A Bayesian coregionalization approach for multivariate pollutant data

Alexandra M. Schmidt

Department of Statistical Methods, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Alan E. Gelfand

Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, USA

[1]   Spatial data collection increasingly turns to vector valued measurements at spatial locations. An example is the observation of pollutant measurements. Typically, several different pollutants are observed at the same sampled location, referred to as a monitoring station or gauged site. Usually, interest lies in the modeling of the joint process for the levels of the different pollutants and in the prediction of pollutant levels at ungauged sites. In this case, it is important to take into account not only the spatial correlation but also the correlation among the different variables at each gauged site. Since, conceptually, there is a potentially observable measurement vector at every location in the study region, a multivariate spatial process becomes a natural modeling choice. In using a Gaussian process, the main challenge is the specification of a valid and flexible cross-covariance function. This paper proposes a rich class of covariance functions developed through the so-called linear coregionalization model [see, e.g., *Wackernagel*, 1998] for multivariate spatial observations. Following the ideas in the work of, for example, *Royle and Berliner* [1999], we can reparameterize a multivariate spatial model using suitable univariate conditional spatial processes, facilitating the computation. We provide explicit details, including the computation of the range associated with the different component processes. As an example, we fit our model to a particular day average of CO, NO, and NO_2 for a set of monitoring stations in California, USA.   *INDEX TERMS:* 0345 Atmospheric Composition and Structure: Pollution—urban and regional (0305); 1610 Global Change: Atmosphere (0315, 0325); 9810 General or Miscellaneous: New fields not (classifiable under other headings); 9820 General or Miscellaneous: Techniques applicable in three or more fields; *KEYWORDS:* coregionalization model, Gaussian processes, hierarchical spatial models, multivariate spatial processes, spatial range

Citation: Schmidt, A. M., and A. E. Gelfand, A Bayesian coregionalization approach for multivariate pollutant data, *J. Geophys. Res.*, *108*(D24), 8783, doi:10.1029/2002JD002905, 2003.

1.   Introduction

[2]   Multivariate spatial problems are frequently encountered in environmental applications where we usually have a set of, say, n monitoring stations or gauged sites, and at each of those sites we have measurements of p pollutants. Therefore, for a given time period, concatenating the measurements into a single column vector, the observations comprise a $np \times 1$ vector. (In fact, in practice, often not all pollutants are observed at all locations, creating misalignment or missing data issues.) Here we will concentrate on the spatial aspect of these observations. Usually, on the basis of the information we have from the gauged sites, the interest lies in predicting the p different processes at locations where they are not measured.

[3]   Multivariate spatial modeling can be implemented in various ways. For instance, one could employ a Markov random field approach, such as applied by *Mardia* [1988] and, more recently, by *Gelfand and Vounatsou* [2003, and references therein]. This approach is most commonly used for areal data and for regular grid or lattice data. It results in a joint multivariate distribution for the data, which is typically assumed to be normal.

[4]   A different formulation is to envision the data as arising from a multivariate spatial process. This approach presumes that a realization of the process consists of p dependent surfaces over the study region. We only observe the values of the surface at a set of n locations. Again, a multivariate distribution results for the observed data. However, specification of a valid multivariate process is more demanding in that it is determined through its associated finite dimensional distributions. Consistency of such determinations with regard to the entire uncountable dimensional joint distribution is a nontrivial matter. However, with Gaussian processes, consistency only requires the use of a valid cross-covariance function [see, e.g., *Wackernagel*, 1998, and references therein]. With interest in prediction

at arbitrary new locations, multivariate processes seem to be the more attractive distributional framework.

[5] In what follows, we work with a multivariate Gaussian process employing a stationary cross-covariance specification. In a series of papers by N. Le, J. Zidek, and colleagues [see, e.g., *Brown et al.*, 1994], this restriction is removed by viewing the joint covariance matrix for the data as varying through an inverse Wishart distribution centered at the covariance matrix which will arise using a stationary cross-covariance function. While the restriction is removed, so is the process notion; we merely achieve a multivariate distribution for the data. In fact, the randomly realized covariance matrix has no connection to a covariance function. Connection with spatial separation is sacrificed, e.g., we expect to see many pairs of entries where, say, d_{ij}, the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, is less than $d_{i'j'}$ and $\mathrm{cov}[\mathbf{Y}(\mathbf{s}_i), \mathbf{Y}(\mathbf{s}_j)] < \mathrm{cov}[\mathbf{Y}(\mathbf{s}_{i'}), \mathbf{Y}(\mathbf{s}_{j'})]$. Whether this consequence is desirable or not is likely application dependent. Alternative approaches to remove stationarity are mentioned by *Gelfand et al.* [2003, section 4]. Currently, approaches to nonstationarity through deformation [see *Sampson and Guttorp*, 1992; *Schmidt and O'Hagan*, 2003, and references therein] have only been implemented for univariate data with independent replications of the process.

[6] To formalize the definition of a valid cross-covariance function, we need to model not only the dependence of measurements across locations (the usual spatial setting) but also the dependence at each location. That is, if $\mathbf{Y}^T(\mathbf{s}) = [Y_1(\mathbf{s}), Y_2(\mathbf{s}), \ldots, Y_p(\mathbf{s})]$ and we assume stationarity, we need to specify a $p \times p$ matrix function $C(\mathbf{s} - \mathbf{s}')$, where $C(\mathbf{s} - \mathbf{s}')_{ll'} = \mathrm{cov}[Y_l(\mathbf{s}), Y_{l'}(\mathbf{s}')]$ such that for any n and locations $\mathbf{s}_1$, $\mathbf{s}_2, \ldots, \mathbf{s}_n$, the resulting $np \times np$ covariance matrix for $\mathbf{Y}^T = [\mathbf{Y}(\mathbf{s}_1), \mathbf{Y}(\mathbf{s}_2), \ldots, \mathbf{Y}(\mathbf{s}_n)]$ is positive definite. Note that C need not be symmetric. Moreover, the goal is to enable flexible and computationally tractable covariance structures. Since environmental data are rarely found to follow a Gaussian distribution, we assume that suitable transformation is carried out in order to achieve approximate normality. Our approach can be extended to accommodate heavier-tailed distributions through scale mixing of Gaussian processes.

[7] With regard to the problem of prediction using multivariate spatial processes, let $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\}$ represent a multivariate spatial random field, where $\mathbf{Y}(\mathbf{s}) \in \mathbb{R}^p$ and $D \subset \mathbb{R}^d$, with, usually, $d = 1, 2$ or 3. A well-known approach for prediction at an ungauged location is cokriging, where the predictor of the process at, say, $\mathbf{s}_0$ is described by a linear combination of all the available data values of all the p variables [*Cressie*, 1993]. In cokriging the prediction is based on a known cross-covariance matrix, which, in practice, is not realistic. Usually, the matrix is estimated from the data so that the prediction does not take into account any uncertainty associated with estimating the covariance structure. In this regard, *Cressie* [1993] discusses potential problems with cross variograms with respect to the scaling of the different variables in the vector $\mathbf{Y}(.)$.

