Variance-Covariance Modeling and Estimation for Multi-Resolution Spatial Models

G. Johannesson, *The Ohio State University*
N. Cressie, *The Ohio State University*

Technical Report No. 700

December, 2002

Department of Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43210-1247

VARIANCE-COVARIANCE MODELING AND ESTIMATION FOR MULTI-RESOLUTION SPATIAL MODELS

GARDAR JOHANNESSON (gardar@stat.ohio-state.edu)
and NOEL CRESSIE (ncressie@stat.ohio-state.edu)
Department of Statistics
The Ohio State University
Columbus, OH 43210
USA

Abstract. The tree-structured multi-resolution spatial models (MRSMs) yield optimal and computationally feasible spatial smoothers of massive spatial data with nonstationary behavior. The nonstationary spatial correlation structure of MRSMs is the result of inhomogeneous stochastic parent-child relationships at adjacent resolutions. Likelihood-based methods are presented for the estimation and modeling of variance-covariance parameters associated with the parent-child relationships, resulting in data-adaptive, nonstationary covariance structure. An application of the MRSMs is given to total column ozone (TCO) data obtained from a polar-orbiting satellite.

Key words: Nonstationarity, RESL estimation, RESREL estimation, total column ozone, tree-structured models, covariance-parameter estimation

1. Introduction

As a consequence of new remote-sensing technology, spatio-temporal environmental data have become more massive in their raw form. Provided with such rich datasets, scientists eye new opportunities, but at the same time they are faced with new challenges. The massiveness of the data is in most cases due to both *fine-resolution* sampling and a *large* spatial domain. An example is Total Column Ozone (TCO), sampled remotely

by satellites over the entire globe on a daily basis. Due to the large size of the spatial domain, stationarity assumptions about the process of interest do not typically hold. Hence, computationally tractable spatial models for massive data, with nonstationary spatial dependence, are in great demand.

Tree-structured multi-resolution spatial models (MRSMs) (see e.g., Huang et al., 2002) are able to handle massive spatial data with nonstationary spatial correlation structure. In Section 2, we shall review the MRSM and the associated fast, change-of-resolution Kalman-filter algorithm for optimal spatial prediction. At the core of the MRSM is the specification of the spatial covariance structure through a coarse-to-fine-resolution process model. Section 3 considers such models and proposes a parameterization that allows one to capture smooth changes in (nonstationary) spatial covariance structure. We also show in Section 3 how to estimate the model parameters using resolution-specific likelihood-based methods. An application to a day's worth of TCO satellite data is presented in Section 4.

2. Multi-resolution Spatial Models

In this section, we review briefly the multi-resolution spatial model (MRSM) as given in Huang et al. (2002). Let D be the spatial domain of interest. The domain D is partitioned into n_0 grid cells, which make up the coarsest resolution (resolution-0). Each grid cell at resolution $r = 0, \ldots, R-1$, is then successively partitioned into m_r smaller grid cells. Thus, we obtain a nested partition of D at $(R+1)$ resolutions. At the r-th resolution, there are $n_r = n_0 m_0 \cdots m_{r-1}$ grid cells given by $\{D(i,r)\}_{i=1}^{n_r}$. We call $(i^*, r+1)$ a child of (i,r) if $D(i^*, r+1) \subset D(i,r)$, and we denote the set of the children of (i,r) by $ch(i,r) \equiv \{ch(i,r)_1, \ldots, ch(i,r)_{m_r}\}$. Then

$$D(i,r) = \bigcup_{j=1}^{m_r} D(ch(i,r)_j)\,; \quad (i,r) \in \mathcal{N}_{R-1},$$

where $\mathcal{N}_u \equiv \{(i,r) : i = 1, \ldots, n_r,\ r = 0, \ldots, u\}$. Figure 1 shows an example of a multi-resolution partition at resolutions $r = 0, 1, 2$.

Let $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ be a Gaussian spatial process of interest defined on D, and define the multi-resolution aggregated Y-process as

$$Y(i,r) \equiv \frac{1}{v(i,r)} \int_{D(i,r)} Y(\mathbf{s}) d\mathbf{s}\,; \quad (i,r) \in \mathcal{N}_R,$$

