data, logistic Gaussian fields, spatial Bayesian model, logistic transformation.

1 Introduction

Compositional data are vectors of proportions, specifying D fractions of a whole. Thus, for $x = (x_1, \ldots, x_D)^T$ to be a compositional vector one must have $x_i > 0$ for $i = 1, \ldots, D$ and $x_1 + \ldots + x_D = 1$. In Aitchison and Shen (1980) and Aitchison (1980, 1982) the logistic normal distribution is introduced as a flexible parametric model for analysis of compositional data. The model is defined from a multi-normal distribution via a logistic transformation. Many of the nice properties of the multi-normal class is inherited by the logistic normal family and this allows extensive analytical treatment also for logistic normal distributions.

Spatial models for proportions are also of interest in many situations. Let $x(u) = (x_1(u), \ldots, x_D(u))^T$ be a stochastic vector field, where $u \in \Re^k$ denotes a location. For this to be a spatial compositional field the above conditions must apply for each position u, i.e. $x_i(u) > 0$ for all $u \in \Re^k$ and $i = 1, \ldots, D$ and $x_1(u) + \ldots, x_D(u) = 1$ for all locations $u \in \Re^k$. For example, with $D = 4$ and $k = 3$, $x_1(u), \ldots, x_4(u)$ may represent volume proportions of oil, gas, water and rock as a function of location in a petroleum reservoir. Alternatively, as in our example in Section 6 where sediments in a lake is of interest, $x_1(u), \ldots, x_D(u)$ may represent proportions of sand, silt and clay where $u \in \Re^2$ is geographical location.

Some spatial generalisations of the logistic normal family are discussed in the literature. In Pawlowsky and Burger (1992), spatial interpolation of compositional data is considered via logistic transformation in each location to obtain data in $\Re^{D-1}$, followed by a cokriging procedure. Thus, a model of correlated Gaussian fields is adopted after the transformation. In Billheimer et al. (1997a, 1997b) discrete spatially correlated compositional data is modelled via underlying (unobserved) Gaussian Markov random fields. However, this formulation allows very little analytical computations and for inference they resort to Markov chain Monte Carlo (MCMC) procedures updating only one or a small number of variables at a time.

In the present paper we revisit the formulation in Pawlowsky and Burger (1992), but considers it in a broader setting. First, we discuss also the problem of parameter estimation by including the model in a hierarchical Bayesian framework. Second, Pawlowsky and Burger (1992) limits the attention to complete observations, i.e. all proportions are observed in a number of locations, whereas in this paper we also consider two forms of partial observations; namely observation of subcompositions and of individual proportions. A subcomposition is the relative proportions of only some of the components. For example, for the oil, gas, water and rock situation mentioned above, one may easily observe the relative volumes of oil, water and gas, whereas to get the volume proportion of rock is a more expensive task. As discussed in Aitchison (1986), subcompositional data fits very nicely into the logistic normal family and allows extensive analytical treatment. Observations of individual proportions are more problematic and MCMC is necessary to handle this case. However, the Gaussian base can be used to define efficient algorithms updating large blocks of variables simultaneously.

The paper is organised as follows. In the next section we introduce some basic notation and discuss appropriate sample spaces. In Section 3 logistic Gaussian fields are defined as a generalisation of the logistic normal model of Aitchison (1986) and, in Section 4, we discuss properties of the corresponding conditional fields given complete or partial observations. In Section 5 a fully Bayesian model for spatial compositional data is defined by introducing appropriate hyper-priors. Moreover, efficient simulation algorithms for the resulting posterior is defined. In Section 6, we revisit a dataset of sediments in an arctic lake (Coakley and Rust, 1968). The dataset is previously analysed (Aitchison, 1986; Iyengar and Dey, 1996), but then ignoring the spatial structure of the data. Finally, Section 7 provides conclusions.

2 Preliminaries

In this section we introduce some basic notation and discuss the logistic transformation. Most of the treatment is based on the presentation in Aitchison (1986) and a more thorough discussion can be found there.

Let $x(u) = (x_1(u), \dots, x_D(u))^T; u \in \Re^k$ be a compositional field, i.e. $x_i(u)$ gives the proportion of substance i in location u. Thus, the sample space of the D-dimensional composition vector $x(u)$ is

$$\mathcal{L}^D = \left\{ x(u) = (x_1(u), \dots, x_D(u))^T \,\middle|\, x_i(u) > 0, i = 1, \dots, D \text{ and } \sum_{i=1}^{D} x_i(u) = 1 \right\} \subset \Re^D. \quad (1)$$

The unit sum constraint implies that any $d = D - 1$ dimensional sub-vector uniquely specifies the last component. Of the same reason the probability density for $x(u)$ is degenerate. Thereby it is natural to focus on a vector, $z(u)$ say, containing only the d first components of $x(u)$. The one-to-one relation between $x(u)$ and $z(u)$ becomes

$$z(u) = A_d x(u) \qquad \Leftrightarrow \qquad x(u) = B_d z(u) + c_d, \quad (2)$$

where $A_d \in \Re^{d \times D}$ consists of a $d \times d$ identity matrix with an extra column of zeros,

$$(A_d)_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

$B_d \in \Re^{D\times d}$ is a $d \times d$ identity matrix with an extra row of -1's,

$$(B_d)_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \text{ and } i \leq d, \\ -1 & \text{if } i = D \end{cases} \tag{4}$$

and $c_d = (0, 0, \ldots, 0, 1)^T \in \Re^D$. The sample space of $z(u)$ becomes

$$\mathcal{S}^d = \left\{ z(u) = (z_1(u), \ldots, z_d(u))^T \,\middle|\, z_i(u) > 0, i = 1, \ldots, d \text{ and } \sum_{i=1}^{d} z_i(u) < 1 \right\}. \tag{5}$$

The logistic transformation defines a one-to-one relation between $\mathcal{S}^d$ and $\Re^d$. With $z(u) \in \mathcal{S}^d$ and $y(u) = (y_1(u), \ldots, y_d(u))^T \in \Re^d$, the logistic transformation reads

$$z_i(u) = \frac{e^{y_i(u)}}{1 + \sum_{k=1}^{d} e^{y_k(u)}} \quad \text{for} \quad i = 1, \ldots, d \tag{6}$$

and the corresponding inverse transformation is

$$y_i(u) = \ln \left[\frac{z_i(u)}{1 - \sum_{k=1}^{d} z_k(u)} \right] \quad \text{for} \quad i = 1, \ldots, d. \tag{7}$$

To simplify notation, we write these and similar relations in a vector format. Thus, (6) and (7) become

$$z(u) = \frac{e^{y(u)}}{1 + j_d^T e^{y(u)}} \qquad \Leftrightarrow \qquad y(u) = \ln \left[\frac{z(u)}{1 - j_d^T z(u)} \right], \tag{8}$$

where $j_d \in \Re^d$ is a vector with all elements equal to one and the exponential and logarithm functions should be interpreted as acting component-wise. Throughout this paper we consistently use $x(u)$, $y(u)$ and $z(u)$ to denote stochastic fields on $\mathcal{L}^D$, $\Re^d$ and $\mathcal{S}^d$, respectively, and let their relation be given by (2) and (8). Thus, if the distribution of one of the fields is specified, the distribution of the other two is also defined.