[8] We adopt a Bayesian approach here in order to more accurately capture the uncertainty in our model specification. In particular, all unknowns in the model are assumed to arise at random from some (prior) distribution. Using Bayes's theorem, this uncertainty is propagated to (posterior) inference about process unknowns and predictions under the process. Wider interval estimates will usually result, but we would argue that they more realistically express the variability associated with the inference. We choose fairly vague specifications for the prior distributions of our unknowns. Thus our inference becomes fairly insensitive to these choices. Still, a thorough analysis of the data and associated prediction requires some examination of prior sensitivity.

[9] As a result, we are left with the selection of a specific multivariate Gaussian spatial process model, i.e., with the selection of a valid cross-covariance function. We adopt an approach based on the linear coregionalization model first introduced by *Matheron* [1982] with cross covariograms of the form $\Sigma_{m=1}^r \mathbf{T}_m g_m(\|\mathbf{s} - \mathbf{s}'\|)$, where the g_ms are known variograms, the $\mathbf{T}_m$s are unknown nonnegative matrices, and $r(<p)$ gives the number of structures. Each matrix $\mathbf{T}_m$ may be thought of as a scalar product matrix between variables. Therefore covariation between variables is captured in a lower-dimensional space ($r < p$), defined by principal components analysis of $\mathbf{T}_m$ [*Goulard and Voltz*, 1992]. *Goulard and Voltz* [1992] describe a least squares technique to fit such models. Their aim is to estimate the unknown matrices $\mathbf{T}_m$. Therefore they plug in empirical variogram estimates for the functions g_m, ignoring the uncertainty in the estimation of these variogram functions. For more details on coregionalization models with associated classical estimation procedures, see *Wackernagel* [1998]. Our use for the linear coregionalization model is not dimension reduction. Rather, we set $r = p$ to obtain a rich constructive class of valid cross-covariance functions, as we describe in section 3.1.

[10] *Mardia and Goodall* [1993] describe modeling of multivariate spatiotemporal data. They propose a separable covariance structure; in order words, the covariance structure of the multivariate process can be factorized as the product of the correlation across locations by the covariance among variables. That is, for the data vector $\mathbf{Y}$, $\Sigma_{\mathbf{Y}} = \mathbf{R} \otimes \mathbf{T}$, where $R_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \phi)$, with ρ being a valid correlation function in two dimensions and where $\mathbf{T}$ is the covariance matrix for $\mathbf{Y}(\mathbf{s})$. The linear coregionalization model includes this form as the special case when all of the g_m are identical. One implication of this separable form is that the cross covariances are symmetric, i.e., $\mathrm{cov}[Y_k(\mathbf{s}_i), Y_l(\mathbf{s}_j)] = \mathrm{cov}[Y_l(\mathbf{s}_i), Y_k(\mathbf{s}_j)]$, which can be restrictive. Also, $\mathrm{cov}[Y_l(\mathbf{s}), Y_{l'}(\mathbf{s}')]/\sqrt{\mathrm{cov}[Y_l(\mathbf{s}), Y_l(\mathbf{s}')]\mathrm{cov}[Y_{l'}(\mathbf{s}), Y_{l'}(\mathbf{s}')]}$ does not depend on $\mathbf{s} - \mathbf{s}'$, i.e., on the spatial scale. Another limitation is that the spatial range for each component of the process $\mathbf{Y}(\mathbf{s})$ is the same. That is, if ρ is isotropic, say, $\rho(\mathbf{s}_i - \mathbf{s}_j) = \rho(d_{ij}; \phi)$, where ϕ is a decay parameter (equivalently a range parameter), then $\mathrm{cov}[Y_l(\mathbf{s}_i), Y_l(\mathbf{s}_j)] = T_{ll}\rho(d_{ij}, \phi)$. As l varies, we have a different process variability but a common process range. Finally, *Mardia and Goodall* [1993] suggest inference based on the maximum likelihood estimator, which will be problematic if the likelihood is multimodal.

[11] This paper is organized as follows. Section 2 briefly describes the source for our data set. Then, in section 3 we present the Bayesian model based on the linear combination of independent univariate spatial processes, the above linear coregionalization model. The main advantage of this model is that each of the p different processes are allowed to have a different spatial range. We then discuss prior distributions which might be assigned to the parameters of this model.

We can think of the multivariate distribution of our data as arising from a suitable sequence of conditioning, following *Royle and Berliner* [1999]. In fact, we identify univariate spatial conditional models to do this. Section 3.3 presents the reparameterization of our multivariate model in terms of univariate conditional processes and discusses its implementation in the latter form. In section 4 we apply the proposed model to a data set comprising day average measurements of CO, NO and NO_2 from a set of gauged sites in California, USA. Finally, section 5 offers some concluding discussion.

2. Environmental Data

[12] Researchers involved in environmental problems usually face the problem of dealing with data sets of unsatisfying quality. It is quite common that not all of the monitoring networks have observations for the same period of time, or they do not measure the same pollutants and/or covariates which might help in explaining levels of pollutants of interest. In section 4 we will make use of data from the California Air Resources Board, which is a part of the California Environmental Protection Agency. Among the board's aims is the effective and efficient reduction of air pollutants in the state of California.

[13] The Air Resources Board has cosponsored several special field measurement studies for the purpose of collecting air quality, meteorological, and emission data for data analysis and modeling. Descriptions and data are available from the Central California Air Quality Studies.

[14] The board has available 21 years of air quality data. In particular, the data comprises 21 years of criteria pollutant air quality data (1980–2000), 11 years of toxics air quality data (1990–2000), 13 years of dichotomous sampler data (1988–2000), and 7 years of nonmethane organic compound data (1994–2000). (These data can be ordered from the board itself or downloaded directly from http://www.arb.ca.gov/aqd/aqdcd/aqdcddld.htm. On this web site, one can also find files which provide detailed information about the data.)

3. A Bayesian Linear Coregionalization Model

[15] In section 3.1 we propose a multivariate spatial model based on a linear combination of independent spatial processes. Section 3.2 discusses how to perform inference on this model under a Bayesian framework. Then, in section 3.3 we discuss the reparameterization of this model in terms of univariate conditional spatial models. It will be seen that inference based on the conditional univariate models is computationally easier to implement. Finally, in section 3.4, under the proposed model, we show how to obtain the posterior distribution of the spatial range for each component of $Y(s)$.