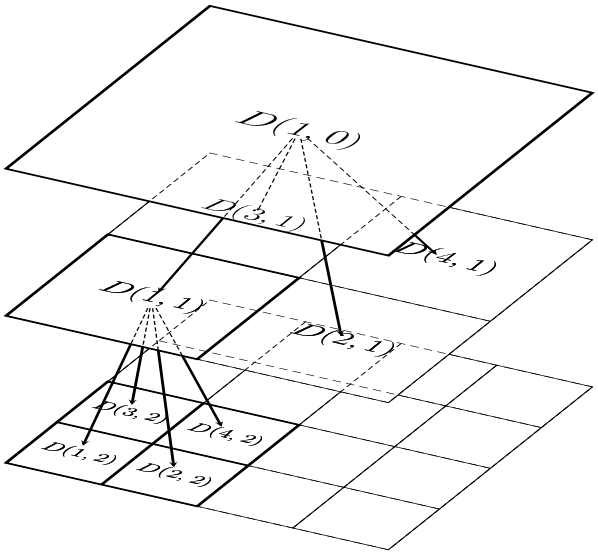

Figure 1. An example of a spatial multi-resolution tree-structure partition.

where $v(i,r) \equiv |D(i,r)|$ denotes the area (volume) of $D(i,r)$. The aggregated Y-process is not observed directly, but indirectly through the additive-measurement-error model,

$$Z(i,r) = Y(i,r) + \nu(i,r)\,; \quad (i,r) \in \mathcal{N}_R, \tag{1}$$

where $\{Z(i,r)\}$ are (potentially) observed data, and the measurement errors $\nu(i,r) \sim \text{Gau}(0, \sigma^2 V(i,r))$ are independent with $\{V(i,r)\}$ known. Henceforth, we refer to (1) as the data model. It should be noted that observations are not needed at all resolutions and can be missing for some cells within a resolution. For example, in the ozone example considered in Section 4, the Y-process is taken to be the underlying TCO process at different resolutions and the data are noisy satellite observations of TCO, reported (incompletely) at the finest resolution, resolution-R.

The spatial variance-covariance structure associated with the Y-process is specified indirectly through the following coarse-to-fine-resolution model:

$$\mathbf{Y}(i,r) = \mathbf{1}Y(i,r) + \omega(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1}, \tag{2}$$

where $\mathbf{Y}(i,r) \equiv (Y(ch(i,r)_1), \ldots, Y(ch(i,r)_{m_r}))'$ and $\omega(i,r) \sim \text{Gau}(\mathbf{0}, \sigma^2 \mathbf{W}(i,r))$, independently. Henceforth, we refer to (2) as the process model. Hence, the

Y-process at the children cells is just taken to be equal the the Y-process at the parent cell plus an error term. The process model is completed by specifying the distribution of the Y-process at the coarsest resolution; here we simply assume that $(Y(1,0),\ldots,Y(n_r,0))' \sim \text{Gau}(\mathbf{a}(0),\sigma^2\mathbf{R}(0))$.

To match the notation style used for the process model in (2), it will be more convenient to write the data model in (1) as

$$\mathbf{Z}(i,r) = \mathbf{Y}(i,r) + \nu(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1}, \tag{3}$$

where $\mathbf{Z}(i,r) \equiv (Z(ch(i,r)_1),\ldots,Z(ch(i,r)_{m_r}))'$, and $\nu(i,r) \sim \text{Gau}(\mathbf{0},\sigma^2\mathbf{V}(i,r))$, independently, with $\mathbf{V}(i,r) \equiv \text{diag}(V(ch(i,r)_1),\ldots,V(ch(i,r)_{m_r}))$.

2.1. CONSTRAINED Y-PROCESS

Note that the process model in (2) does not have a one-to-one mapping between $\mathbf{Y}(i,r)$ and $\{Y(i,r),\omega(i,r)\}$; $\mathbf{Y}(i,r)$ is a vector of length m_r, but $\{Y(i,r),\omega(i,r)\}$ has a total of (m_r+1) elements. Consequently, different configurations of $\{Y(i,r),\omega(i,r)\}$ can yield the same $\mathbf{Y}(i,r)$. However, by placing a single linear constraint on the error term $\omega(i,r)$, a one-to-one mapping is achieved. That is, we constrain

$$\mathbf{q}(i,r)'\omega(i,r) = 0\,; \quad (i,r) \in \mathcal{N}_{R-1}, \tag{4}$$

for some chosen constraining vectors $\{\mathbf{q}(i,r)\}$. To satisfy (4), let $\mathbf{Q}(i,r)$ be any $m_r \times (m_r-1)$ orthonormal matrix with columns that span the space orthogonal to $\mathbf{q}(i,r)$ (i.e., $\mathbf{q}(i,r)'\mathbf{Q}(i,r) = \mathbf{0}$ and $\mathbf{Q}(i,r)'\mathbf{Q}(i,r) = \mathbf{I}$). Then any $\omega(i,r)$ satisfying (4) can be written as

$$\omega(i,r) = \mathbf{Q}(i,r)\omega^*(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1},$$

for some unconstrained $\omega^*(i,r) \in \mathbb{R}^{m_r-1}$. The constrained Y-process can therefore be written as:

$$\mathbf{Y}(i,r) = \mathbf{1}Y(i,r) + \mathbf{Q}(i,r)\omega^*(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1}, \tag{5}$$

where $\omega^*(i,r) \sim \text{Gau}(\mathbf{0},\sigma^2\mathbf{W}^*(i,r))$, independently. In terms of the process model in (2), we have constrained $\mathbf{W}(i,r)$ to be of the form:

$$\mathbf{W}(i,r) = \mathbf{Q}(i,r)\mathbf{W}^*(i,r)\mathbf{Q}(i,r)'\,; \quad (i,r) \in \mathcal{N}_{R-1}. \tag{6}$$