Subcompositions and individual proportions can be obtained from $x(u)$ via selection matrices. A matrix $S \in \Re^{C\times D}$ is called a selection matrix if it has C unity elements, of which exactly one is in each row and at most one in each column, and all other elements are equal to zero. For a given selection matrix, S, the corresponding vector of individual proportions at location u is given by $Sx(u)$. A rescaling to obtain unit sum is necessary to get the corresponding subcomposition, $x_S(u)$, i.e.

$$x_S(u) = \frac{Sx(u)}{j_C^T Sx(u)}. \tag{9}$$

Note that $x_S(u) \in \mathcal{L}^C$, so that $x_S(u)$ is a C-dimensional compositional field. Thus, corresponding quantities $z_S(u) \in \mathcal{S}^c$ and $y_S(u) \in \Re^c$ where $c = C - 1$ can be defined by (2) and (8) by replacing $x(u)$, $y(u)$ and $z(u)$ by $x_S(u)$, $y_S(u)$ and $z_S(u)$, respectively. Then, quite remarkably, a linear relation between $y(u)$ and $y_S(u)$ results,

$$y_S(u) = Q_S y(u) \quad \text{where} \quad Q_S = B_c^T S B_d H_d^{-1} \in \Re^{c\times d} \ \ , \ \ H_d = I_d + J_d, \tag{10}$$

where I_d is the $d \times d$ identity matrix and J_d is a $d \times d$ matrix with all elements equal to one. It should be noted that the definition of $y_S(u)$ from $y(u)$ goes via a series of transformations, several of which are non-linear, so this simple linear relation is not at all obvious. The result is by no means new, it is discussed in Aitchison (1986) and forms there the main motivation for definition of logistic normal distributions. Likewise, this property is essential for the tractability of logistic Gaussian fields discussed in this paper. Since (10) is so fundamental in the following sections and as Aitchison (1986) contains no clearly written proof of the result we include one in Appendix A.

A particular subclass of selection matrices occurs for $C = D$. Then the denominator in (10) becomes one and the elements of $x_S(u)$ is just a permuted version of the elements in $x(u)$. Thus, a selection matrix with $C = D$ is called a permutation matrix, we denote it by P and the permuted variant of $x(u)$ by $x_P(u) = Px(u)$. One should note that $y_P(u) \neq Py(u)$ unless $P = I_D$, the relation between $y(u)$ and $y_P(u)$ is given by (10) with P replacing S.

3 Logistic Gaussian fields

Correspondingly to how logistic normal distributions is defined from multi-normal distributions in Aitchison (1986), we now define logistic Gaussian fields from correlated Gaussian fields.

Definition 1 *Let $y(u) = (y_1(u), \ldots, y_d(u))^T; u \in \Re^k$ be d (correlated) Gaussian fields with mean function $\mu(u) = E[y(u)] = (\mu_1(u), \ldots, \mu_d(u))^T$ and covariance structure given by $c(u, u') = Cov[y(u), y(u')] = (c_{ij}(u, u'))_{i,j=1}^d$. Furthermore, for each $u \in \Re^k$, let $z(u)$ and $x(u)$ be defined from $y(u)$ by (8) and (2), respectively. We then call $x(u) = (x_1(u), \ldots, x_D(u))^T; u \in \Re^k$, or equivalently $z(u) = (z_1(u), \ldots, z_d(u))^T$; $u \in \Re^k$, a logistic Gaussian field with parameter functions $\mu(\cdot)$ and $c(\cdot, \cdot)$.*

Logistic Gaussian fields have several important properties which are direct generalisations of corresponding features for logistic normal distributions. We discuss in the following.

In the definition of logistic Gaussian fields the last proportion, $x_D(u)$, is obviously treated differently from the other components. However, it should be noted that a fully symmetric definition gives the same model class. More precisely, let $\{v(u) = (v_1(u), \ldots, v_D(u))^T;\ u \in \Re^k\}$ be D correlated Gaussian fields with mean function $\mu_0(u) \in \Re^D$ and correlation structure $c_0(u, u') \in \Re^{D \times D}$, and set $x(u) = e^{v(u)}/(j_D^T e^{v(u)})$ for each $u \in \Re^k$. Then $x(u)$ is a logistic Gaussian field with parameter functions $\mu(u) = B_d^T \mu_0(u)$ and $c(u, u') = B_d^T c_0(u, u') B_d$. This result is immediate by transforming $x(u)$ via $z(u)$ to the corresponding $y(u)$ by (2) and (8) and noting that this gives $y(u) = B_d v(u)$.

A logistic Gaussian field may be interpreted as a prior distribution in a Bayesian framework. In some situations it is then natural to restrict attention to fields where the different proportions are a priori exchangeable, i.e. to require

$$\{x(u); u \in \Re^k\} \stackrel{d}{=} \{Px(u); u \in \Re^k\} \tag{11}$$

for any permutation matrix P. This is fulfilled if and only if $\mu(u) = 0$ for all $u \in \Re^k$ and $c(u, u')$ has the form

$$c(u, u') = \sigma^2 \alpha(u, u') H_d/2, \tag{12}$$

where $\sigma^2 > 0$ is the variance of the Gaussian fields, $\alpha(\cdot,\cdot)$ is a scalar spatial correlation function and H_d is as specified in (10). It should be noted that this $c(\cdot,\cdot)$ falls within the class of intrinsic correlations structures (see for example Wackernagel (1995), Ch. 22), for which it is required that

$$c(u, u') = \alpha(u, u')\Psi, \tag{13}$$

where $\alpha(\cdot,\cdot)$ is as above and Ψ is a $d \times d$ covariance matrix. Intrinsic correlation structures can often be reasonable in applications and has several computational advantages. In the next section we limit the attention to this situation when defining a fully Bayesian model for compositional data.

Unconditional simulation of logistic Gaussian fields in a set of grid locations is straightforward from its definition; first simulate the correlated Gaussian field $y(u)$ in the given array and then use the logistic transformation in (8) for each grid location. To simulate correlated Gaussian fields many different methods exist, see Cressie (1993) for a overview. Simulation of correlated Gaussian fields become especially simple when an intrinsic covariance structure is assumed, as it can then be expressed as a pointwise linear combination of independent Gaussian fields, see Wackernagel (1995).

4 Conditional logistic Gaussian fields

Properties of logistic Gaussian fields conditioned to data depend on the type of the observations available. In the following, we first discuss complete and subcompositional data and thereafter observation of individual proportions.

4.1 Complete and subcompositional observations

Let $x(u)$ be a logistic Gaussian field with parameter functions $\mu(\cdot)$ and $c(\cdot,\cdot)$. Assume that complete or subcompositional observations are available in n locations, $u_1, \ldots, u_n$. Thus, to each u_i there is an associated selection matrix $S_i \in \Re^{C_i \times D}$ and a corresponding observation vector $x_{S_i}(u_i)$. Complete observations in location u_i corresponds to setting S_i equal to the identity matrix (or some other permutation matrix).

From the definition and properties discussed in the previous sections it follows directly that $x(u)|x_{S_1}(u_1), \ldots, x_{S_n}(u_n)$ is also logistic Gaussian; The one-to-one relation between $x_{S_i}(u_i)$ and $y_{S_i}(u_i)$ gives that conditioning to the vectors $x_{S_1}(u_1), \ldots, x_{S_n}(u_n)$ is equivalent to conditioning on $y_{S_1}(u_1), \ldots, y_{S_n}(u_n)$. From the linear relation in (10) is then follows that $y(u)|y_{S_1}(u_1), \ldots, y_{S_n}(u_n)$ is correlated Gaussian fields and thereby, from the definition of logistic Gaussian fields, that $x(u)|x_{S_1}(u_1), \ldots, x_{S_n}(u_n)$ is a logistic Gaussian field. The parameter functions of the conditional field is also readily available as the mean and covariance functions of the corresponding Gaussian y-field. Conditional simulation of $x(u)$ in a set of grid locations is thereby also straightforward; first simulate the Gaussian distributed $y(u)$ conditioned to $y_{S_1}(u_1), \ldots, y_{S_n}(u_n)$ and then use the logistic transformation in (8) for each grid location. A conditional simulation of $y(u)$ is most efficiently obtained from an unconditional sample via a kriging procedure, see for example Cressie (1993), Ch. 3.6.2.

