Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

Bayesian modelling of spatial compositional data

Håkon Tjelmeland [a]; Kjetill Vassmo Lund [a]
[a] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Online Publication Date: 01 January 2003

To cite this Article Tjelmeland, Håkon and Lund, Kjetill Vassmo(2003)'Bayesian modelling of spatial compositional data',Journal of Applied Statistics,30:1,87 — 100
To link to this Article: DOI: 10.1080/0266476022000018547
URL: http://dx.doi.org/10.1080/0266476022000018547

Journal of Applied Statistics, Vol. 30, No. 1, 2003, 87–100

Bayesian modelling of spatial compositional data

HÅKON TJELMELAND & 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 generalize 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 subcompositions or individual proportions, and discuss the resulting posterior distributions. In general, the posterior cannot 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 have previously been considered, but then without taking the spatial aspect into account.*

1 Introduction

Compositional data are vectors of proportions, specifying D fractions of a whole. Thus, for $x = (x_1, \ldots, x_D)^{\mathrm{T}}$ 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 & Shen (1980) and Aitchison (1982, 1986) 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 are inherited by the logistic normal family and this allows extensive analytical treatment for logistic normal distributions.

Spatial models for proportions are also of interest in many situations. Let

Correspondence: H. Tjelmeland, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

ISSN 0266-4763 print; 1360-0532 online/03/010087-14
DOI: 10.1080/0266476022000018547

$x(u) = (x_1(u), \ldots, x_D(u))^{\mathrm{T}}$ be a stochastic vector field, where $u \in \mathscr{R}^k$ denotes 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 $u \in \mathscr{R}^k$ and $i = 1, \ldots, D$ and $x_1(u) + \ldots, x_D(u) = 1$ for all locations $u \in \mathscr{R}^k$. 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 are of interest, $x_1(u), \ldots, x_D(u)$ may represent proportions of sand, silt and clay and $u \in \mathscr{R}^2$ is geographical location.

Some spatial generalizations of the logistic normal family are discussed in the literature. In Pawlowsky & Burger (1992), spatial interpolation of compositional data is considered via logistic transformation in each location to obtain data in $\mathscr{R}^{D-1}$, followed by a co-kriging procedure. Thus, a model of correlated Gaussian fields is adopted after the transformation. In Billheimer *et al.* (1997a, b) discrete spatially correlated compositional data are modelled via underlying (unobserved) Gaussian Markov random fields. However, this formulation allows very little analytical computations and for inference Billheimer *et al.* (1997a, b) 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 of Pawlowsky & Burger (1992), but we consider it in a broader setting. First, we discuss the problem of parameter estimation by including the model in a hierarchical Bayesian framework. Second, Pawlowsky & Burger (1992) limit the attention to complete observations, i.e. all proportions are observed in a number of locations, whereas in this paper we also consider observation of subcompositions and 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 Aitchson (1986), subcompositional data fit very nicely into the logistic normal family and allow extensive analytical treatment. Observation of individual proportions is more problematic and MCMC is necessary to handle this case. However, the Gaussian base can be used to define efficient algorithms, simultaneously updating large blocks of variables.

The paper is organized as follows. In the next section we introduce the basic notation and discuss appropriate sample spaces. In Section 3, logistic Gaussian fields are defined as a generalization 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 discussed. In Section 6, we revisit a data set of sediments in an Arctic lake (Coakley & Rust, 1968). The data set has been previously analysed (Aitchison, 1986; Iyengar & Dey, 1996), but the spatial aspect of the problem was then ignored. 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 $x(u) = (x_1(u), \ldots, x_D(u))^{\mathrm{T}}; u \in \mathscr{R}^k$ be a compositional field, i.e. $x_i(u)$ gives the proportion of substance i in location u. Thus, the sample space of $x(u)$ is

$$\mathscr{L}^D = \left\{ x(u) = (x_1(u), \ldots, x_D(u))^{\mathrm{T}} \,\middle|\, x_i(u) > 0, i = 1, \ldots, D \quad \text{and} \quad \sum_{i=1}^{D} x_i(u) = 1 \right\} \subset \mathscr{R}^D \tag{1}$$

The unit sum constraint implies that any $d = D - 1$ dimensional sub-vector uniquely specifies the last component and the probability density for $x(u)$ is thereby degenerate. We therefore 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

$$z(u) = A_d x(u) \quad \Leftrightarrow \quad x(u) = B_d z(u) + c_d \tag{2}$$

where $A_d \in \mathscr{R}^{d \times D}$ consists of a $d \times d$ identity matrix with an extra column of zeros, $B_d \in \mathscr{R}^{D \times d}$ is a $d \times d$ identity matrix with an extra row of -1s added, and $c_d = (0, 0, \ldots, 0, 1)^{\mathrm{T}} \in \mathscr{R}^D$. The sample space of $z(u)$ becomes

$$\mathscr{S}^d = \left\{ z(u) = (z_1(u), \ldots, z_d(u))^{\mathrm{T}} \,|\, z_i(u) > 0, i = 1, \ldots, d \quad \text{and} \quad \sum_{i=1}^{d} z_i(u) < 1 \right\} \tag{3}$$