Huang et al. (2002) proposed choosing $\mathbf{q}(i,r) = \mathbf{v}(i,r)$, where $\mathbf{v}(i,r) \equiv (v(ch(i,r)_1), \ldots, v(ch(i,r)_{m_r}))'$. This choice results in a physically *mass-balanced* process model, since it follows that

$$v(i,r)Y(i,r) = \sum_{j=1}^{m_r} v(ch(i,r)_j)Y(ch(i,r)_j)\,; \quad (i,r) \in \mathcal{N}_{R-1}. \tag{7}$$

2.2. POSTERIOR INFERENCE

Given all the variance-covariance parameters associated with the data model in (1) and the process model in (2), our goal is to predict the hidden process $\{Y(i,r)\}$ from noisy and incomplete data $\{Z(i,r)\}$. Optimal prediction is obtained from the *posterior distribution* of $\{Y(i,r)\}$, which can be calculated rapidly using the change-of-resolution Kalman-filter algorithm (Chou et al., 1994; Huang and Cressie, 2001). The algorithm consists of two major steps, namely the leaves-to-root step and the root-to-leaves step. The leaves-to-root step consists of recursively deriving the distribution of $Y(i,r)$ conditional on all data observed at all descendents of (i,r) and at (i,r) itself. At the end of the leaves-to-root recursion, we obtain the distribution of $\{Y(i,0)\}$ conditional on all the data (i.e., the posterior distribution of $\{Y(i,0)\}$). The root-to-leaves step starts at the root node, and then traces down the tree, recursively computing the posterior distribution of $Y(i,r)$ at every node in the tree. The algorithm is fast; it requires computations only proportional to the number of nodes in the tree, with a small computational overhead at each node. Computation times are discussed in Section 4.

3. Variance-Covariance Modeling and Estimation

In Section 2, the scalars $\{V(i,r)\}$ associated with the measurement errors in (1), and the parameters $\{\mathbf{W}^*(i,r)\}$, $\mathbf{a}(0)$, and $\mathbf{R}(0)$ associated with the process model in (5), were assumed known. This assumption is realistic for the $\{V(i,r)\}$, since they reflect the relative accuracy (weight) of each observation. On the other hand, the matrices $\{\mathbf{W}^*(i,r)\}$ and $\mathbf{R}(0)$ determine the variance-covariance structure of the hidden process $\{Y(i,r)\}$, *a priori*. A common approach in spatial statistics is to use the data to assist in specifying the variance-covariance structure of the Y-process, which can be thought of as an empirical Bayes approach. For example, when

doing kriging (e.g., Cressie, 1993, Chapter 3), the data are typically used to estimate variance-covariance parameters using, for example maximum likelihood (ML) or restricted maximum likelihood (REML) estimation (e.g., Cressie, 2002). We follow a similar approach here by parameterizing the $\{\mathbf{W}^*(i,r)\}$ matrices and then estimating any unknown parameters using ML- and REML-based methods. Estimation of $\mathbf{a}(0)$ and $\mathbf{R}(0)$ is discussed in Section 4.

3.1. VARIANCE-COVARIANCE MODELING

The $(m_r-1)\times(m_r-1)$ matrix $\mathbf{W}^*(i,r)$ has at most $(m_r-1)m_r/2$ unknown parameters associated with it that need to be estimated. Denote by $\theta(i,r)$ the unknown parameter vector associated with $\mathbf{W}^*(i,r)$, and write

$$\mathbf{W}^*(i,r) = \mathbf{W}^*_r(\theta(i,r))\,; \quad (i,r) \in \mathcal{N}_{R-1}. \tag{8}$$

An example of a $\mathbf{W}^*$-model is the single-parameter-per-scale (SPPS) model:

$$\mathbf{W}^*(i,r) = \theta(i,r)\mathbf{C}_0(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1}, \tag{9}$$

where $\{\mathbf{C}_0(i,r)\}$ are known positive-definite matrices and $\{\theta(i,r)\}$ are unknown, positive, scaling parameters.