3.1. Model

[16] We start by developing the separable (or intrinsic) correlation specification of $Y(s)$ (from section 1) through a linear function of random processes. In particular, let $w_j(s)$, $j = 1, \ldots, p$ be p-independent but identically distributed univariate spatial processes, each with unit variance and parametric correlation function $\rho(s - s'; \phi)$. When ρ is isotropic and strictly decreasing for a given ϕ, the distance which makes ρ negligible, i.e., 0.05, is referred to as the range. Now, let $Y_l(s) = \Sigma_{j=1}^{p} a_{lj} w_j(s)$. It is straightforward to see that $C[Y(s), Y(s')] = \rho(s - s'; \phi)T$, where $T = AA^T$, with A being $p \times p$ with elements $A_{lj} = a_{lj}$. The matrix T represents the covariance matrix of the process at any location s. It is referred to as a coregionalization matrix, and the model for $Y(s)$ is referred to as a linear coregionalization model to suggest that the components of $Y(s)$ covary over the region. T is of full rank if and only if A is of full rank. In modeling $Y(s)$ we are only interested in a fully p-dimensional process. We are not interested in the case $j = 1, \ldots, k < p$, i.e., in extracting a lower-dimensional representation for the p variables.

[17] Assuming a common parameter ϕ for each of the spatial processes $w_j(s)$ results in a common range for all the component processes in $Y(s)$. A natural extension is to assume that the $w_j(s)$s are still independent with unit variance but now with correlation function $\rho(s - s'; \phi_j)$. Again, let $Y_l(s) = \Sigma_{j=1}^{p} a_{lj} w_j(s)$. Now we have that $\text{cov}[Y_l(s), Y_{l'}(s')] = \text{cov}[\Sigma_{j=1}^{p} a_{lj} w_j(s), \Sigma_{j=1}^{p} a_{l'j} w_j(s')] = \Sigma_{j=1}^{p} a_{lj} a_{l'j} \rho(s - s'; \phi_j)$. The latter equality can be written in matrix notation as $\Sigma_{Y(s), Y(s')} = C(s - s') = AD_{s,s'}A^T = \Sigma_{j=1}^{p} \rho(s - s'; \phi_j) \ T_j$ with $T_j = a_j a_j^T$, where a_j represents the jth column vector of A and where D is a $p \times p$ diagonal matrix with elements $\rho(s - s'; \phi_j)$, $j = 1, 2, \ldots, p$. By the construction, $C(s - s')$ is a valid cross-covariance function which is nonseparable for the data vector Y at n spatial locations. The covariance structure of the multivariate spatial process is given by

$$\Sigma_Y = \begin{bmatrix} AA' & AD_{1,2}A' & \cdots & AD_{1,n}A' \\ AD_{2,1}A' & AA' & \cdots & AD_{2,n}A' \\ \vdots & \vdots & \ddots & \vdots \\ AD_{n,1}A' & AD_{n,2}A' & \cdots & AA' \end{bmatrix} = \sum_{j=1}^{p} R(\phi_j) \otimes a_j a_j^T,$$

(1)

where $\otimes$ denotes the Kronecker product and $R(\phi_j)$ is $n \times n$ with $[R(\phi_j)]_{ii'} = \rho(s_i - s_{i'}; \phi_j)$. Writing $Y(s) = Aw(s)$, it is clear that the $w_l(s)$ are latent processes which generate the $Y(s)$ given A. Notice that the model has parameters $a_1, a_2, \ldots, a_p, \phi_1, \phi_2, \ldots, \phi_p$. There are $p \times p$ parameters in A when only $p(p + 1)/2$ parameters are required since $AA^T = T$. A convenient reduction is to make A lower triangular. Then, the number of parameters in the model reduces to $p(p + 1)/2 + p$. In this case the model is given by $Y_j(s) = \Sigma_{j=1}^{p} a_{jl} \omega_l(s)$, $j = 1, \ldots, p$:

$$\begin{pmatrix} Y_1(s) \\ Y_2(s) \\ \vdots \\ Y_p(s) \end{pmatrix} = \begin{pmatrix} a_{11} w_1(s) \\ a_{21} w_1(s) + a_{22} w_2(s) \\ \vdots \\ \sum_{l=1}^{p} a_{pl} w_l(s) \end{pmatrix}.$$

(2)

[18] In summary, the $Y(s)$ process is still stationary, has a symmetric cross-covariance matrix, has a different variance for each component, and, when the correlation function ρ is isotropic and monotonic, has a different range for each component. Note that a measurement or nonspatial error term can be added to equation (2). We do not elaborate details here since for our example in section 4 we do not

anticipate nugget effects. However, see *Gelfand et al.* [2003] in this regard.

3.2. Prior Specification and Posterior Distribution

[19] If, for the moment, we assume that $\mathbf{Y}(\mathbf{s})$ has mean 0 and let $\mathbf{\Theta} = (\mathbf{A}, \mathbf{\Phi})$, where $\mathbf{\Phi} = (\phi_1, \ldots, \phi_p)$, then

$$\mathbf{Y}|\mathbf{\Theta} \sim N(\mathbf{0}, \Sigma_Y).$$

Under a Bayesian framework we need to assign prior distributions to all parameters in $\mathbf{\Theta}$, say $p(\mathbf{\Theta})$. Then the posterior distribution $p(\mathbf{\Theta}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{\Theta})p(\mathbf{\Theta})$.

[20] First, notice that there is a 1 to 1 transformation between the elements of the matrices $\mathbf{T}$ and $\mathbf{A}$. For example, if $p = 2$, it is easy to show that

$$\begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} T_{11}^{1/2} & 0 \\ \frac{T_{21}}{T_{11}^{1/2}} & T_{22} - \frac{T_{21}}{T_{11}^{1/2}} \end{pmatrix}.$$

The transformation becomes explicitly messy for larger p but can be developed recursively. For future reference, we record the relationship for $p = 3$: we add

$$a_{31} = \frac{T_{13}}{\sqrt{T_{11}}}, a_{32} = \frac{T_{11}T_{23} - T_{12}T_{13}}{\sqrt{T_{11}T_{22} - T_{12}^2}\sqrt{T_{11}}}$$

and

$$a_{33} = \sqrt{T_{33} - \frac{T_{13}^2}{T_{11}} - \frac{(T_{11}T_{23} - T_{12}T_{13})^2}{T_{11}(T_{11}T_{22} - T_{12}^2)}}.$$

Hence, rather than assign a prior to $\mathbf{A}$, we can assign a prior to $\mathbf{T}$. Since $\mathbf{T}$ is a covariance matrix, a natural choice is an inverse Wishart prior distribution for $\mathbf{T}$ [see *Box and Tiao*, 1992] with low precision and mean, say, $\mathbf{D}_{\sigma^2} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, where σ_l^2 is a prior expected variance for $Y_l(\mathbf{s})$.