The logistic transformation defines a one-to-one relation between $\mathscr{S}^d$ and $\mathscr{R}^d$. With $y(u) = (y_1(u), \ldots, y_d(u))^{\mathrm{T}}$, the logistic transformation is in vector format given as

$$z(u) = \frac{\mathrm{e}^{y(u)}}{1 + j_d^{\mathrm{T}} \mathrm{e}^{y(u)}} \quad \Leftrightarrow \quad y(u) = \ln\left[\frac{z(u)}{1 - j_d^{\mathrm{T}} z(u)} \right] \tag{4}$$

where $j_d \in \mathscr{R}^d$ is a column 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 $\mathscr{L}^D$, $\mathscr{R}^d$ and $\mathscr{S}^d$, respectively, and let their relation be given by equations (2) and (4). Thus, whenever the distribution of one of the fields is specified, the distributions of the other two are also given.

Subcompositions and individual proportions can be obtained from $x(u)$ via selection matrices. A matrix $S \in \mathscr{R}^{C \times D}$, where $C \leqslant D$, 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 $Sx(u)$. A rescaling to obtain unit sum gives the corresponding subcomposition, $x_S(u)$, i.e.

$$x_S(u) = \frac{Sx(u)}{j_C^{\mathrm{T}} Sx(u)} \tag{5}$$

Note that $x_S(u) \in \mathscr{L}^C$, so $x_S(u)$ is a C-dimensional compositional field. Thus, corresponding quantities $z_S(u) \in \mathscr{S}^c$ and $y_S(u) \in \mathscr{R}^c$ where $c = C - 1$ can be defined by equations (2) and (4) 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,

$$y_S(u) = Q_S y(u) \quad \text{where} \quad Q_S = B_c^{\mathrm{T}} S B_d H_d^{-1} \in \mathscr{R}^{c \times d}, H_d = I_d + \mathfrak{J}_d \tag{6}$$

where I_d is the $d \times d$ identity matrix and $\mathcal{J}_d$ 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 the definition of logistic normal distributions. Likewise, this property is essential for the tractability of logistic Gaussian fields discussed in this paper. Since equation (6) is so fundamental for the following sections, and as Aitchison (1986) contains no clearly written proof of the result, we include one in the Appendix.

A particular subclass of selection matrices is permutation matrices, for which $C = D$. The denominator in equation (5) is then unity and the elements of $x_S(u)$ are just a permuted version of the elements in $x(u)$. We denote a permutation matrix 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 $P = I_D$.

3 Logistic Gaussian fields

Corresponding to how Aitchison (1986) defines logistic normal distributions from multi-normal distributions, we here define logistic Gaussian fields from a multivariate Gaussian process.

Definition 1:
Let $y(u) = (y_1(u), \ldots, y_d(u))^{\mathrm{T}}$; $u \in \mathcal{R}^k$ be a multivariate Gaussian process with mean function $\mu(u) = E[y(u)] = (\mu_1(u), \ldots, \mu_d(u))^{\mathrm{T}}$ and covariance structure given by $c(u,u') = \mathrm{Cov}[y(u), y(u')] = (c_{ij}(u, u'))_{i,j=1}^{d}$. Furthermore, for each $u \in \mathcal{R}^k$, let $z(u)$ and $x(u)$ be defined from $y(u)$ by equations (4) and (2), respectively. We then call $x(u) = (x_1(u), \ldots, x_D(u))^{\mathrm{T}}$; $u \in \mathcal{R}^k$, or equivalently $z(u) = (z_1(u), \ldots, z_d(u))^{\mathrm{T}}$; $u \in \mathcal{R}^k$, a logistic Gaussian field with parameter functions $\mu(\cdot)$ and $c(\cdot, \cdot)$.