4.2 Observation of individual proportions

Again let $x(u)$ be a logistic Gaussian field with parameter functions $\mu(\cdot)$ and $c(\cdot,\cdot)$, but now assume that observations of individual proportions are available in the n locations, $u_1,\dots,u_n$. Thus, to each u_i there is again associated a selection matrix, $S_i \in \Re^{C_i\times D}$, but now the observed values are the vectors $S_1x(u_1),\dots,S_nx(u_n)$. In this situation there is no easy transformation of the data to corresponding values for $y(u)$ and the resulting conditional distribution for $x(u)$ is neither logistic Gaussian nor analytically tractable. Thus, properties of the conditional field must be obtained by generating conditional realisations of $x(u)$ by Markov chain Monte Carlo (MCMC). It should be noted that it suffices to use MCMC to generate samples of $x(u_1),\dots,x(u_n)|S_1x(u_1),\dots,S_nx(u_n)$. If samples of the logistic field $x(u)$ for locations $u \notin \{u_1,\dots,u_n\}$ are of interest, these can thereafter be generated directly by conditioning to simulated values for $x(u_1),\dots,x(u_n)$ by the procedure discussed in Section 4.1.

As some components of $x(u_1),\dots,x(u_n)$ are identical to the elements of $S_1x(u_1),\dots,$ $S_nx(u_n)$, sampling $x(u_1),\dots,x(u_n)|S_1x(u_1),\dots,S_nx(u_n)$ actually means to generate samples of the elements in $x(u_1),\dots,x(u_n)$ which are not specified by $S_1x(u_1),\dots,S_nx(u_n)$. Therefore, let T_i be a selection matrix which picks out the non-observed elements of $x(u_i)$, so that

$$P_i = \begin{bmatrix} S_i \\ T_i \end{bmatrix} \tag{14}$$

is a permutation matrix, by which one can obtain $x(u_i)$ from $S_ix(u_i)$ and $T_ix(u_i)$ via

$$x(u_i) = P_i^T \begin{bmatrix} S_ix(u_i) \\ T_ix(u_i) \end{bmatrix}. \tag{15}$$

Moreover, from the unit sum constraint in each location it follows that the last component in $T_ix(u_i)$ is given from the first $D-C_i-1$ elements of $T_ix(u_i)$ together with $S_ix(u_i)$. Thus, variables to be simulated are $\widetilde{z}_1 = A_{b_1}T_1x(u_1),\dots,\widetilde{z}_n = A_{b_n}T_nx(u_n)$, where $b_i = D-C_i-1$ for $i=1,\dots,n$, which has conditional distribution

$$f(\widetilde{z}_1,\dots,\widetilde{z}_n|S_1x(u_1)=\widetilde{x}_1,\dots,S_nx(u_n)=\widetilde{x}_n) \propto f(z_1,\dots,z_n), \tag{16}$$

where $f(z_1,\dots,z_n)$ denotes the unconditional distribution of $z(u_1),\dots,z(u_n)$ and

$$z_i = (z_{i1},\dots,z_{id})^T = A_dP_i^T\left(B_d\begin{bmatrix}\widetilde{x}_i \\ \widetilde{z}_i\end{bmatrix} + c_d\right) \quad \text{for } i=1,\dots,n. \tag{17}$$

Using the transformation formula, an expression for $f(z_1,\dots,z_n)$ is easily found from the corresponding joint Gaussian distribution of $y(u_1),\dots,y(u_n)$. It reads

$$f(z_1,\dots,z_n) = \frac{\exp\left\{-\frac{1}{2}(y(z_1,\dots,z_n)-\mu)^T\Sigma^{-1}(y(z_1,\dots,z_n)-\mu)\right\}}{(2\pi)^{\frac{nd}{2}}\sqrt{|\Sigma|}\prod_{i=1}^n\left[\left(\prod_{j=1}^d z_{ij}\right)(1-j_d^Tz_i)\right]}, \tag{18}$$

where $y(z_1,\dots,z_n)$ is a vector of y_i's corresponding to the z_i's, i.e.

$$y(z_1,\dots,z_n) = \left(\ln\left[\frac{z_1^T}{1-j_d^Tz_1}\right],\dots,\ln\left[\frac{z_n^T}{1-j_d^Tz_n}\right]\right)^T, \tag{19}$$

and $\mu \in \Re^{nd}$ and $\Sigma \in \Re^{nd \times nd}$ are corresponding expectation vector and covariance matrix for $y(z_1, \dots, z_n)$, respectively, i.e.

$$\mu = (\mu(u_1)^T, \dots, \mu(u_n)^T)^T \tag{20}$$

and Σ is made up of $d \times d$ blocks

$$\Sigma = \begin{bmatrix} c(u_1, u_1) & \cdots & c(u_1, u_n) \\ \vdots & & \vdots \\ c(u_n, u_1) & \cdots & c(u_n, u_n) \end{bmatrix}. \tag{21}$$

With an intrinsic covariance structure, as in (13), the computation of $f(z_1, \dots, z_n)$ can be simplified by writing $\Sigma = \Phi \otimes \Psi$, where

$$\Phi = \begin{bmatrix} \alpha(u_1, u_1) & \cdots & \alpha(u_1, u_n) \\ \vdots & & \vdots \\ \alpha(u_n, u_1) & \cdots & \alpha(u_n, u_n) \end{bmatrix} \in \Re^{n \times n} \tag{22}$$

and $\otimes$ denotes the Kronecker product. Thus, the inverse and determinant of Σ is given by $\Sigma^{-1} = \Phi^{-1} \otimes \Psi^{-1}$ and $|\Sigma| = |\Phi|^d |\Psi|^n$, respectively.

Direct simulation from (16) is not feasible and one has to resort to MCMC and the Metropolis-Hastings algorithm. For an introduction to this, see Besag et al. (1995) and references therein. Our target distribution $f(\widetilde{z}_1, \dots, \widetilde{z}_n | S_1 x(u_1) = \widetilde{x}_1, \dots, S_n x(u_n) = \widetilde{x}_n)$ typically has strong correlations between variables. To avoid a very slow mixing Markov chain, one therefore ought to propose changes in all variables simultaneously. The Gaussian base of the model enables construction of this as an independent proposal (Tierney, 1994) algorithm. To see how to do this, first observe that to propose new values for $\widetilde{z}_1, \dots, \widetilde{z}_n$ is equivalent to proposing values for $z_{T_1}(u_1), \dots, z_{T_n}(u_n)$ or for $y_{T_1}(u_1), \dots, y_{T_n}(u_n)$, where the one-to-one relation between the two first is

$$\widetilde{z}_i = (1 - j_{C_i}^T \widetilde{x}_i) z_{T_i}(u_i) \quad \text{and} \quad z_{T_i}(u_i) = \frac{\widetilde{z}_i}{1 - j_{C_i}^T \widetilde{x}_i}. \tag{23}$$