[21] As for the choice of correlation function $\rho(.;\phi_j)$, we illustrate with an exponential, $\exp(-\phi_j\|\mathbf{s} - \mathbf{s}'\|)$. We assign an independent Gamma prior distribution to each ϕ_j, which has a large variance and mean based on a crude estimated range, $3/\phi_j$, taken to be half of the maximum interlocation distance. More informative priors for the ϕ_j could be employed if prior knowledge regarding the associated ranges was available. Also, more flexible correlation functions, e.g., the powered exponential or Matérn classes, are available.

[22] Then, the posterior distribution for $\mathbf{\Theta}$ becomes

$$p(\mathbf{\Theta}|\mathbf{Y}) \propto |\Sigma_\mathbf{Y}|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{Y}^T\Sigma_\mathbf{Y}^{-1}\mathbf{Y}\right)p(\mathbf{T})p(\phi_1)\cdots p(\phi_p). \quad (3)$$

Inference can be made using Markov Chain Monte Carlo methods (MCMC), and samples from the posterior distributions of the parameters (or any function of them) could be obtained. See *Gelfand et al.* [2003] for more details on implementing a MCMC algorithm for this model. Unlike previous approaches which use linear coregionalization models [e.g., *Wackernagel*, 1998] uncertainty regarding all the quantities in the model is taken into account, and inference is made simultaneously through the posterior.

Prediction at an ungauged location $\mathbf{s}_{new}$ is based upon the posterior predictive distribution $p(\mathbf{Y}(\mathbf{s}_{new})|\mathbf{Y})$. This distribution is a posterior mixture of multivariate normals and is routinely estimated or sampled (see section 3.3 below).

3.3. Hierarchical Spatial Conditional Modeling

[23] From basic probability theory a joint probability density function can be written as a product of conditional densities. For example, if $p = 3$, the full joint distribution for $\mathbf{Y}(\mathbf{s})$, $p[Y_1(\mathbf{s}), Y_2(\mathbf{s}), Y_3(\mathbf{s})]$, can be written as, for example,

$$p[Y_1(\mathbf{s}), Y_2(\mathbf{s}), Y_3(\mathbf{s})] = p[Y_1(\mathbf{s})]\, p[Y_2(\mathbf{s})|Y_1(\mathbf{s})]\, p[Y_3(\mathbf{s})|Y_2(\mathbf{s}), Y_1(\mathbf{s})].$$
$$(4)$$

The idea of writing a multivariate spatial process as the product of conditional distributions and the potential computational advantage is described, for example, in the work of *Royle and Berliner* [1999]. They call this approach hierarchical because of the parallel with hierarchical modeling in the Bayesian literature and to differentiate it from approaches which condition on auxiliary variables but do not model them spatially. Further, they suggest that in some cases the sequence of "conditioning" in the conditional model be made in a "cause/effect" fashion. They do note that the conditional parameterization is not restricted to this "cause/effect" setting.

[24] We now write the linear coregionalization model in terms of conditional models and show how inference can be simplified. For simplicity, we continue to assume that $p = 3$, though similar results can be obtained, after some algebra, for any value of p. Following equation (4), one can write the joint distribution of $\mathbf{Y}(\mathbf{s})$, for example, as

$$\begin{aligned} Y_1(\mathbf{s}) &= \mu_1 + \sigma_1\tilde{w}_1(\mathbf{s}); \\ Y_2(\mathbf{s})|Y_1(\mathbf{s}) &= \mu_2 + \alpha Y_1(\mathbf{s}) + \sigma_2\tilde{w}_2(\mathbf{s}); \quad (5) \\ Y_3(\mathbf{s})|Y_1(\mathbf{s}), Y_2(\mathbf{s}) &= \mu_3 + \gamma Y_1(\mathbf{s}) + \beta Y_2(\mathbf{s}) + \sigma_3\tilde{w}_3(\mathbf{s}), \end{aligned}$$

where $\tilde{w}_j(\mathbf{s})$ is a mean 0, unit variance Gaussian random field with correlation function $\rho(\mathbf{s} - \mathbf{s}'; \phi_j)$ and where σ_j is a scale parameter. Here μ_j denotes the constant mean of the process $Y_j(.)$. More general mean structure $\mu_j(\mathbf{s})$, capturing, say, a trend surface, could be introduced. The connection between the conditional specification in equation (5) and the associated unconditional one is given by

$$\begin{pmatrix} Y_1(\mathbf{s}) \\ Y_2(\mathbf{s}) \\ Y_3(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 + \alpha\mu_1 \\ \mu_3 + (\gamma + \alpha\beta)\mu_1 + \beta\mu_2 \end{pmatrix} \\ + \begin{pmatrix} \sigma_1 & 0 & 0 \\ \alpha\sigma_1 & \sigma_2 & 0 \\ (\gamma + \alpha\beta)\sigma_1 & \beta\sigma_2 & \sigma_3 \end{pmatrix} \begin{pmatrix} \tilde{w}_1(\mathbf{s}) \\ \tilde{w}_2(\mathbf{s}) \\ \tilde{w}_3(\mathbf{s}) \end{pmatrix}, \quad (6)$$

or, in matrix notation, $\mathbf{Y}(\mathbf{s}) = \mathbf{\mu} + \mathbf{A}\tilde{\mathbf{w}}(\mathbf{s})$, revealing the reparameterization from one to the other. This is clearly a version of the model of equation (2). In fact, $a_{11} = \sigma_1$, $a_{21} = \alpha\sigma_1$, $a_{22} = \sigma_2$, $a_{31} = (\gamma + \alpha\beta)\sigma_1$, $a_{32} = \beta\sigma_2$, and $a_{33} = \sigma_3$. Since $\mathbf{T} = \mathbf{A}\mathbf{A}^T$, the coregionalization matrix can therefore be obtained in terms of α, β, γ, σ_1, σ_2, and σ_3.

[25] Bayesian inference under equation (6) requires the likelihood, which will be written in product form using equations (4) and (5). Under the reparameterization a prior is required for the μs, for the ϕs, and for α, β, γ, σ_1^2, σ_2^2, and σ_3^2. For the μs we used normal priors with a large variance which could be centered at 0 or the sample mean of the corresponding $\mathbf{Y}$s with little sensitivity. α, γ, and β are assumed to be independent, again, normal, with mean 0 and a large variance. With an exponential correlation function a gamma prior distribution is assigned to each ϕ_j, and inverse gamma prior distributions are assigned to the scale parameters σ_j^2. Notice that we can obtain the posterior distribution for each of these three models separately since the likelihood for $\mathbf{Y}$ factors and the prior distributions are assumed to be independent.