As presented above, the different $\{\theta(i,r)\}$ in (8) are not related in any way. However, one could expect that cells within the same resolution that are nearby (in space) will have similar θ-parameters. Let $\{\mathbf{s}(i,r)\}$ be a set of representative point locations for $\{D(i,r)\}$ (e.g., using the centroids of each cell). Then, in the case of the SPPS model (9), for example, one could assume

$$\log\theta(i,r) = \textstyle\sum_{j=1}^{p_r} \psi_j(\mathbf{s}(i,r))\beta(r)_j, \tag{10}$$

within each resolution r, where $\psi_1(\cdot),\ldots,\psi_{p_r}(\cdot)$ are known, smooth basis-functions of spatial locations, $\beta(r) \equiv (\beta(r)_1,\ldots,\beta(r)_{p_r})'$ are unknown parameters to be estimated, and $p_r \in \{1,2,\ldots\}$. We now present likelihood-based methods for estimating $\{\beta(r)\}$.

3.2. LIKELIHOOD-BASED PARAMETER ESTIMATION

Denote by

$$p(\mathbf{Z}(i,r) \,|\, \mathbf{Y}(i,r)) \quad \text{and} \quad p(\mathbf{Y}(i,r) \,|\, Y(i,r); \theta(i,r)), \tag{11}$$

the conditional Gaussian probability densities associated with the data model (3) and the process model (5), respectively, and assume for the moment that σ^2, the variance-scaling parameter in (3) and (5) is known. With very little loss of generality, assume further that the data are only observed at the finest resolution, resolution-R. Due to the conditional structure of the MRSM, the joint density of $\{Z(i,R)\}$ and $\{Y(i,r)\}$ is given simply by a product of conditional densities. That is,

$$\begin{aligned} p(\{Z(i,R)\}, \{Y(i,r)\}; \{\theta(i,r)\}) &= \left(\textstyle\prod_{i=1}^{n_{R-1}} p(\mathbf{Z}(i,R-1) \,|\, \mathbf{Y}(i,R-1))\right) \\ &\times \left(\textstyle\prod_{r=1}^{R-1} \prod_{i=1}^{n_r} p(\mathbf{Y}(i,r) \,|\, Y(i,r); \theta(i,r))\right) p(\{Y(i,0)\}), \end{aligned} \tag{12}$$

where $\{\mathbf{Z}(i,R-1)\}$ is equivalent to $\{Z(i,R)\}$ and recall that the last factor is the density of the multivariate Gau$(\mathbf{a}(0), \sigma^2\mathbf{R}(0))$. However, for maximum-likelihood inference, the marginal distribution of the data $\{Z(i,R)\}$ is needed, which is the integral the joint distribution above with respect to $\{Y(i,r)\}$. This integration is not at all straightforward, and it leaves us with a likelihood that has to be simultaneously maximized with respect to all variance-covariance parameters. However, as we shall see, it is possible to extract information from the data that is relevant to each resolution separately, leading to fast, resolution-specific likelihood inference. One such approach, given by Kolaczyk and Huang (2001), is to combine a recursive integration of (12) with recursive aggregation and transformation of the data. The resulting marginal distribution of the transformed data factors into resolution-specific likelihood (RESL) components, with each component being only informative for the variance-covariance parameters associated with that particular resolution. Another such approach, which mirrors REML estimation in mixed-effects models (e.g., McCulloch and Searle, 2001), is to form contrasts among the data such that the distribution of the contrasted data only depends on the variance-covariance parameters associated with a single resolution. The resolution-restricted likelihood-based (RESREL-based) estimates derived using this latter approach are

not in general the same as the RESL-based estimates obtained from the first approach. However, when estimating variance-covariance parameters in Gaussian mixed-effect models and in Gaussian spatial models, REML estimators are in many cases preferred (see, e.g., McCulloch and Searle, 2001, Section 6.10; Cressie, 2002). We now present briefly both estimation approaches; see Johannesson (2003) for full details.

The RESL is derived by effectively integrating (12), resolution-by-resolution, with the help of a recursive decomposition of the data. Let $r = R - 1$. Integrating (12) with respect to $\{\mathbf{Y}(i,r)\}_{i=1}^{n_r}$, results in most terms coming outside the integral, leaving behind

$$\prod_{i=1}^{n_r} \int_{\mathbf{Y}(i,r)} p(\mathbf{Z}(i,r) \mid \mathbf{Y}(i,r))\, p(\mathbf{Y}(i,r) \mid Y(i,r); \theta(i,r))\, d\mathbf{Y}(i,r), \qquad (13)$$

for $r = R-1$. The i-th integral in (13) is easily seen to be $p(\mathbf{Z}(i,r) \mid Y(i,r); \theta(i,r))$, which can be obtained from the additive model,

$$\mathbf{Z}(i,r) = \mathbf{1}Y(i,r) + \mathbf{Q}(i,r)\omega^*(i,r), +\nu(i,r)\,; \quad i = 1, \ldots, n_r. \qquad (14)$$