As an approximation to sampling $y_{T_1}(u_1), \dots, y_{T_n}(u_n)$ conditioned to the observed individual proportions, one can then sample the $y_{T_i}(u_i)$'s conditioned only to the corresponding subcompositions $y_{S_1}(u_1), \dots, y_{S_n}(u_n)$. One then remains within Gaussianity and this can be used as a proposal distribution in a Metropolis-Hastings algorithm. Thus, to generate potential new values $\widetilde{z}_1', \dots, \widetilde{z}_n'$ one first sample $y_{T_1}(u_1)', \dots, y_{T_n}(u_n)'$ from the Gaussian distribution $f(y_{T_1}(u_1), \dots, y_{T_n}(u_n) | y_{S_1}(u_1), \dots, y_{S_n}(u_n))$ and thereafter compute corresponding transformed values $z_{T_i}(u_i)'$ and $\widetilde{z}_i'; i = 1, \dots, n$ by (8) and (23). The resulting proposal density for $\widetilde{z}_1', \dots, \widetilde{z}_n'$, which is needed to compute the Metropolis-Hastings acceptance probability, is easily obtained by the transformation formula from the Gaussian density for $y_{T_1}(u_1)', \dots, y_{T_n}(u_n)'$ and reads

$$q(\widetilde{z}_1', \dots, \widetilde{z}_n') \propto \frac{\exp\left\{-\frac{1}{2}(y(\widetilde{z}_1', \dots, \widetilde{z}_n') - \mu_{T|S})^T \Sigma_{T|S}^{-1} (y(\widetilde{z}_1', \dots, \widetilde{z}_n') - \mu_{T|S})\right\}}{\sqrt{|\Sigma_{T|S}|} \prod_{i=1}^n \left[\left(\prod_{j=1}^{c_i} \widetilde{z}_{ij}'\right)(1 - j_{c_i}^T \widetilde{z}_i')\right]}, \tag{24}$$

where $y(\tilde{z}'_1,\dots,\tilde{z}'_n)$ is a vector of the $y'_{T_i}(u_i)$'s generated, i.e.

$$y(\tilde{z}'_1,\dots,\tilde{z}'_n) = \left(\ln\left[\frac{\tilde{z}'_1}{1-j_{C_1}^T\tilde{x}_1 - j_{b_i}^T\tilde{z}'_1}\right]^T, \dots, \ln\left[\frac{\tilde{z}'_n}{1-j_{C_n}^T\tilde{x}_n - j_{b_n}^T\tilde{z}'_n}\right]^T \right)^T, \tag{25}$$

and $\mu_{S|T}$ and $\Sigma_{S|T}$ are the conditional mean vector and covariance matrix for $y(\tilde{z}'_1,\dots,\tilde{z}'_n)$. The $\mu_{T|S}$ and $\Sigma_{T|S}$ are given from standard multi-normal theory and, introducing the block diagonal matrices

$$Q_S = \begin{bmatrix} Q_{S_1} & 0 & \cdots & 0 \\ 0 & Q_{S_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & Q_{S_n} \end{bmatrix} \in \Re^{(\sum_{i=1}^n c_i)\times nd} \tag{26}$$

and

$$Q_T = \begin{bmatrix} Q_{T_1} & 0 & \cdots & 0 \\ 0 & Q_{T_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & Q_{T_n} \end{bmatrix} \in \Re^{(\sum_{i=1}^n b_i)\times nd}, \tag{27}$$

they can be expressed as

$$\mu_{T|S} = \mu_T + \Sigma_{TS}\Sigma_{SS}^{-1}(y_S - \mu_S) \quad \text{and} \quad \Sigma_{T|S} = \Sigma_{TT} - \Sigma_{TS}\Sigma_{SS}^{-1}\Sigma_{ST}, \tag{28}$$

where $\mu_T = Q_T\mu$, $\mu_S = Q_S\mu$, $\Sigma_{SS} = Q_S\Sigma Q_S^T$, $\Sigma_{TT} = Q_T\Sigma Q_T^T$, $\Sigma_{TS} = Q_T\Sigma Q_S^T$ and $\Sigma_{ST} = Q_S\Sigma Q_T$, with μ and Σ given by (20) and (21), respectively.

The motivation for the above defined independent proposal Metropolis-Hastings algorithm is to approximate $f(y_{T_1}(u_1),\dots,y_{T_n}(u_n)|S_1x(u_1),\dots,S_nx(u_n))$ with $f(y_{T_1}(u_1),\dots,y_{T_n}(u_n)|x_{S_1}(u_1),\dots,x_{S_n}(u_n))$. However, it is clear that this approximation will deteriorate as n grows, with lower acceptance probabilities and slower convergence as a result. For n large enough, a better strategy for proposing potential new values $y_{T_1}(u_1)',\dots,y_{T_n}(u_n)'$, and thereby $\tilde{z}'_1,\dots,\tilde{z}'_n$, is to let $y_{T_i}(u_i)'$ be a linear combination of the current value and the value used above. More precisely, to set

$$y_{T_i}(u_i)' = \sqrt{1-\gamma^2}\,(y_{T_i}(u_i) - \mu_c(u_i)) + \gamma(y_{T_i}(u_i)^\star - \mu_c(u_i)) \quad \text{for} \quad i = 1,\dots,n, \tag{29}$$

where $\gamma \in (0,1]$, $\mu_c(u_i) = \mathrm{E}[y_{T_i}(u_i)|y_{S_1}(u_1),\dots,y_{S_n}(u_n)]$ and $y_{T_1}(u_1)^\star,\dots,y_{T_n}(u_n)^\star$ are a sample from the Gaussian density $f(y_{T_1}(u_1),\dots,y_{T_n}(u_n)|x_{S_1}(u_1),\dots,x_{S_n}(u_n))$. It should be noted that the weighting used in (29) would have given unity acceptance probability if the goal had been to sample from $f(y_{T_1}(u_1),\dots,y_{T_n}(u_n)|x_{S_1}(u_1),\dots,x_{S_n}(u_n))$. Even if this is not our target distribution, this still seems like the natural linear combination to use. The algorithmic parameter γ specifies how large change to propose and $\gamma = 1$ gives back the independent proposal algorithm discussed above. The resulting proposal density for $\tilde{z}'_1,\dots,\tilde{z}'_n$, which now depends on the current values $\tilde{z}_1,\dots,\tilde{z}_n$ is again easily found and becomes

$$q(\tilde{z}'_1,\dots,\tilde{z}'_n|\tilde{z}_1,\dots,\tilde{z}_n) \propto$$

$$\frac{\exp\left\{-\frac{1}{2}(y(\tilde{z}'_1,\dots,\tilde{z}'_n) - \mu(\tilde{z}_1,\dots,\tilde{z}_n))^T\Sigma_{T|S}^{-1}(y(\tilde{z}'_1,\dots,\tilde{z}'_n) - \mu(\tilde{z}_1,\dots,\tilde{z}_n))\right\}}{\sqrt{|\Sigma_{T|S}|}\ \prod_{i=1}^n\left[\left(\prod_{j=1}^{c_i}\tilde{z}'_{ij}\right)(1 - j_{c_i}^T\tilde{z}'_i)\right]}. \tag{30}$$

5 A fully Bayesian model

In the previous sections we have assumed the parameter functions of the logistic Gaussian field to be known. In practice, this will of course typically not be the situation. In stead, they have to estimated from data, or a Bayesian formulation can be adopted. Here, we concentrate on the latter approach. No prior distributions for the parameter functions seem to exist which allow full analytical treatment of the resulting posterior. Thus, we have to resort to MCMC to explore the posterior. In the following, our focus is on how to choose prior distributions which are both flexible and give a posterior which allows effective updating of large blocks of parameters.

5.1 Prior distribution

For the mean function, $\mu(u)$, we follow the Bayesian kriging (Omre, 1987; Hjort and Omre, 1994) framework and assume the following regression form,

$$\mu_\beta(u) = F(u)\beta, \tag{31}$$

where $\beta \in \Re^p$ is a vector of unknown parameters and $F(u) \in \Re^{d\times p}$ is a matrix of known regressor functions. Via the choice of $F(u)$ one may for example model linear trends and include the effect of explanatory variables, see also the discussion in Billheimer and Guttorp (1995). Still following the choice in Bayesian kriging, we assume a Gaussian prior distribution for the vector β,

$$\beta \sim \mathrm{N}_p(\mu_0, \Sigma_0), \tag{32}$$

where μ_0 and Σ_0 are parameters to be specified.