[26] A MCMC algorithm can be used in order to make inference regarding the posterior distribution. More specifically, Gibbs sampling [*Gelfand and Smith*, 1990] can be applied. It is straightforward to see that the full conditionals for α, β, and γ are normal and that the full conditionals for the σ_j^2 are inverse gamma. The ϕ_js have nonstandard, messy full conditional distribution since ϕ_j enters through the correlation matrix associated with the jth conditional model. A slice sampler [*Neal*, 2003] can be used to sample from such distributions. Computationally, the fitting of the three univariate conditional models is much simpler than the fitting of the multivariate one as we avoid, at each iteration of the Gibbs sampler, calculation of a likelihood involving a $3n \times 3n$ covariance matrix. Instead, we have three likelihoods, each involving an $n \times n$ matrix.

[27] Notice that prediction of the processes at ungauged locations of interest follows results from the multivariate normal distribution [see, e.g., *Mardia et al.*, 1979]. For instance, for a new location, $\mathbf{s}_{new} p(\mathbf{Y}(\mathbf{s}_{new})|\mathbf{Y}) = \int p(\mathbf{Y}(\mathbf{s}_{new})|\mathbf{Y}, \Theta) p(\Theta|\mathbf{Y})$. Each sample Θ^* from $p(\Theta|\mathbf{Y})$ yields a realization from the predictive distribution by drawing from the multivariate normal distribution $p(\mathbf{Y}(\mathbf{s}_{new})|\mathbf{Y}, \Theta^*)$.

3.4. Ranges of Each of the p Spatial Processes

[28] For any isotropic, monotonic ρ, there is a range associated with $\rho(.;\phi_j)$ which is the range of the $Y_j(\mathbf{s})$ in the corresponding conditional model. More specifically, ϕ_1 denotes the decay of the process $Y_1(\mathbf{s})$ unconditionally. However, for $Y_2(\mathbf{s})$, ϕ_2 represents the decay of the correlation function of the conditional process $Y_2(\mathbf{s})|Y_1(\mathbf{s})$ and not of $Y_2(\mathbf{s})$ marginally. The same idea applies to ϕ_3 and $Y_3(\mathbf{s})|Y_1(\mathbf{s})$, $Y_2(\mathbf{s})$. However, we can obtain the posterior distribution of the ranges of the spatial processes of the unconditional processes in equation (6). We know already that the range of $Y_1(\mathbf{s})$ is the range of $\tilde{w}_1(\mathbf{s})$. The range of $Y_2(\mathbf{s})$ solves $0.05 = \text{corr}[w_2(\mathbf{s}), w_2(\mathbf{s}')]$, where $w_2(\mathbf{s})$ is the spatial process associated with the unconditional variable $Y_2(\mathbf{s})$, that is, from equation (6), $w_2(\mathbf{s}) = \alpha\tilde{w}_1(\mathbf{s}) + \tilde{w}_2(\mathbf{s})$ and therefore

$$\text{corr}[w_2(\mathbf{s}), w_2(\mathbf{s}')] = \frac{\alpha^2\sigma_1^2\rho(\mathbf{s} - \mathbf{s}'; \phi_1) + \sigma_2^2\rho(\mathbf{s} - \mathbf{s}'; \phi_2)}{\alpha^2\sigma_1^2 + \sigma_2^2}. \quad (7)$$

The same idea is applied to obtain the range of $Y_3(\mathbf{s})$. In this case, the range solves $0.05 = \text{corr}[w_3(\mathbf{s}), w_3(\mathbf{s}')]$, where $w_3(\mathbf{s}) = \alpha\beta w_1(\mathbf{s}) + \beta\tilde{w}_2(\mathbf{s}) + \tilde{w}_3(\mathbf{s})$ and
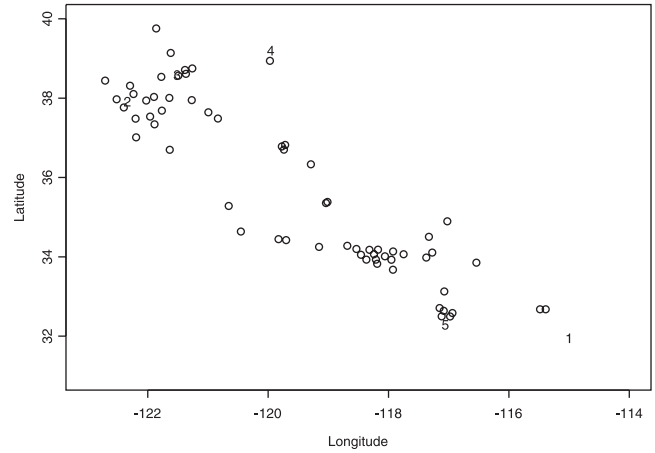
Figure 1. Locations of the 63 monitoring stations used to fit the model and of the 5 stations chosen for prediction of NO_2 (labeled as points $1-5$).

$$\text{corr}[w_3(\mathbf{s}), w_3(\mathbf{s}')]$$
$$= \frac{(\alpha\beta + \gamma)^2\sigma_1^2\rho(\mathbf{s} - \mathbf{s}'; \phi_1) + \beta^2\sigma_2^2\rho(\mathbf{s} - \mathbf{s}'; \phi_2) + \sigma_3^2\rho(\mathbf{s} - \mathbf{s}'; \phi_3)}{(\alpha\beta + \gamma)^2\sigma_1^2 + \beta^2\sigma_2^2 + \sigma_3^2}. \quad (8)$$

Notice that given α, β, γ, σ_1^2, σ_2^2, σ_3^2, ϕ_1, ϕ_2, and ϕ_3, equations (7) and (8) are monotonic functions, so one can use a standard root-finder algorithm to solve equations (7) and (8) for the range of each of the marginal processes. Hence each posterior realization of Θ produces a posterior realization for the range of $Y_2(\mathbf{s})$ and for the range of $Y_3(\mathbf{s})$.

4. Analysis of the California Data

[29] In this section we apply the model proposed in section 3 to a limited data set obtained for a collection of monitoring stations in California. The data were obtained from the California Air Resources Board. We chose to analyze the daily average of carbon monoxide (CO), nitrous oxide (NO), and nitrogen dioxide (NO_2) based on hourly measurements on 16 July 1999. During this year, there were nearly 700 monitoring stations over California, but not all of them measured the above pollutants. After removing missing data and considering only the sites which have measurements for all three of these pollutants we wound up with 68 monitoring stations. Figure 1 shows the locations of these 68 monitoring sites on a degrees latitude by degrees longitude scale. Since, in the sampling area, $1°$ of latitude ≈ 65 km while $1°$ of longitude ≈ 110 km, we projected the locations using the Lambert 2 parallel projection method to produce accurate interlocation distances. Figure 1 also shows five numbered sites which were held out and used for prediction of NO_2. The observed correlations between these pollutants were 0.46 (CO and NO), 0.56 (CO and NO_2), and 0.77 (NO and NO_2). In order to achieve approximate normality, we use the logarithm of the daily average of each of these variables.