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 $\{v(u) = (v_1(u), \ldots, v_D(u))^{\mathrm{T}}; u \in \mathcal{R}^k\}$ be a multivariate Gaussian process with mean function $\mu_0(u) \in \mathcal{R}^D$ and correlation structure $c_0(u, u') \in \mathcal{R}^{D \times D}$, and set $x(u) = \mathrm{e}^{v(u)} / (j_D^{\mathrm{T}} \mathrm{e}^{v(u)})$ for each $u \in \mathcal{R}^k$. Then $x(u)$ is a logistic Gaussian field with parameter functions $\mu(u) = B_d^{\mathrm{T}} \mu_0(u)$ and $c(u, u') = B_d^{\mathrm{T}} c_0(u, u') B_d$. This result is immediate by transforming $x(u)$ via $z(u)$ to the corresponding $y(u)$ by equations (2) and (4) 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 $\{x(u); u \in \mathcal{R}^k\} \stackrel{d}{=} \{Px(u); u \in \mathcal{R}^k\}$ for any permutation matrix P. This is fulfilled if and only if $\mu(u) = 0$ for all $u \in \mathcal{R}^k$ and $c(u,u')$ has the form $c(u,u') = \sigma^2 \alpha(u, u')\, H_d/2$, 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 equation (6). It should be noted that this $c(\cdot, \cdot)$ falls within the class of intrinsic correlations structures (see, for example, Wackernagel, 1995, Ch. 22), which is reasonable in some applications and which has important computational advantages.

Unconditional simulation of logistic Gaussian fields in a set of grid locations is straightforward from its definition; first simulate the multivariate Gaussian process

$y(u)$ in the given array and then use logistic transformation (4) for each grid location. To simulate a Gaussian process many different methods exist, see Cressie (1993) for an overview. Simulation of the Gaussian process becomes especially simple with an intrinsic covariance structure.

4 Conditional logistic Gaussian fields

Properties of logistic Gaussian fields conditioned to data depend on the type of observations available. In the following, we first discuss complete and subcompositional data and thereafter observations 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 $S_i \in \mathscr{R}^{C_i \times D}$ and a corresponding observation vector $x_{S_i}(u_i)$. Complete observations in location u_i correspond to setting S_i equal to the identity matrix (or any 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 $y_{S_1}(u_1),\ldots,y_{S_n}(u_n)$. The linear relation in equation (6) gives that $y(u)\,|\,y_{S_1}(u_1),\ldots,y_{S_n}(u_n)$ is a multivariate Gaussian process and, thereby, by definition of logistic Gaussian fields, that $x(u)\,|\,x_{S_1}(u_1),\ldots,x_{S_n}(u_n)$ is also a logistic Gaussian field. The parameter functions of the conditional field are also readily available as the mean and covariance functions of the corresponding Gaussian y-process. 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 equation (4) for each grid location. A conditional realisation 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, $u_1,\ldots,u_n$. Thus, to each u_i there is again associated a selection matrix, $S_i \in \mathscr{R}^{C_i \times D}$, but now the observed values are the vectors $S_1x(u_1),\ldots,S_nx(u_n)$. 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 realizations of $x(u)$ by Markov chain Monte Carlo (MCMC) procedures. It should be noted that it is sufficient to use MCMC to generate samples of $x(u_1),\ldots,x(u_n)\,|\,S_1x(u_1),\ldots,S_nx(u_n)$. If samples of the logistic field $x(u)$ for locations $u \notin \{u_1,\ldots,u_n\}$ are of interest, these can thereafter be generated directly by conditioning to simulated values for $x(u_1),\ldots,x(u_n)$ by the procedure discussed in Section 4.1.

As some components of $x(u_1),\ldots,x(u_n)$ are identical to the elements of

$S_1x(u_1), \ldots, S_nx(u_n)$, sampling $x(u_1), \ldots, x(u_n) \,|\, S_1x(u_1), \ldots, S_nx(u_n)$ actually means generating samples of the elements in $x(u_1), \ldots, x(u_n)$ that are not specified by $S_1x(u_1), \ldots, S_nx(u_n)$. Therefore, let T_i be a selection matrix that picks out the non-observed elements of $x(u_i)$, so that

$$P_i = \begin{bmatrix} S_i \\ T_i \end{bmatrix} \tag{7}$$

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

$$x(u_i) = P_i^{\mathrm{T}} \begin{bmatrix} S_ix(u_i) \\ T_ix(u_i) \end{bmatrix} \tag{8}$$

Moreover, from the unit sum constraint in each location, it follows that the last component in $T_ix(u_i)$ is given by the first $D - C_i - 1$ elements of $T_ix(u_i)$ together with $S_ix(u_i)$. Thus, the variables to be simulated are $\tilde{z}_1 = A_{b_1} T_1 x(u_1), \ldots, \tilde{z}_n = A_{b_n} T_n x(u_n)$, where $b_i = D - C_i - 1$. The conditional density of interest is

$$f(\tilde{z}_1, \ldots, \tilde{z}_n \,|\, S_1x(u_1) = \tilde{x}_1, \ldots, S_nx(u_n) = \tilde{x}_n) \propto f(z_1, \ldots, z_n) \tag{9}$$

where

$$z_i = (z_{i1}, \ldots, z_{id})^{\mathrm{T}} = A_d P_i^{\mathrm{T}} \left(B_d \begin{bmatrix} \tilde{x}_i \\ \tilde{z}_i \end{bmatrix} + c_d \right) \quad \text{for} \quad i = 1, \ldots, n \tag{10}$$

and $f(z_1, \ldots, z_n)$ denotes the unconditional distribution of $z(u_1), \ldots, z(u_n)$. Using the transformation formula, an expression for $f(z_1, \ldots, z_n)$ is readily available from the corresponding joint Gaussian distribution for $y(u_1), \ldots, y(u_n)$.