To obtain a flexible parametric form for the covariance function, $c(u, u')$, we assume an intrinsic structure with a parametric form for the spatial term, i.e.

$$c_{\theta,\Psi}(u, u') = \alpha_\theta(u, u')\Psi. \tag{33}$$

Different choices for the spatial correlation function exist, the generalised exponential and the Matern correlation functions are two possibilities (Cressie, 1993). Moreover, as computational considerations does not favour a specific choice, we do no closer specification at this stage. To obtain a fully specified Bayesian model, it remains to specify priors for θ and Ψ. The natural candidate for Ψ is to let Ψ^{-1} have a Wishart distribution,

$$\Psi^{-1} \sim \mathcal{W}_d(p, P), \tag{34}$$

where the scalar $p > d$ and the positive definite $d \times d$ matrix P are parameters to be specified. This is the conjugate prior distribution in the Gaussian case and, as our model also has a Gaussian base, facilitates construction of an MCMC algorithm for the posterior distribution. For θ we assume some prior $\pi(\theta)$.

5.2 Simulation algorithm

In this section we discuss how to sample from the posterior distribution corresponding to the prior defined above. As in Section 4, we assume the data to be exact and discuss conditioning on complete observations, subcompositional observations and observation of individual proportions. Unlike the situation in Section 4, all three cases must be treated differently and in the following we discuss each situation in turn

5.2.1 Complete observations

Let $x(u)|\beta, \Psi, \theta$ be a logistic Gaussian field and let the parameters β, Ψ and θ have prior distributions as specified above. Assume complete observations in n locations; $x(u_1), \ldots, x(u_n)$. As discussed in Section 4.1, conditioning on $x(u_1), \ldots, x(u_n)$ is equivalent to conditioning on the corresponding $y(u_1), \ldots, y(u_n)$. Thus, defining $y = [y(u_1)^T \ldots y(u_n)^T]^T$, the posterior distribution of interest can be expressed as

$$\pi(\beta, \Psi, \theta | x(u_1), \ldots, x(u_n)) = \pi(\beta, \Psi, \theta | y) \propto \pi(\theta)\pi(\Psi)\pi(\beta) f(y|\beta, \Psi, \theta), \tag{35}$$

where $f(y|\beta, \Psi, \theta)$ is a multi-normal density with expectation vector and covariance matrix

$$\mathrm{E}[y|\beta, \Psi, \theta] = F\beta \quad \text{where} \quad F = \begin{bmatrix} F(u_1) \\ \vdots \\ F(u_n) \end{bmatrix} \tag{36}$$

and

$$\Phi(\theta) \otimes \Psi \quad \text{where} \quad \Phi(\theta) = \begin{bmatrix} \alpha_\theta(u_1, u_1) & \cdots & \alpha_\theta(u_1, u_n) \\ \vdots & & \vdots \\ \alpha_\theta(u_n, u_1) & \cdots & \alpha_\theta(u_n, u_n) \end{bmatrix} \in \Re^{n \times n}, \tag{37}$$

respectively. To construct an MCMC algorithm for $\pi(\beta, \Psi, \theta | x(u_1), \ldots, x(u_n))$, it is natural to update the three β, Ψ and θ separately. In the following we discuss updates for each of them.

To update β it is natural to consider the corresponding full conditional, $\pi(\beta|\Psi, \theta, y)$. The situation is identical to Bayesian kriging, see Hjort and Omre (1994), Ch. 3.1.3, and the conditional distribution is easily seen to be multi-normal with expectation vector and covariance matrix given by

$$\mathrm{E}[\beta|\Psi, \theta, y] = \mu_0 + \Sigma_0 F^T (F\Sigma_0 F^T + \Phi(\theta) \otimes \Psi)^{-1}(y_S - F\mu_0) \tag{38}$$

and

$$\mathrm{Cov}[\beta|\Psi, \theta, y] = \Sigma_0 - \Sigma_0 F^T (F\Sigma_0 F^T + \Phi(\theta) \otimes \Psi)^{-1} F\Sigma_0, \tag{39}$$

respectively. Thus, it is natural to use a Gibbs step (Smith and Roberts, 1993) for β. It should be noted that if the product nd is large, direct inversion of the matrix $F\Sigma_0 F^T + \Phi(\theta) \otimes \Psi$ is computationally expensive and it becomes preferable to take use of the matrix identity $(F\Sigma_0 F^T + \Phi(\theta) \otimes \Psi)^{-1} = \Phi(\theta)^{-1} \otimes \Psi^{-1} - (\Phi(\theta)^{-1} \otimes \Psi^{-1})(F^T(\Phi(\theta)^{-1} \otimes \Psi^{-1})F + \Sigma_0^{-1})^{-1} F^T (\Phi(\theta)^{-1} \otimes \Psi^{-1})$.

To update Ψ, it is again most natural to use a Gibbs step as the conditional distribution of Ψ^{-1} given β, θ, y is the Wishart distribution

$$\Psi^{-1} | \beta, \theta, y \sim \mathcal{W}(\widetilde{p}, \widetilde{P}), \tag{40}$$

where $\widetilde{p} = p + n$ and

$$\widetilde{P} = \left[P^{-1} + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\Phi(\theta)^{-1}\right)_{ij} (y(u_i) - F(u_i)\beta)(y(u_j) - F(u_j)\beta)^T \right]^{-1}. \tag{41}$$

As for the updating of β above, also the proposal of Ψ is similar to the pure Gaussian case with conditioning on y.

Unlike the situation for β and Ψ above, there is no most natural proposal distribution for θ and, specifically, a Gibbs step is not computationally viable. However, θ is typically low dimensional and to propose a small change and accept with the standard Metropolis-Hastings acceptance probability should therefore give satisfactory convergence.

5.2.2 Subcompositional observations

Let the $x(u)$, β, Ψ and θ be as in the previous section, but consider now the situation where the available observations are the subcompositions $x_{S_1}(u_1), \ldots, x_{S_n}(u_n)$. It should be noted that this includes the possibility of complete observations in some locations by letting the corresponding $S_i = I_D$. Correspondingly to the situation in the previous section, conditioning to the observed subcompositions is equivalent to conditioning to $y_{S_1}(u_1), \ldots, y_{S_n}(u_n)$. Thus, introducing $y_S = [y_{S_1}(u_1)^T, \ldots, y_{S_n}(u_n)^T]^T$, the posterior of interest can be expressed as

$$\pi(\beta, \Psi, \theta | x_{S_1}(u_1), \ldots, x_{S_n}(u_n)) = \pi(\beta, \Psi, \theta | y_S) \propto \pi(\theta)\pi(\Psi)\pi(\beta) f(y_S | \beta, \Psi, \theta). \tag{42}$$

However, to simulate from this distribution by a Metropolis-Hasting algorithm similar to the one discussed in the previous section is not viable as the full conditional $\Psi^{-1} | \beta, \theta, y_S$ is not a Wishart or any other tractable distribution. To avoid this complication it is preferable in stead to use Metropolis-Hastings to sample from

$$\pi(\beta, \Psi, \theta, y | x_{S_1}(u_1), \ldots, x_{S_n}(u_n)) = \pi(\beta, \Psi, \theta, y | y_S) \tag{43}$$

and update the four groups β, Ψ, θ and y separately. To update β, Ψ and θ, the procedures discussed in the previous section is still the natural choice. To update y a Gibbs step can be used, i.e. to sample the new value for y from $\pi(y|\beta, \Psi, \theta)$, see the discussion in Section 4.1.