Instead of proceeding to next resolution and taking a second integral of (12), now with respect to $\{\mathbf{Y}(i, r-1)\}_{i=1}^{n_{r-1}}$, we decompose the $\{\mathbf{Z}(i,r)\}$ into aggregated global components $\{Z(i,r)\}$ and detail local components $\{\mathbf{d}(i,r)\}$. Define

$$\begin{bmatrix} Z(i,r) \\ \mathbf{d}(i,r) \end{bmatrix} \equiv \begin{bmatrix} \tilde{\mathbf{q}}(i,r)' \\ \mathbf{P}(i,r)' \end{bmatrix} \mathbf{Z}(i,r)\,; \quad i = 1, \ldots, n_r,\ r = R-1, \qquad (15)$$

where $\mathbf{q}(i,r)$ is given in (4), $\tilde{\mathbf{q}}(i,r) \equiv \mathbf{q}(i,r)(\mathbf{1}'\mathbf{q}(i,r))^{-1}$, assuming that $\mathbf{1}'\mathbf{q}(i,r) \neq 0$; and $\mathbf{P}(i,r)$ is any $m_r \times (m_r - 1)$ matrix satisfying $\mathbf{P}(i,r)'(\mathbf{1} - \mathbf{k}(i,r)) = \mathbf{0}$, $\mathbf{k}(i,r) \equiv \mathbf{V}(i,r)\tilde{\mathbf{q}}(i,r)V(i,r)^{-1}$, and $V(i,r) \equiv \tilde{\mathbf{q}}(i,r)'\mathbf{V}(i,r)\tilde{\mathbf{q}}(i,r)$. Given that the transformation in (15) is one-to-one and does not depend on $\theta(i,r)$, the joint density of $\{Z(i,r), \mathbf{d}(i,r)\}$ provides identical likelihood inference for $\theta(i,r)$, conditional on $Y(i,r)$. Its advantage over using the conditional density of $\mathbf{Z}(i,r)$ given $Y(i,r)$, follows from the fact that

$$p(Z(i,r), \mathbf{d}(i,r) \mid Y(i,r); \theta(i,r)) = p(\mathbf{d}(i,r) \mid Z(i,r); \theta(i,r))\, p(Z(i,r) \mid Y(i,r)),$$

where $p(Z(i,r) \mid Y(i,r))$ is a Gaussian density with mean $Y(i,r)$ and variance $\sigma^2 V(i,r)$, and $p(\mathbf{d}(i,r) \mid Z(i,r); \theta(i,r))$ is a multivariate Gaussian den-

sity with mean $\mathbf{P}(i,r)'\mathbf{k}(i,r)Z(i,r)$ and variance-covariance matrix

$$\mathbf{P}(i,r)'(\mathbf{Q}(i,r)\mathbf{W}_r^*(\theta(i,r))\mathbf{Q}(i,r)' + \mathbf{V}(i,r) - V(i,r)\mathbf{k}(i,r)\mathbf{k}(i,r)')\mathbf{P}(i,r).$$

That is, (15) factorizes the information content of $\mathbf{Z}(i,r)$ into what is relevant to $\theta(i,r)$, through $\mathbf{d}(i,r)$, and what is relevant to all coarser-resolution θ-parameters, through $Z(i,r)$; $r = R-1$. Note that $\{Y(i,R-1)\}$ is equivalent to $\{\mathbf{Y}(i,R-2)\}$ and hence the second integration of (12) yields a term equivalent to (13) with $r = R-2$. By repeating the integration-factorization process outlined above, until the final integration with respect to $\{Y(i,0)\}_{i=1}^{n_0}$, we obtain the likelihood of $\{\theta(i,r)\}$ to be proportional to

$$\prod_{r=0}^{R-1}\prod_{i=1}^{n_r} p(\mathbf{d}(i,r) \mid Z(i,r);\theta(i,r)), \tag{16}$$

where $\mathbf{d}(i,r)$ and $Z(i,r)$ are obtained from (15), generalized for all $r = R-1,\ldots,0$. The estimation of $\theta(r) \equiv \{\theta(i,r) : i = 1,\ldots,n_r\}$ (or equivalently $\beta(r)$) is then carried out using the resolution-specific likelihood (RESL),

$$L_r^{(d)}(\theta(r)) \equiv \prod_{i=1}^{n_r} p(\mathbf{d}(i,r) \mid Z(i,r);\theta(i,r)) \; ; \quad r = R-1,\ldots,0, \tag{17}$$

resulting in a fast, resolution-specific estimation procedure.