Direct simulation from equation (9) is not feasible and one has to resort to MCMC and the Metropolis–Hastings algorithm. For a general introduction to this topic, see Besag *et al.* (1995) and references therein. Our target density $f(\tilde{z}_1, \ldots, \tilde{z}_n \,|\, S_1x(u_1) = \tilde{x}_1, \ldots, S_nx(u_n) = \tilde{x}_n)$ 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 suggests construction of an independent proposal (Tierney, 1994) algorithm. To see how to do this, first observe that to propose new values for $\tilde{z}_1, \ldots, \tilde{z}_n$ is equivalent to proposing values for $z_{T_1}(u_1), \ldots, z_{T_n}(u_n)$ or for $y_{T_1}(u_1), \ldots, y_{T_n}(u_n)$, where the one-to-one relations between the first two are

$$\tilde{z}_i = (1 - j_{C_i}^{\mathrm{T}} \tilde{x}_i) z_{T_i}(u_i) \quad \Leftrightarrow \quad z_{T_i}(u_i) = \frac{\tilde{z}_i}{1 - j_{C_i}^{\mathrm{T}} \tilde{x}_i} \tag{11}$$

As an approximation to sampling $y_{T_1}(u_1), \ldots, y_{T_n}(u_n)$ conditioned to the observed individual proportions, one can then sample the $y_{T_i}(u_i)$s conditioned to the corresponding subcompositions $y_{S_1}(u_1), \ldots, y_{S_n}(u_n)$. 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 $\tilde{z}_1', \ldots, \tilde{z}_n'$ one first samples $y_{T_1}(u_1)', \ldots, y_{T_n}(u_n)'$ from the Gaussian distribution $f(y_{T_1}(u_1), \ldots, y_{T_n}(u_n) \,|\, y_{S_1}(u_1), \ldots, y_{S_n}(u_n))$ and thereafter one computes corresponding transformed values $z_{T_1}(u_i)'$ and $\tilde{z}_i'; i = 1, \ldots, n$ by equations (4) and (11). For mathematical details, see Tjelmeland & Lund (2001).

The main motivation for the above independent-proposal Metropolis–Hastings algorithm is to approximate $f(y_{T_1}(u_1), \ldots, y_{T_n}(u_n) \mid S_1 x(u_1), \ldots, S_n x(u_n))$ with $f(y_{T_1}(u_1), \ldots, y_{T_n}(u_n) \mid x_{S_1}(u_1), \ldots, x_{S_n}(u_n))$. It is clear that this approximation deteriorates as n increases, with lower acceptance probabilities and slower convergence as results. For n large, a better strategy is therefore to generate potential new values as a linear combination of the current state vector and the values used above. Mathematical details can again be found in Tjelmeland & Lund (2001).

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. Instead, these functions 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 that 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 that are both flexible and give a posterior that 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 & Omre, 1994) framework and assume the following regression form,

$$\mu_\beta(u) = F(u)\,\beta \tag{12}$$

where $\beta \in \mathscr{R}^p$ is a vector of unknown parameters and $F(u) \in \mathscr{R}^{d \times p}$ 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 & Guttorp (1995). Still following the choice in Bayesian kriging, we assume a Gaussian prior distribution for the vector β,

$$\beta \sim \mathrm{N}_p(\mu_0, \Sigma_0) \tag{13}$$

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.

$$c_{\theta,\Psi}(u,u') = \alpha_\theta(u,u')\,\Psi \tag{14}$$

Different choices for the spatial correlation function exist—the generalized exponential and the Matern correlation functions are two possibilities (Cressie, 1993). However, as computational considerations do not favour any 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, $\Psi^{-1} \sim \mathscr{W}_d(q, Q)$, where the scalar $q > d$ and the positive definite $d \times d$ matrix Q are hyper-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 $\pi(\theta)$.

5.3 Simulation algorithms

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 consider each situation in turn.