5.2.3 Observation of individual proportions

Let $x(u)$, β, Ψ and θ be as in the two previous sections, but consider now the situation where individual proportions as observed in some locations. More precisely, let the observed values be $S_1 x(u_1), \ldots, S_n x(u_n)$, where $S_1, \ldots, S_n$ are selection matrices. Setting $S_i = I_D$, again corresponds to complete observation in location u_i. The posterior distribution of interest becomes

$$\pi(\beta, \Psi, \theta | S_1 x(u_1), \ldots, S_n x(u_n)) \propto \pi(\beta)\pi(\Psi)\pi(\theta) f(S_1 x(u_1), \ldots, S_n x(u_n)). \tag{44}$$

Similar to the situation for subcompositional observations, there exist no natural Metropolis-Hastings algorithm for this density. Here, neither $\beta | \Psi, \theta, S_1 x(u_1), \ldots, S_n x(u_n)$ nor $\Psi | \beta, \theta$, $S_1 x(u_1), \ldots, S_n x(u_n)$ are tractable distributions. As in the previous section, this complication disappears by in stead considering

$$\pi(\beta, \Psi, \theta, y | S_1 x(u_1), \ldots, S_n x(u_n)) \tag{45}$$

and define a Metropolis-Hastings algorithm updating β, Ψ, θ and y separately. To update the three first variables, the procedures used in Sections 5.2.1 and 5.2.2 still apply. To update y the type of Metropolis-Hastings step discussed in Section 4.2 is a natural choice.

6 Example: Sediments in an Arctic lake

In this section we revisit a data set of sediments in Stanwell-Fletcher lake in the Canadian Arctic Archipelago (Coakley and Rust, 1968). The data set is previously discussed in Aitchison (1986) and Iyengar and Dey (1996), but without taking the spatial aspect of the problem into account. Our goal is to analyse also the spatial structure of the data and for this we use the fully Bayesian model discussed in Section 5. We start with a brief introduction to the data set.

The data consists of observations in 39 locations in Stanwell-Fletcher lake (see Figure 1), in each of which the water depth and the compositions of sand, silt and clay in the upper part

—

Figure 1 about here.

—

of the bottom sediments have been measured. The proportions of sand, silt and clay show a clear trend with depth, see Figure 2.

—

Figure 2 about here.

—

To analyse the data, we adopt the fully Bayesian model discussed in Section 5. We let $x(u) = (x_1(u), x_2(u), x_3(u))$ where $x_1(u)$, $x_2(u)$ and $x_3(u)$ are proportions of sand, silt and clay, respectively, at location u. We let the mean function be given by equation (31) and use water depth as covariate in $F(u)$. More precisely, we set $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ and

$$F(u) = \begin{bmatrix} 1 & 0 & \ln(d(u)) & 0 \\ 0 & 1 & 0 & \ln(d(u)) \end{bmatrix}, \tag{46}$$

where $d(u)$ denotes water depth at location u. Thus, each element in the two dimensional $y(u)$ has expectation consisting of a constant term and a term proportional to $\ln(d(u))$. As spatial correlation function, $\alpha_\theta(u, u')$, we take a generalised exponential correlation function and includes a nugget effect. Moreover, we also include the possibility for no spatial correlation. Thus, with $\theta = (R, \nu, \varepsilon, \chi)$ and

$$\alpha_\theta(u, u') = \begin{cases} 1 & \text{if } u = u', \\ (1-\varepsilon)\exp\left\{-\left(\frac{||u-u'||}{3R}\right)^\nu\right\} & \text{if } u \neq u' \text{ and } \chi = 1, \\ 0 & \text{if } u \neq u' \text{ and } \chi = 0, \end{cases} \tag{47}$$

where $||\cdot||$ denotes Euclidean distance and $R > 0$, $\nu \in [0, 2]$, $\varepsilon \in [0, 1]$ and $\chi \in \{0, 1\}$ are spatial correlation length, form parameter for correlation function, nugget effect and indicator for spatial correlation, respectively. It should be noted that with the inclusion of χ, the situation can be interpreted as a model choice problem with two competing models, one with and one without spatial correlation present, see also the discussions in Gelfand and Dey (1994) and Carlin and Chib (1995).

We assign diffuse, but proper, prior distributions to the parameters β, Ψ and θ. More precisely, we let β be Gaussian with vanishing expectation and diagonal covariance matrix with all variances equal to 100^2. The Ψ^{-1} is assigned a Wishart distribution with $p = 4$ and $P = I_2/(p-3)$, so that $\mathrm{E}[\Psi^{-1}] = I_2$. The elements of θ are assumed a priori independent,

R is assigned an exponential distribution with expectation 15, ν and ε are assigned uniform distributions on the intervals $[0, 2]$ and $[0, 1]$, respectively, and $\mathrm{P}(\chi = 0) = \mathrm{P}(\chi = 1) = 1/2$.

To sample the posterior distribution we use the MCMC algorithm discussed in Section 5.2.1, but update θ five times for each update of β and Ψ. The total of one update for each of β and Ψ and five updates for θ we call one iteration. To update β and Ψ we use Gibbs steps as discussed in Section 5.2.1. A satisfactory updating scheme for θ is found after some try and error; with probability 0.25 we propose to change the value of χ to its opposite value and keep R, ν and ε unchanged, with probability 0.75 we keep χ unchanged and draw potential new values for R, ν and ε from their respective prior distributions.

The MCMC procedure seems quickly the reach convergence and to have good mixing properties, see Figure 3 for trace plots of R, ν, β_1 and $\Psi_{1,1}$ for the first 500 iterations of a run. Trace plots for other simulated components have similar appearances. To quantify the

—

Figure 3 about here.

—

mixing properties of the simulated Markov chain, we estimate the auto-correlation functions for each of the eleven parameters simulated, four of which are shown in Figure 4. The longest auto-correlation range appear for the elements of Ψ, where a lag of about 20 iterations is necessary to obtain uncorrelated samples.

—

Figure 4 about here.

—

We estimate the posterior probability for presence of spatial correlation by the fraction of iterations where $\chi = 1$. We get $\widehat{\mathrm{P}}(\chi = 1|\text{data}) = 0.73$, which corresponds to a Bayes factor (Gelfand and Dey, 1994) of 1.46. Thus, the data favour the model with spatial correlation, but the tendency is not very strong. In Figure 5, the estimated posterior densities for each of

—

Figure 5 about here.

—

the 11 parameters, when conditioned to $\chi = 1$, is presented. One observe that the correlation length, R, with high probability is quite large, but so is the nugget effect, ε. However, in stead of considering the parameters R, ν and ε, individually, it is more informative to study the corresponding induced correlations at different lags. The last two subplots in Figure 5 give the posterior densities for the correlation at lags $1.6km$, which is the minimum distance between observations, and at lag $3.2km$. One can observe that the spatial correlations at lags $1.6km$ and larger are, with high probability, quite small. This also explains the quite small Bayes factor of 1.46 and it is reason to believe that a more dense sampling of data would have given a higher Bayes factor.

In Figure 5 one can observe that all four elements of β are significantly different from zero. To evaluate our choice of the transformation $\ln(d(u))$ in the definition of $F(u)$, we also run a simulation where $\beta = (\beta_1, \ldots, \beta_6)$ and

$$F(u) = \begin{bmatrix} 1 & 0 & \ln(d(u)) & 0 & d(u) & 0 \\ 0 & 1 & 0 & \ln(d(u)) & 0 & d(u) \end{bmatrix}. \tag{48}$$

This gave posterior distributions for β_5 and β_6 approximately centered at the origin, whereas the densities of β_1 to β_4 were essentially equal to the one in Figure 5. Thus, this gives a clear support for our choice of using a logarithmic transformation of depth.