[30] There was no information on potential covariates, such as temperature or wind direction, at these gauged sites. Therefore we fit a model with a constant mean structure, as in equation (5). Owing to the observed correlations, we

Table 1. Posterior Summaries of All the Parameters in the Joint Model for CO (Y_1), NO (Y_2), and NO$_2$ (Y_3)

Parameter	Mean	2.50%	Median	97.50%
α	0.317	0.086	0.3123	0.575
β	0.309	0.196	0.308	0.419
γ	0.192	0.088	0.190	0.301
μ_{CO}	−0.896	−1.093	−0.892	−0.722
μ_{NO}	−5.278	−5.668	−5.268	−4.923
μ_{NO_2}	−4.366	−4.705	−4.363	−4.029
ϕ_1	0.088	0.039	0.072	0.178
ϕ_2	0.027	0.012	0.025	0.052
ϕ_3	0.013	0.005	0.012	0.026
σ_1^2	0.346	0.246	0.337	0.489
σ_2^2	0.629	0.397	0.602	1.034
σ_3^2	0.206	0.117	0.191	0.411
CO range, km	43.38	16.89	41.77	76.69
NO range, km	126.86	55.59	118.09	255.01
NO$_2$ range, km	113.02	50.29	102.46	237.63

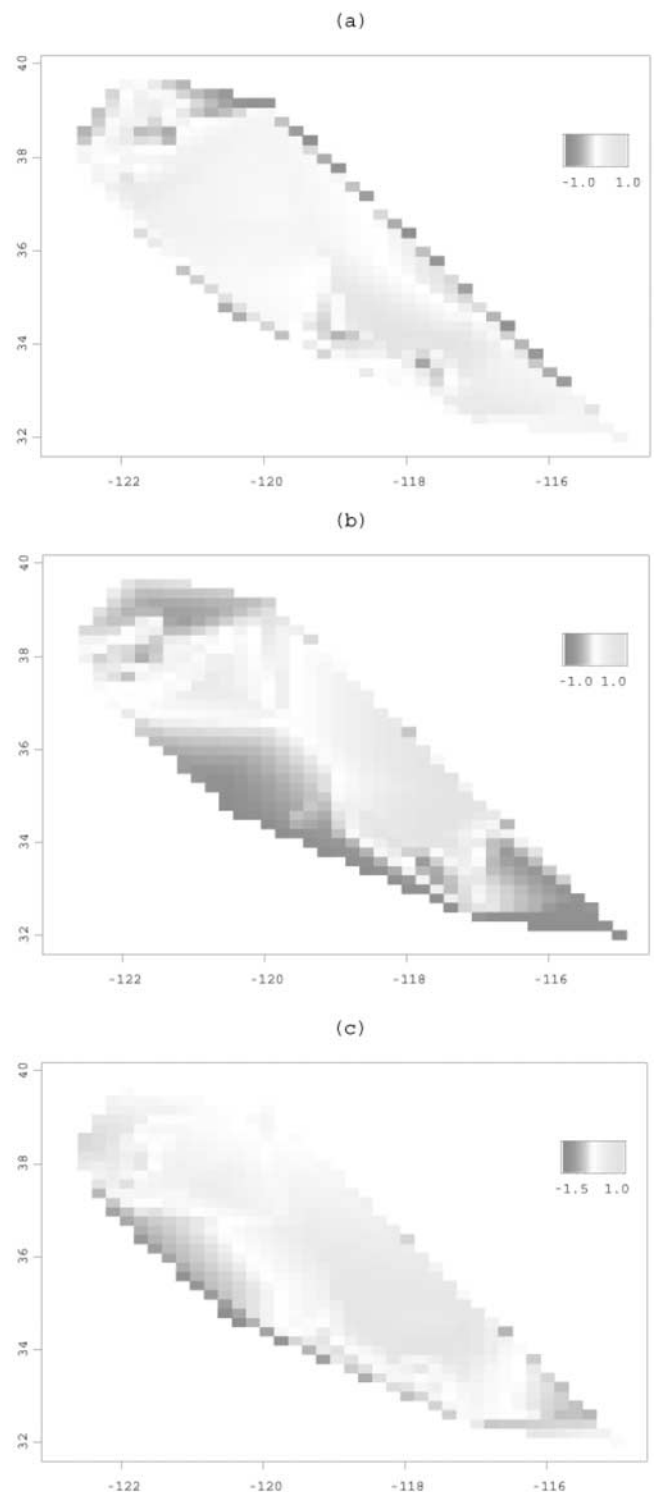decided to order the condition in the following way: $Y_1(\mathbf{s}) =$ CO$(\mathbf{s})$, $Y_2(\mathbf{s}) =$ NO$(\mathbf{s})$, and $Y_3(\mathbf{s}) =$ NO$_2(\mathbf{s})$; however, any other order could have been used leading to the same joint distribution. Following section 3.3, we assumed that α, β, γ, μ_1, μ_2, and μ_3 are all independent and normally distributed a priori. They are centered at 0 with a variance 10 times the variance of the coefficient estimate obtained under an ordinary least squares (OLS) fit, i.e., under equation (5), with $\rho(\mathbf{s} - \mathbf{s}'; \phi_j) = 0$ if $\mathbf{s} \neq \mathbf{s}'$. For each σ_j^2 we use an inverse gamma having infinite variance, with a mean equal to the OLS estimate. For the ϕ_j's parameters we use gamma priors arising from a mean for the associated range of one half the maximum interlocation distance, with a large variance.

[31] We made use of the software GeoBugs available through Bugs [*Spiegelhalter et al.*, 1996] in order to obtain samples from the posterior distributions of all the parameters involved in the model. The ranges of the processes of CO, NO, and NO$_2$ were obtained following equations (7) and (8). These equations were rapidly solved using the "false position method" as described by *Press et al.* [1999]. Table 1 shows the posterior summaries for all parameters in the model and also for the ranges of the processes. The benefit of the added flexibility of the coregionalization model is clear, as is the Bayesian learning. Starting from a common prior, we observe that the posterior distribution of the ranges for the three pollutants are all different, with CO having the smallest range among the three. Figure 2 also shows the unconditional process means, i.e., $EY_1(\mathbf{s}) = \mu_{CO}$, $EY_2(\mathbf{s}) = \mu_{NO}$, and $EY_3(\mathbf{s}) = \mu_{NO_2}$, where the forms for these means are given in equation (6). Figure 3 presents a

grayscale plot showing the mean-adjusted estimated spatial surface for each unconditional process. In particular, these are developed from the posterior means of the $w_j(\mathbf{s})$s over a fine grid obtained as functions of the $\tilde{w}_j(\mathbf{s})$s, again using equation (6). Resulting from the use of log $Y_j(\mathbf{s})$ as the response,

$$
\begin{array}{ccc|ccc}
\multicolumn{3}{c}{\mathbf{T}} & \multicolumn{3}{c}{\mathbf{R}} \\
0.34 & 0.11 & 0.10 & 1 & 0.23 & 0.32 \\
(0.25;0.49) & (0.02;0.21) & (0.05;0.17) & & (0.06;0.43) & (0.17;0.48) \\
 & 0.64 & 0.22 & & 1 & 0.51 \\
 & (0.43;1.08) & (0.13;0.40) & & & (0.33;0.69) \\
 & & 0.29 & & & 1 \\
 & & (0.19;0.50) & & & \\
\end{array}
$$

Figure 2. Posterior median with the associated 95% credible interval (in parentheses) of the elements of the coregionalization matrix and the correlation matrix for each location $\mathbf{s}$.