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 $y=[y(u_1)^{\mathrm{T}}\ldots\ y(u_n)^{\mathrm{T}}]^{\mathrm{T}}$, the posterior distribution of interest can be expressed as

$$\pi(\beta,\Psi,\theta\,|\,x(u_1),\ldots,x(u_n))=\pi(\beta,\Psi,\theta\,|\,y)\propto\pi(\theta)\,\pi(\Psi)\,\pi(\beta)f(y\,|\,\beta,\Psi,\theta) \quad (15)$$

where $f(y\,|\,\beta,\Psi,\theta)$ is a multi-normal density. To construct an MCMC algorithm it is natural to update the three β, Ψ and θ separately.

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 & Omre (1994, Ch. 3.1.3), and the conditional distribution is easily seen to be multi-normal. Thus, it is natural to use a Gibbs step (Smith & Roberts, 1993) for β. 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 again a Wishart distribution. As for the updating of β, the proposal of Ψ is also similar to the pure Gaussian case with conditioning on y. Unlike the situation for β and Ψ, 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.

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 $x_{S_1}(u_1),\ldots,x_{S_n}(u_n)$. 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 $y_S=[y_{S_1}(u_1)^{\mathrm{T}},\ldots,y_{S_n}(u_n)^{\mathrm{T}}]^{\mathrm{T}}$, the posterior of interest can be expressed as

$$\pi(\beta,\Psi,\theta\,|\,x_{S_1}(u_1),\ldots,x_{S_n}(u_n))=\pi(\beta,\Psi,\theta\,|\,y_S)\propto\pi(\theta)\,\pi(\Psi)\,\pi(\beta)f(y_S\,|\,\beta,\Psi,\theta) \quad (16)$$

However, simulating 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 to include y as a latent variable, i.e. to sample from $\pi(\beta,\Psi,\theta,y\,|\,x_{S_1}(u_1),\ldots,x_{S_n}(u_n))=\pi(\beta,\Psi,\theta,y\,|\,y_S)$ and update the four groups β, Ψ, θ and y separately. To update β, Ψ and θ, the procedures discussed in the previous section are still natural choices, whereas for y a Gibbs step is applicable via the procedure defined in Section 4.1.

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 $S_1x(u_1), \ldots, S_nx(u_n)$, where $S_1, \ldots, S_n$ are selection matrices. The posterior distribution of interest is

$$\pi(\beta, \Psi, \theta \mid S_1x(u_1), \ldots, S_nx(u_n)) \propto \pi(\beta)\pi(\Psi)\pi(\theta)f(S_1x(u_1), \ldots, S_nx(u_n)) \quad (17)$$

Similar to the situation for subcompositional observations, there exist no natural Metropolis–Hastings algorithms for this density. Here, neither $\beta \mid \Psi, \theta, S_1x(u_1), \ldots, S_nx(u_n)$ nor $\Psi \mid \beta, \theta, S_1x(u_1), \ldots, S_nx(u_n)$ are tractable distributions. Again this complication disappears by introducing y as a latent variable. Thus, one should sample from $\pi(\beta, \Psi, \theta, y \mid S_1x(u_1), \ldots, S_nx(u_n))$. To update β, Ψ and θ, the procedures discussed above still apply and to update y, the Metropolis–Hastings step discussed in Section 4.2 is the most 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 & Rust, 1968). The data set was previously discussed in Aitchison (1986) and Iyengar & Dey (1996), but without taking the spatial aspect of the problem into account. Our goal is to analyse the spatial structure of the data and for this we use the fully Bayesian model defined in Section 5. We start with a brief introduction to the data set.

The data consist of observations in $n = 39$ locations in Stanwell–Fletcher lake (see Fig. 1), in each of which the water depth and the compositions of sand, silt and clay in the upper part of the bottom sediments have been measured. The proportions of sand, silt and clay show a clear trend with depth, see Fig. 2.

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 (12) and use water depth as covariate in $F(u)$. More precisely, we set $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^{\mathrm{T}}$ and

$$F(u) = \begin{bmatrix} 1 & 0 & \ln(d(u)) & 0 \\ 0 & 1 & 0 & \ln(d(u)) \end{bmatrix} \quad (18)$$

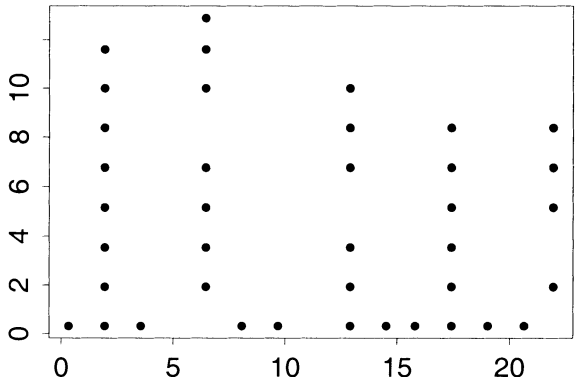

FIG. 1. Locations with depth and sediment observations in the Stanwell–Fletcher lake. Distances are in kilometres.