RESL-based estimates of $\{\theta(i,r)\}$ are identical to maximum-likelihood estimates if the transformation in (15) is one-to-one. Kolaczyk and Huang (2001) point out that a necessary and sufficient condition for this is $\mathbf{q}(i,r) = (V(ch(i,r)_1)^{-1},\ldots,V(ch(i,r)_{m_r})^{-1})'$. Generally, this is different from the mass-balance constraint $\mathbf{q}(i,r) = \mathbf{v}(i,r)$, but is the same when the measurement-error variance is inversely proportional to the area of the cell. However, if the transformation in (15) is not one-to-one, the likelihood decompositon in (16) is not exact, and hence the RESL estimates derived using (17) are only approximately ML estimates. We therefore consider an alternative likelihood-type quantity to maximize, namely the resolution-specific restricted likelihood (RESREL).

In place of maximum likelihood estimation of the $\{\theta(i,r)\}$, the fine-resolution data $\{Z(i,R)\}$ and the aggregated data $\{Z(i,r)\}$; $r = R-1,\ldots,0$, can be used to construct a sequence of resolution-specific restricted likelihoods (RESRELs), such that the r-th likelihood is used to estimate $\theta(r)$; $r = R-1,\ldots,0$. Just as for REML, let $\mathbf{E}(r)$ be any $m_r \times (m_r - 1)$

matrix such that $\mathbf{E}(r)'\mathbf{1} = \mathbf{0}$, and define the contrasts,

$$\mathbf{e}(i,r) \equiv \mathbf{E}(r)'\mathbf{Z}(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1}. \tag{18}$$

Then, using (14),

$$\mathbf{e}(i,r) = \mathbf{Q}_e(i,r)\omega^*(i,r) + \nu_e(i,r)\,; \quad (i,r) \in \mathcal{N}_{R-1},$$

where $\mathbf{Q}_e(i,r) \equiv \mathbf{E}(r)'\mathbf{Q}(i,r)$ and $\nu_e(i,r) \equiv \mathbf{E}(r)'\nu(i,r)$. That is,

$$\mathbf{e}(i,r) \sim \text{Gau}(\mathbf{0}, \sigma^2(\mathbf{Q}_e(i,r)\mathbf{W}_r^*(\theta(i,r))\mathbf{Q}_e(i,r)' + \mathbf{V}_e(i,r))),$$

where $\mathbf{V}_e(i,r) \equiv \mathbf{Q}_e(i,r)\mathbf{V}(i,r)\mathbf{Q}(i,r)'$. Note that within each resolution r, the $\{\mathbf{e}(i,r)\}$ are independent. One can then use the resolution-specific restricted likelihood (RESREL),

$$L_r^{(e)}(\theta(r)) \equiv \textstyle\prod_{i=1}^{n_r} p(\mathbf{e}(i,r); \theta(i,r)), \tag{19}$$

for inference on $\theta(r)$, where $p(\mathbf{e}(i,r); \theta(i,r))$ is the Gaussian density associated with $\mathbf{e}(i,r)$; $i = 1, \ldots, n_r$, $r = R-1, \ldots, 0$.

Note that the RESREL is not tied to any particular set of constraining vectors $\{\mathbf{q}(i,r)\}$, as is the case for RESL. However, the choice of $\{\mathbf{q}(i,r)\}$ does determine how the fine-resolution data $\{Z(i,R)\}$ will be aggregated.

Hitherto, we have assumed that σ^2 is known and there is no missing data in $\{Z(i,R)\}$. In the more realistic situation where σ^2 is unknown, one can estimate σ^2 at a fixed resolution, say the finest-resolution (using either RESL or RESREL), and use the resulting σ^2 estimate when estimating $\{\theta(i,r)\}$.

If some of the elements of $\mathbf{Z}(i, R-1)$ are missing (unobserved), it is not possible to decompose $\mathbf{Z}(i, R-1)$ into the two components, $Z(i, R-1)$ and $\mathbf{d}(i, R-1)$ with the right factorization properties needed for RESL. One solution is to ignore those i for which $\mathbf{Z}(i, R-1)$ has any missing elements. A similar strategy can be taken for the RESREL approach.