Figure 3. Adjusted posterior mean surfaces for the unconditional spatial processes of (a) CO, (b) NO, and (c) NO$_2$. See text for details. See color version of this figure at back of this issue.

Table 2. Prediction of NO_2 Based on the Three Different Models

Site	Mean	2.50%	Median	97.50%	Observed
(1) Unconditional Model for NO_2					
1	−4.869	−5.986	−4.839	−3.802	**−4.342**
2	−4.624	−5.082	−4.632	−4.127	**−4.585**
3	−4.294	−4.679	−4.294	−3.896	**−4.100**
4	−4.55	−5.386	−4.565	−3.734	**−4.342**
5	−4.184	−4.756	−4.18	−3.588	**−4.906**
(2) Model for NO_2 Conditioned on CO					
1	−4.722	−5.712	−4.733	−3.73	**−4.342**
2	−4.7	−5.106	−4.702	−4.301	**−4.585**
3	−4.132	−4.471	−4.131	−3.794	**−4.100**
4	−4.49	−5.196	−4.485	−3.816	**−4.342**
5	−4.304	−4.797	−4.306	−3.807	**−4.906**
(3) Model for NO_2 Conditioned on CO and NO					
1	−4.5	−5.313	−4.508	−3.679	**−4.342**
2	−4.585	−4.964	−4.587	−4.22	**−4.585**
3	−3.966	−4.26	−3.964	−3.653	**−4.100**
4	−4.57	−5.158	−4.575	−3.937	**−4.342**
5	−4.17	−4.594	−4.169	−3.721	**−4.906**

these spatial effects are generally between -1.5 and 1.0. The spatial patterns are different; CO tends to be elevated along the Pacific coast, while NO is more depressed there.

[32] Using the 1 to 1 transformation between the univariate conditional parameterization and the multivariate one (equation (6) shows the matrix A), we can obtain the posterior distribution of the elements of **T**, the within-location covariance matrix of the pollutants. The posterior mean entries and 95% credible interval are shown in Figure 2. Evidently, the NO measurements are more variable than those for CO and NO_2. Also shown in Figure 2 are the entries in the posterior correlation matrix, **R**, based on **T**. Figure 2 confirms the information from the data that NO_2 and NO tend to be more correlated, with the posterior median being 0.52. Additionally, a posteriori, CO is more correlated with NO_2 than with NO, although the interval estimates overlap.

[33] To perform some validation of the model, we return to the five hold out sites (see Figure 1) and predict the levels of NO_2 at these locations under three different models: (1) $NO_2(\mathbf{s})$ is modelled unconditionally as $NO_2(\mathbf{s}) = \mu_1 + \sigma_1 w_1(\mathbf{s})$; (2) $NO_2(\mathbf{s})$ is modelled conditionally given the levels of $CO(\mathbf{s})$, i.e., $NO_2(\mathbf{s})|CO(\mathbf{s}) = \mu_2 + \alpha CO(\mathbf{s}) + \sigma_2 w_2(\mathbf{s})$; and (3) $NO_2(\mathbf{s})$ is modelled conditionally given the levels of both $CO(\mathbf{s})$ and $NO(\mathbf{s})$ as $NO_2(\mathbf{s}) = \mu_3 + \gamma CO(\mathbf{s}) + \beta NO(\mathbf{s}) + \sigma_3 w_3(\mathbf{s})$. It is expected that models 2 and 3 will present successively better prediction than will model 1. We fit each of the models with the same priors as discussed in section 3.3, and then we perform the prediction for the five sites left out. Again, this was done using GeoBugs. The posterior summary of the predictions, using each of the models, is shown in Table 2. The models all fail to predict correctly at site 5. Site 5 may be anomalous or might suggest that the use of a stationary (in fact, isotropic) model for spatial association is inappropriate. However, the predictive intervals become increasingly tighter, as anticipated.

5. Discussion and Future Work

[34] The foregoing methodology can still be used in the presence of missing components in the $\mathbf{Y}(\mathbf{s})$s. The fully Bayesian approach would introduce these missing $Y_j(\mathbf{s}_i)$s as

unknowns in the model. In the simulation-based model fitting they would be simulated along with all of the other model unknowns. This would result in posterior predictive distributions for the missing components. Viewed as a Gibbs sampler, updating all of the parameters given values of the missing $Y_j(\mathbf{s}_i)$s and the observed $Y_j(\mathbf{s}_i)$s is precisely the updating given a full data set that we have proposed in section 3.3. However, then, to update the missing $Y_j(\mathbf{s}_i)$s will be very awkward. Obtaining the full conditional distributions for these Ys will require a considerable additional amount of matrix computation. The simulation will run very slowly and will become ill behaved if too many data are missing. An inexpensive approximation is to employ multiple imputation [*Schafer*, 1997]. Such imputation can be implemented in various ways; we omit the details.

[35] Also of interest would be to build spatiotemporal versions of equations (5) and (6). There are many possibilities here. A simple and computationally convenient choice would be to use a cross-covariance function which is separable in space and time (see, e.g., *Kyriakidis and Journel* [1999] for a review in the univariate case). Again, we omit the details. Of course, if the measurements obtained over time are viewed as independent replications, then the likelihood under conditional or joint modeling just becomes a product over the replications. With an increasing number of components, locations, and time points, the computational demand to fit such models will exceed current capabilities. Suitable approximations to handle such high-dimensional situations are being explored.

[36] The modeling in equations (2) and (5) yields a stationary multivariate process. In few practical situations will stationarity be a reasonable assumption; nonstationary specifications would be sought. Two straightforward approaches are possible here. First, the $w_j(\mathbf{s})$s could be taken to be convenient univariate nonstationary processes. Alternatively, we could allow **A** to vary spatially, i.e., define $\mathbf{Y}(\mathbf{s}) = \mathbf{A}(\mathbf{s}) \, \mathbf{w}(\mathbf{s})$, where the specification of $\mathbf{A}(\mathbf{s})$ can be made in various ways. We omit the details here.