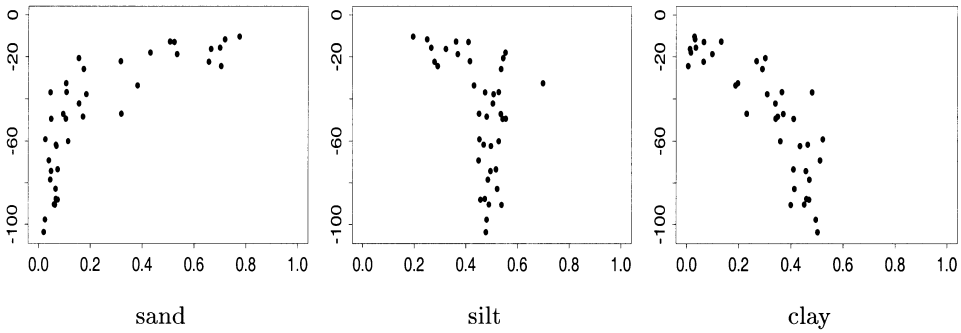

FIG. 2. Proportions of sand, silt and clay as function of depth, proportions along the horizontal axis and negative depth along vertical axis.

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

$$\alpha_\theta(u,u') = \begin{cases} 1 & \text{if} \quad u = u' \\ (1-\varepsilon)\exp\left\{-\left(\dfrac{\|u-u'\|}{3R}\right)^{\nu}\right\} & \text{if} \quad u \neq u' \text{ and } \chi = 1 \\ 0 & \text{if} \quad u \neq u' \text{ and } \chi = 0 \end{cases} \tag{19}$$

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 & Dey (1994) and Carlin & 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 $q = 4$ and $Q = I_2/(p-3)$, 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. After some experimentation with the algorithm we found the elements in θ to have the longest autocorrelation range and this motivated us to 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. A satisfactory updating scheme for θ is found by trial and error; with probability 0.25 we propose changing the value of χ to its opposite value and keeping R, ν and ε unchanged, and with the remaining probability we keep χ unchanged and draw potential new values for R, ν and ε

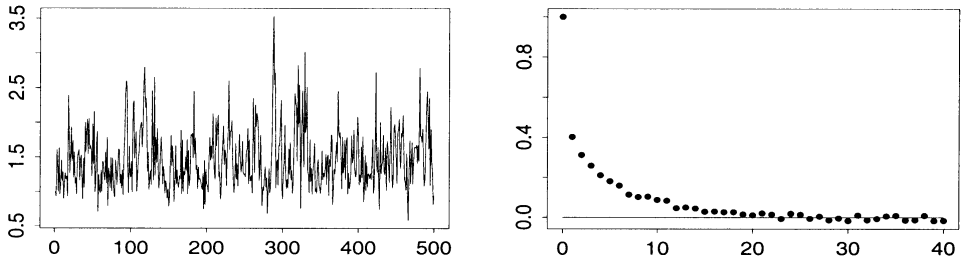

FIG. 3. Trace plot (500 iterations) and estimated autocorrelation function for element (1,1) in Ψ. The basis for the auto-correlation estimation is a 5000 iteration run, of which a 20 iteration 'burn-in' period is discarded.

from their respective prior distributions. We do not claim this algorithm to be optimal in any sense, but we found it to work satisfactory for this data set.

The MCMC procedure seems to reach convergence quickly and to have good mixing properties. Of the simulated variables, element (1,1) of Ψ has the longest auto-correlation range, see Fig. 3. We estimate the posterior probability for the presence of spatial correlation by the fraction of iterations where $\chi = 1$. We get $\hat{P}(\chi = 1 \mid \text{data}) = 0.73$, which corresponds to a Bayes factor (Gelfand & Dey, 1994) of $0.73/0.27 = 2.7$. Thus, the data favour the model with spatial correlation. In Fig. 4, the estimated posterior densities for each of the 11 parameters, when conditioned to $\chi = 1$, are presented. One observes that the correlation length, R, with high probability, is quite large, but there is also a significant nugget effect present. Instead 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 sub-plots in Fig. 4 therefore give the posterior densities for the spatial correlations at lags 1.6 km (which is the minimum distance between observations) and 3.2 km. One can observe that the spatial correlations at lag 3.2 km is, with high probability, rather small.