4. Application: Total Column Ozone (TCO)

Our data consist of spatially and temporally irregular TCO observations sampled on October 2, 1988, by the total ozone mapping spectrometer

(TOMS) instrument on the Nimbus-7 satellite. In a single day, the satellite is able to achieve approximately global coverage, with a slight overlap in consecutive orbits. Under perfect conditions, this generates about 200,000 TCO observations within a single day. In practice, a number of observations are missing and others are removed by a quality-control procedure, resulting in 162,265 valid observations for October 2, 1988. In our analysis of the TCO data, we shall use five spatial resolutions, as in Huang et al. (2002):

Resolution:	R-1	R-2	R-3	R-4	R-5
Cell size (lon × lat):	45°×36°	15°×12°	5°×4°	2.5°×2°	1.25°×1°
Number of cells:	40	360	3,240	12,960	51,840

The TCO data are initially aggregated to the finest resolution, R-5, yielding the (potential) data $\{Z(i,5), V(i,5) : i = 1, \ldots, 51,840\}$, where $Z(i,5)$ is defined as the average of all observations within $D(i,5)$, and $V(i,5)$ is taken to be the reciprocal of the number of observations within $D(i,5)$; $i = 1, \ldots, 51,840$. In our case, 7,382 R-5 cells do not contain any observations, resulting in 7,382 missing observations in the R-5 dataset (Figure 3, top).

To apply the spatial multi-resolution model of Section 2 to the TCO data, the matrices $\{\mathbf{W}(i,r)\}$ and σ^2 need to be estimated. Although the optimal predictor does not depend on σ^2, we need it for prediction variances. We assume that the TCO process follows the mass-balanced, coarse-to-fine-resolution process model (5), with $\{\mathbf{W}^*(i,r)\}$ given by the SPPS in (9) and $\mathbf{C}_0(i,r) = \mathbf{I}$. An exploratory data analysis indicates that most of the between-resolution variation is latitudinal. Based on this, $\{\log\theta(i,r)\}$ is modeled as a smooth function of latitude only, within each resolution r, using a linear combination of B-spline basis functions, as in (10), with 4, 7, 10, and 14 knots at resolutions 1–4, respectively. The B-splines were constrained to have zero derivative at the poles, resulting in a smooth surface on the sphere. Estimation of the parameter vectors $\{\beta(r)\}$ was carried out using both the RESL in (17) and the RESREL in (19), with σ^2 estimated at the finest resolution in each case. Only aggregated data $\{\mathbf{Z}(i,r)\}$ with no missing elements were used in the estimation process. At the coarsest resolution, R-1, recall that $(Y(1,1), \ldots, Y(40,1))' \sim \text{Gau}(\mathbf{a}(0), \sigma^2\mathbf{R}(0))$. We assume that the trend $\mathbf{a}(0)$ is a linear combination of 25 spherical

(a) $\hat{\sigma}\hat{\theta}(i,4)^{1/2}$ (b) $(\widehat{\mathrm{var}}(Y(i,5)))^{1/2}$

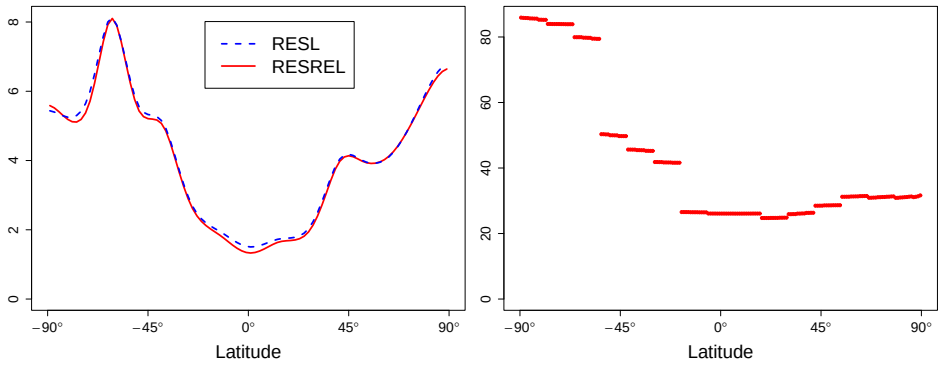

Figure 2. (a) The RESL and RESREL estimates of $\hat{\sigma}\hat{\theta}(i,4)^{1/2}$ as a function of latitude. (b) The estimated standard deviation of $\{Y(i,5)\}$, as a function of latitude, based on RESREL estimation of σ^2 and $\{\theta(i,r)\}$.

harmonics (i.e., $\mathbf{a}(0) = \mathbf{X}\beta(0)$) and $\mathbf{R}(0)$ is given by an exponential covariance function. Unknown parameters of this model were estimated from the coarsest-resolution aggregated data $\{Z(i,0)\}_{i=1}^{40}$ using REML. An alternative approach would be to detrend the original, massive TCO data, as in Johannesson and Cressie (2003).

Figure 2(a) shows both the RESL and the RESREL estimates of $\sigma\theta(i,4)^{1/2}$, plotted versus latitude. We note first that the two estimates are basically identical, both showing that the difference between the aggregated Y-process at R-4 and R-5 has least variability around the equator. Figure 2(b) shows the marginal variance of $\{Y(i,5)\}$, based on the RESREL estimates of σ^2 and $\{\theta(i,r)\}$. The stepwise appearance in Figure 2(b) is due to the change-of-resolution nature of the MRSM. Finally, Figure 3 shows the TCO data $\{Z(i,5)\}$, and the posterior mean and standard deviation given by the MRSM after substituting in RESREL estimates of σ^2 and $\{\theta(i,r)\}$.