7 Closing Remarks

The paper defines a spatial model for compositional data and evaluates the model within a Bayesian setting. Different forms of observations are considered. Complete observations in a number of locations is the simplest variant to handle, but we also define efficient MCMC algorithms to handle situations where subcompositions or individual proportions are available. In addition to the data considered in Section 6, we have used the different algorithms discussed in this paper with several simulated data sets and obtained good convergence rates.

References

Aitchison, J. (1982). The statistical analysis of compositional data, *J. Roy. Statist. Soc. Ser. B* **44**: 139–177. With discussion.

Aitchison, J. (1986). *The statistical analysis of compositional data*, Chapman & Hall, London.

Aitchison, J. and Shen, S. M. (1980). Logistic-normal distributions: some properties and uses, *Biometrika* **67**: 261–272.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**: 3–66.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W. and Silkey, M. (1997). Natural variability of benthic species in the delaware bay, *Env. Ecol. Statist.* **4**: 95–115.

Billheimer, D. and Guttorp, P. (1995). Spatial models for discrete compositional data, *No. 301*, Department of Statistics, University of Washington. Available from http://www.stat.washington.edu/.

Billheimer, D., Guttorp, P. and Fagan, W. F. (1997). Statistical analysis and interpretation of discrete compositional data, *NRCSE-TRS No. 011*, The national research center for statistics and the environment, University of Washington.

Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov-chain Monte-Carlo methods, *Journal of the Royal Statistical Society, Series B* **57**: 473–484.

Coakley, J. P. and Rust, B. R. (1968). Sedimentation in an arctic lake, *Journal of sedimentary petrology* **38**: 1290–1300.

Cressie, N. A. C. (1993). *Statistics for spatial data*, 2 edn, John Wiley, New York.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice – asymptotics and exact calculations, *Journal of the Royal Statistical Society, Series B* **56**: 501–514.

Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics (with discussion), *Scandinavian Journal of Statistics* **21**: 289–357.

Iyengar, M. and Dey, D. K. (1996). Bayesian analysis of compositional data, *No. 31*, Department of Statistics, University of Connecticut.

Omre, H. (1987). Bayesian kriging – merging observations and quallified guesses in kriging, *Mathematical Geology* **19**: 25–39.

Pawlowsky, V. and Burger, H. (1992). Spatial structure-analysis of regionalized compositions, *Mathematical Geology* **24**: 675–691.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion), *Journal of the Royal Statistical Society, Series B* **55**(1): 3–23.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *The Annals of Statistics* **22**(4): 1701–1762.

Wackernagel, H. (1995). *Multivariate geostatistics*, Springer, Berlin.

A Proof of relation (10)

Let $x \in \Re^D$, $y, z \in \Re^d$, $x_S \in \Re^C$, $y_S, z_S \in \Re^c$, where $d = D-1$ and $c = C-1$ and let $S \in \Re^{C\times D}$ be a selection matrix. We will then show that the following sequence of transformations from y to y_S,

$$z = \frac{e^y}{1 + j_d^T e^y} \quad , \quad x = B_d z + c_d \quad , \quad x_S = \frac{Sx}{j_C^T Sx} \quad , \quad z_S = A_c x_S \quad , \quad y_S = \ln\left[\frac{z_S}{1 - j_c^T z_S}\right], \tag{49}$$

gives the linear relation

$$y_S = Q_S y \qquad \text{where} \qquad Q_S = B_c^T S B_d H_d^{-1}. \tag{50}$$

The relations in (49) are one-to-one except the middle one and the corresponding four inverse transformations read

$$y = \ln\left[\frac{z}{1 - j_d^T z}\right] \quad , \quad z = A_d x \quad , \quad x_S = B_c z_s + c_c \quad , \quad z_S = \frac{e^{y_S}}{1 + j_c^T e^{y_S}}. \tag{51}$$

We first observe

$$1 - j_d^T z = 1 - j_d^T A_d x = x_D \qquad \text{and} \qquad \ln(A_d x) = A_d \ln(x), \tag{52}$$

the last because each row of A_d has all zero elements except one which is equal to unity. Thus, combining the two first relations in (51) gives

$$y = \ln\left[\frac{A_d x}{x_D}\right] = \ln(A_d x) - j_d \ln(x_D) = A_d \ln(x) - j_d \ln(x_D) = B_d^T \ln(x). \tag{53}$$

Likewise, from the two last relations in (49), one gets

$$y_S = B_c^T \ln(x_S), \tag{54}$$

from which, by inserting the definition of x_S, its follows

$$y_S = B_c^T \ln\left[\frac{Sx}{j_C^T Sx}\right] = B_c^T(\ln(Sx) - j_C \ln(j_C^T Sx)) = B_c^T S \ln(x), \tag{55}$$

by using the relations $\ln(Sx) = S\ln(x)$ and $B_c^T j_c = 0$.

Multiplying with $B_d H_d^{-1}$ from the left on each side of (53) and using the matrix identity $B_d H_d^{-1} B_d^T = I_D - (1/D) J_D$ gives

$$B_d H_d^{-1} y = B_d H^{-1} B_d^T \ln(x) = \ln(x) - \frac{1}{D} J_d \ln(x) \tag{56}$$

so that

$$\ln(x) = B_d H_d^{-1} y + \frac{1}{D} J_D \ln(x). \tag{57}$$

Thus, inserting this last expression in (55) one gets

$$y_S = B_c^T S \left(B_d^T H_d^{-1} y + \frac{1}{D} J_D \ln(x)\right) = B_c^T S B_d^T H_d^{-1} y \tag{58}$$

because $B_c^T(SJ_D) = B_c^T J_D = 0$. Thus, (50) is established.

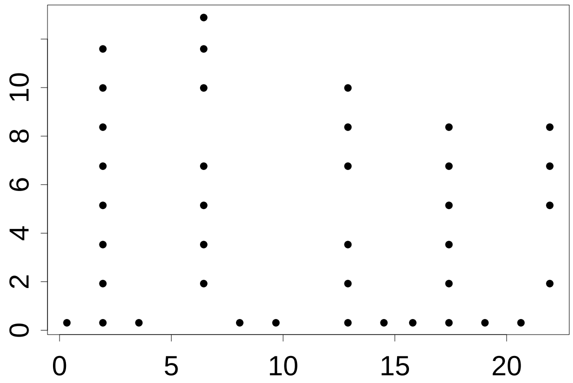

Figure 1: Locations with depth and sediment observations in the Stanwell-Fletcher lake. Distances are in kilometres.

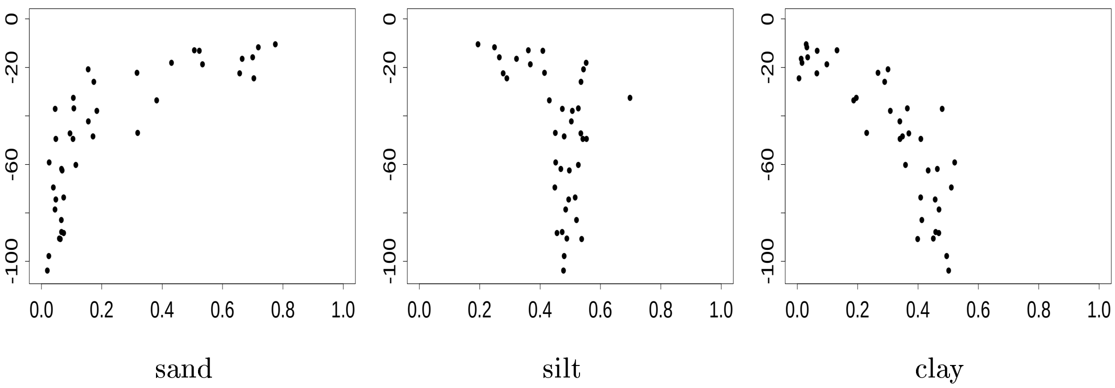

Figure 2: Proportions of sand, silt and clay as function of depth, proportions along the horizontal axis and negative depth along vertical axis.