In Fig. 4 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

$$F(u) = \begin{bmatrix} 1 & 0 & \ln(d(u)) & 0 & d(u) & 0 \\ 0 & 1 & 0 & \ln(d(u)) & 0 & d(u) \end{bmatrix} \tag{20}$$

This gave posterior distributions for β_5 and β_6 approximately centred at the origin, whereas the densities of β_1 to β_4 were essentially equal to the ones in Fig. 4. Thus, this gives clear support for our choice of using a logarithmic transformation for the 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

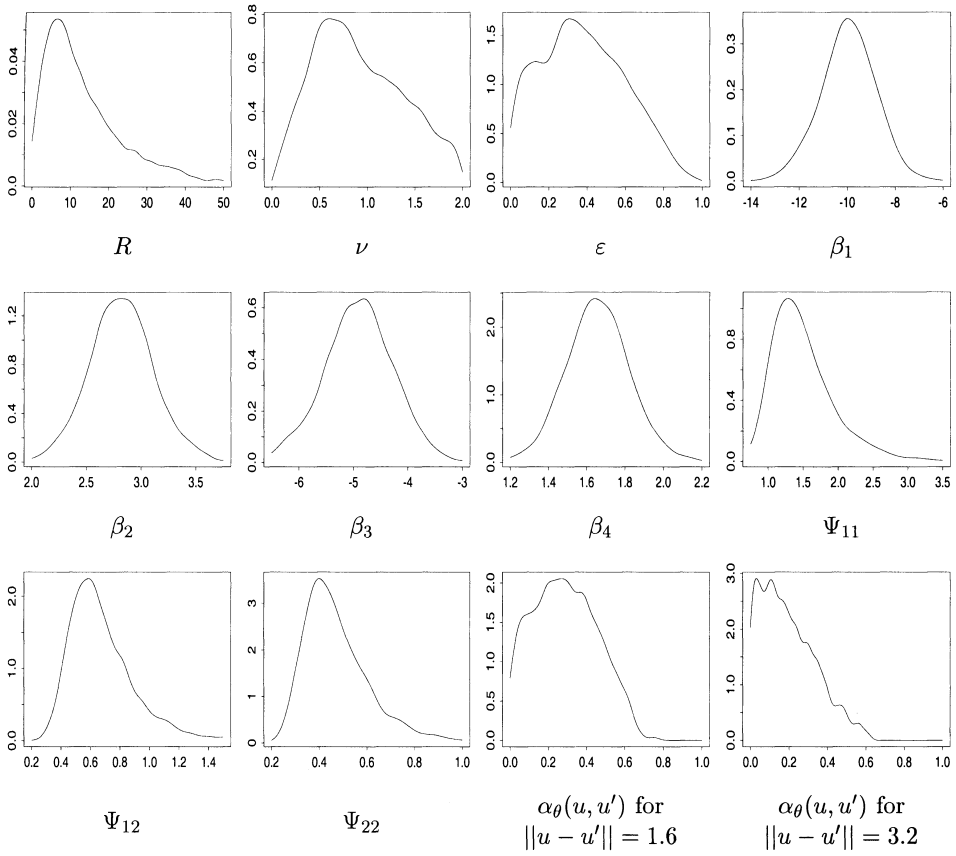

FIG. 4. Marginal posterior densities for each of the 11 parameters present in the model (conditioned on $\chi = 1$) and corresponding induced posterior densities for spatial correlation at distances 1.6 (which is the minimum distance between two locations with observations) and 3.2. The densities presented are produced via kernal density estimation from the output of a 5000 iteration Metropolis–Hastings run, of which a 20 iteration 'burn-in' period is discarded.

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.

In Section 5.1, we assumed an intrinsic covariance structure for the underlying Gaussian process. The motivations for this are twofold. First, it allows a parsimonious parameterization of the covariance function. A more general form for the covariance structure would typically include more parameters and is therefore a viable alternative only if enough data are present or if sufficient prior information is available. The second, and perhaps most important, motivation for the intrinsic structure is computational efficiency. For the Metropolis–Hastings algorithms discussed in Section 5, a general covariance structure would, in each iteration, require matrix operations on $nd \times nd$ matrices, whereas, with an intrinsic structure, operations on $n \times n$ matrices are sufficient. At least with $n = 39$, as in our example,

the necessary inversion and decomposition of $n \times n$ matrices is computationally unproblematic. However, even an $n \times n$ matrix can cause problems when the number of observations, n, is large. This is, of course, a general problem for Gaussian-based models and a full discussion of this topic is beyond the scope of this paper. However, a possible solution can be found in the use of Gaussian Markov random fields. The spatial covariance structure is then specified via a sparse inverse covariance matrix and special algorithms for sparse matrices can be used, see Rue (2001) and references therein. See also Rue & Tjelmeland (2002) for a procedure for fitting the parameters of a Gaussian Markov random field to a specified covariance function.

References

Aitchison, J. (1982) The statistical analysis of compositional data, *Journal of the Royal Statistical Society Series B*, 44, pp. 139–177, with discussion.