[37] Finally, in some settings, there might be a need to introduce a microscale or nugget effect into equations (2) or (5). In this case, there is evidently still an equivalence between the conditional and the unconditional specifications. However, given a residual in the joint model which consists of a spatial component plus a white noise component, the conditional version no longer provides a residual which is composed of a spatial term plus a white noise term. Moreover, the nonspatial components of these residuals are no longer independent across the conditional pieces (see *Gelfand et al.* [2003] for details).
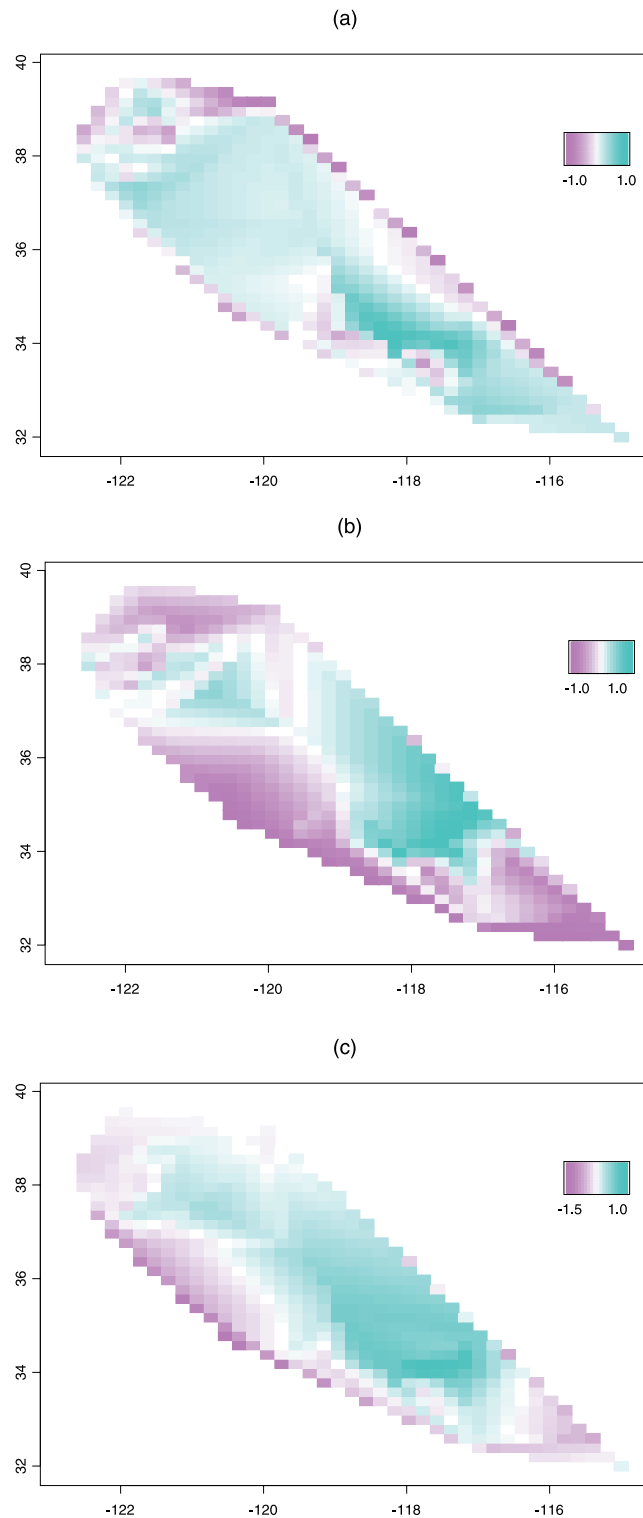
References

Box, S. E. P., and S. C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley, Hoboken, N. J., 1992.

Brown, P. J., N. D. Le, and J. V. Zidek, Multivariate spatial interpolation and exposure to air pollutants, *Can. J. Stat.*, 22, 489–509, 1994.

Cressie, N. A. C., *Statistics for Spatial Data*, revised ed., John Wiley, Hoboken, N. J., 1993.

Gelfand, A. E., and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, *J. Am. Stat. Assoc.*, *85*, 398–409, 1990.

Gelfand, A. E., and P. Vounatsou, Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics*, *4*, 11–25, 2003.

Gelfand, A. E., A. M. Schmidt, and C. F. Sirmans, Multivariate spatial process models: Conditional and unconditional Bayesian approaches using coregionalization, *Tech. Rep. 20*, Inst. of Stat. and Decis. Sci., Duke Univ., Durham, N. C., 2003.

Goulard, M., and M. Voltz, Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix, *Math. Geol.*, *24*, 269–286, 1992.

Kyriakidis, P. C., and A. G. Journel, Geostatistical space-time models: A review, *Math. Geol.*, *31*, 651–683, 1999.

Mardia, K. V., Multi-dimensional multivariate Gaussian Markov random fields with application to image processing, *J. Multivariate Anal.*, *19*, 265–284, 1988.

Mardia, K. V., and C. R. Goodall, Spatial-temporal analysis of multivariate environmental monitoring data, in *Multivariate Environmental Statistics*, edited by G. P. Patil and C. R. Rao, pp. 347–386, Elsevier Sci., New York, 1993.

Mardia, K. V., J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic, Norwell, Mass., 1979.

Matheron, G., Pour une analyse krigeante des données régionaliées, *Publ. N-732*, Cent. de Geostat., Fountainebleau, France, 1982.

Neal, R., Slice sampling (with discussion), *Ann. Stat.*, *31*, 705–767, 2003.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C-The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, New York, 1999.

Royle, J. A., and L. M. Berliner, A hierarchical approach to multivariate spatial modeling and prediction, *J. Agric. Biol. Environ. Stat.*, *4*, 1–28, 1999.

Sampson, P. D., and P. Guttorp, Nonparametric estimation of nonstationary spatial covariance structure, *J. Am. Stat. Assoc.*, *87*, 108–119, 1992.

Schafer, J. L., *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York, 1997.

Schmidt, A. M., and A. O'Hagan, Bayesian inference for non-stationary spatial covariance structure via spatial deformations, *J. R. Stat. Soc.*, *B*, *65*, 743–758, 2003.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks, Bugs 0.5 Bayesian inference using Gibbs sampling—Manual (version ii), *Tech. Rep. XXX*, MRC Biostat. Unit, Inst. of Public Health, Cambridge, U. K., 1996.

Wackernagel, H., *Multivariate Geostatistics—An Introduction With Applications*, 2nd ed., Springer-Verlag, New York, 1998.

———————————

A. E. Gelfand, Institute of Statistics and Decision Sciences, Duke University, Box 90251, Durham, NC 27708-0251, USA. (alan@stat.duke.edu)

A. M. Schmidt, Department of Statistical Methods, Federal University of Rio de Janeiro, Caixa Postal 68530, Rio de Janeiro, RJ CEP 21945-970, Brazil. (alex@im.ufrj.br)

Figure 3. Adjusted posterior mean surfaces for the unconditional spatial processes of (a) CO, (b) NO, and (c) NO_2. See text for details.