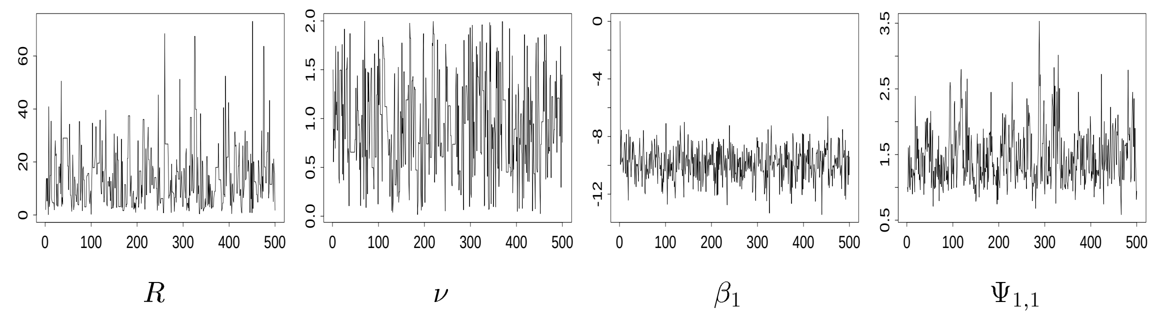

Figure 3: Trace plots of simulated values for R, ν, β_1 and $\Psi_{1,1}$. Horizontal axis is iteration number and vertical axis is parameter value.

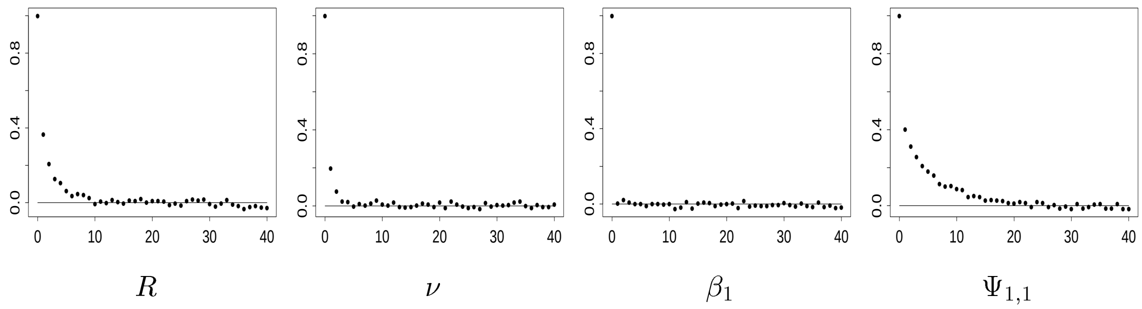

Figure 4: Estimated autocorrelation functions for simulated values of R, ν, β_1 and $\Psi_{1,1}$. Basis for the estimation is a $5,000$ iteration run, of which a 20 iteration "burn-in" period is discarded.

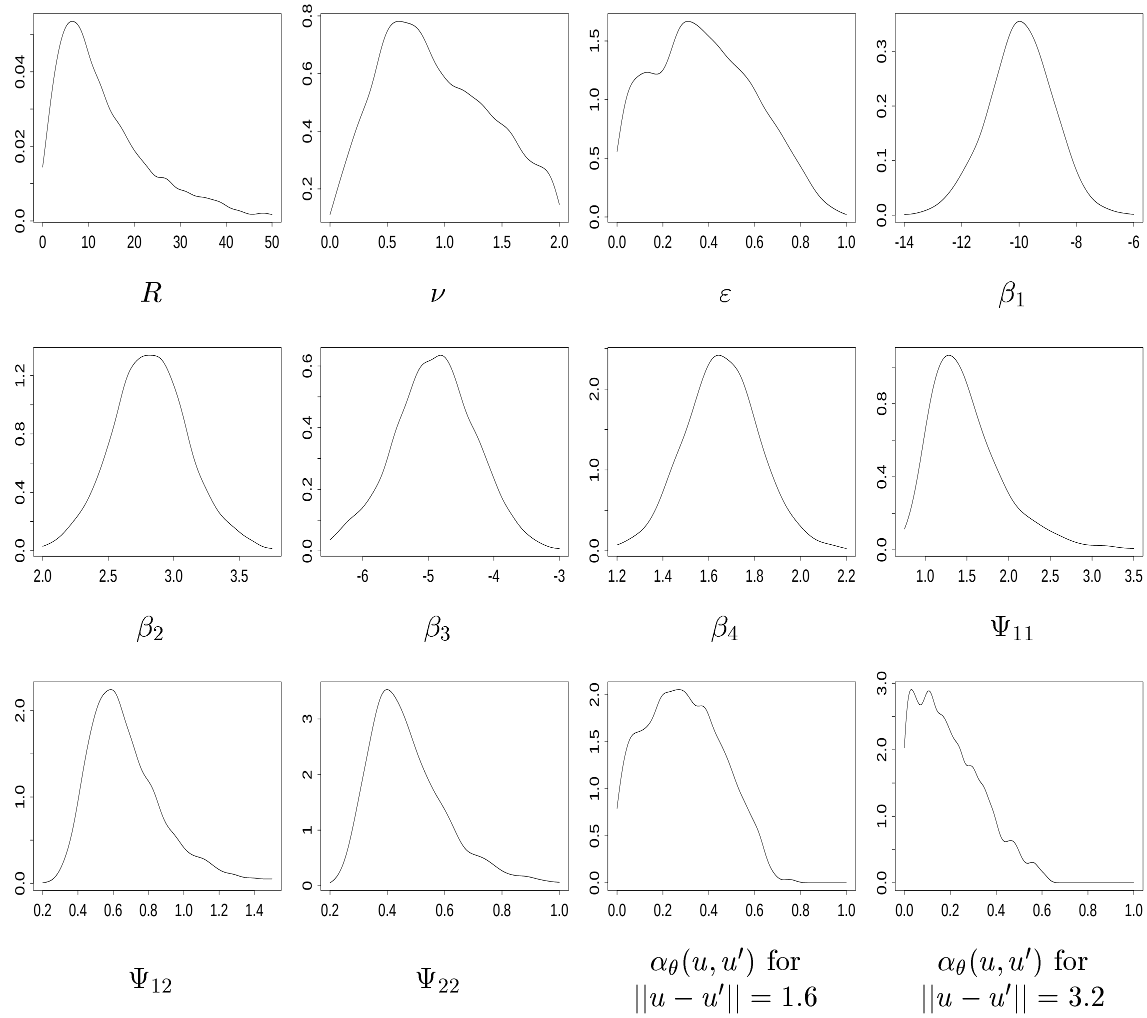

Figure 5: Marginal posterior densities for each of the 11 parameters present in the model (conditioned on $\chi = 1$) and corresponding induced posterior densities for spatial correlation at distances 1.6 (which is the minimum distance between two locations with observations) and 3.2. The densities presented are produced via kernel density estimation from output of a $5,000$ iteration Metropolis-Hastings run, of which a 20 iteration "burn-in" period is discarded.