Aitchison, J. (1986) *The Statistical Analysis of Compositional Data* (London, Chapman & Hall).

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

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

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W. & Silkey, M. (1997a) Natural variability of benthic species in the Delaware Bay, *Journal of Environmental and Ecological Statistics*, 4, pp. 95–115.

Billheimer, D. & 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. & Fagan, W. F. (1997b) 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. & Chib, S. (1995) Bayesian model choice via Markov-chain Monte-Carlo methods, *Journal of the Royal Statistical Society, Series B*, 57, pp. 473–484.

Coakley, J. P. & Rust, B. R. (1968) Sedimentation in an Arctic lake, *Journal of Sedimentary Petrology*, 38, pp. 1290–1300.

Cressie, N. A. C. (1993) *Statistics for Spatial Data*, 2nd edn (New York, Wiley).

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

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

Ivengar, M. & Dey, D. K. (1996) Bayesian analysis of compositional data. *No. 31*, Department of Statistics, University of Connecticutt.

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

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

Rue, H. (2001) Fast sampling of Gaussian Markov random fields, *Journal of the Royal Statistical Society, Series B*, 63(2).

Rue, H. & Tjelmeland, H. (2002) Fitting Gaussian Markov random fields to Gaussian fields, *Scandinavian Journal of Statistics*, 29, pp. 31–49.

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

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

Tjelmeland, H. & Lund, K. V. (2001) Bayesian modelling of spatial compositional data. *Statistics no. 1*, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway. Available from http://www.math.ntnu.no/preprint/.

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

Appendix

Proof of relation (6)

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

$$z = \frac{e^y}{1 + j_d^{\mathrm{T}} \mathrm{e}^y}, x = B_d z + c_d, x_S = \frac{Sx}{j_C^{\mathrm{T}} Sx}, z_S = A_c x_S, y_S = \ln\left[\frac{z_S}{1 - j_c^{\mathrm{T}} z_S}\right] \tag{21}$$

gives the linear relation

$$y_S = Q_S y \quad \text{where} \quad Q_S = B_c^{\mathrm{T}} S B_d H_d^{-1} \tag{22}$$

The relations in equation (21) are one-to-one except the middle one and the corresponding four inverse transformations read

$$y = \ln\left[\frac{z}{1 - j_d^{\mathrm{T}} z}\right], z = A_d x, x_S = B_c z_s + c_c, z_S = \frac{\mathrm{e}^{y_S}}{1 + j_c^{\mathrm{T}} e^{y_S}} \tag{23}$$

We first observe

$$1 - j_d^{\mathrm{T}} z = 1 - j_d^{\mathrm{T}} A_d x = x_D \quad \text{and} \quad \ln(A_d x) = A_d \ln(x) \tag{24}$$

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 equation (23) gives

$$y = \ln\left[\frac{A_d x}{x_D}\right] = \ln(A_d x) - j_d \ln(x_D) = A_d \ln(x) - j_d \ln(x_D) = B_d^{\mathrm{T}} \ln(x) \tag{25}$$

Likewise, from the two last relations in equation (21), one gets

$$y_S = B_c^{\mathrm{T}} \ln(x_S) \tag{26}$$

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

$$y_S = B_c^{\mathrm{T}} \ln\left[\frac{Sx}{j_C^{\mathrm{T}} Sx}\right] = B_c^{\mathrm{T}} (\ln(Sx) - j_C \ln(j_C^{\mathrm{T}} Sx)) = B_c^{\mathrm{T}} S \ln(x) \tag{27}$$

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

Multiplying with $B_d H_d^{-1}$ from the left on each side of equation (25) and using the matrix identity $B_d H_d^{-1} B_d^{\mathrm{T}} = I_D - (1/D) \mathcal{J}_D$ gives

$$B_d H_d^{-1} y = B_d H^{-1} B_d^{\mathrm{T}} \ln(x) = \ln(x) - \frac{1}{D} \mathcal{J}_d \ln(x) \tag{28}$$

so that

$$\ln(x) = B_d H_d^{-1} y + \frac{1}{D} \mathcal{J}_D \ln(x) \tag{29}$$

Thus, inserting this last expression in equation (27) one gets

$$y_S = B_c^{\mathrm{T}} S \left(B_d^{\mathrm{T}} H_d^{-1} y + \frac{1}{D} \mathcal{J}_D \ln(x) \right) = B_c^{\mathrm{T}} S B_d^{\mathrm{T}} H_d^{-1} y \tag{30}$$

because $B_c^{\mathrm{T}} (S \mathcal{J}_D) = B_c^{\mathrm{T}} \mathcal{J}_D = 0$. Thus, equation (22) is established.