The MRSM has enormous advantages, computationally. The program used for the analysis in this paper was written using the statistical programming language R (Ihaka and Gentleman, 1996). The whole execution time of the program, from creating the spatial tree-structure, through to computing the estimates used in Figures 2 and 3, took about 3 minutes on

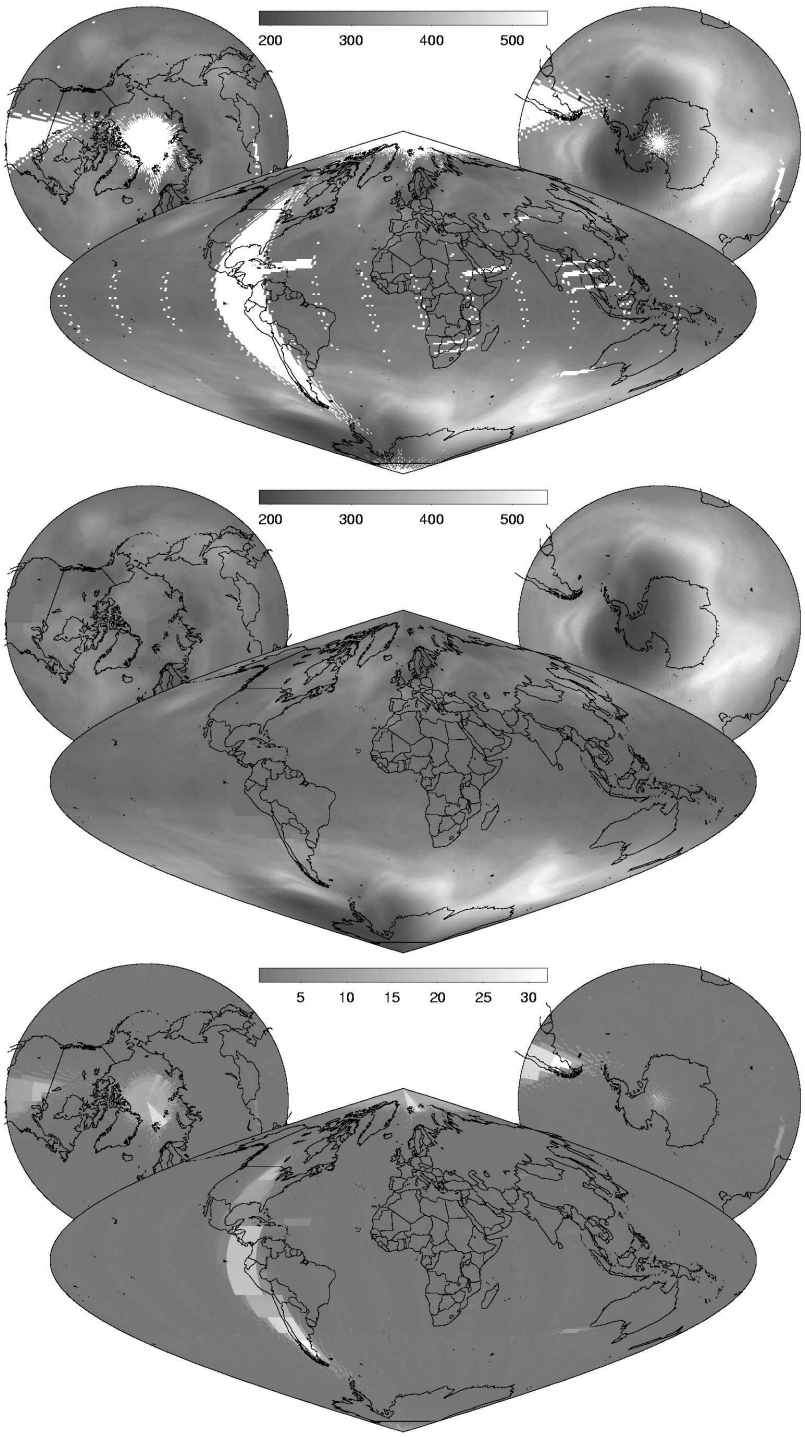

Figure 3. Top: the TCO data at resolution-5 (white denotes missing data). Middle: the posterior mean of the TCO process. Bottom: the posterior standard deviation.

a linux computer with an Atholon MP 1800 processor.

Acknowledgements

This research was supported by the U.S. Environmental Protection Agency under Assistance Agreement R827257-01-0. The authors would like to thank Hsin-Cheng Huang for his helpful comments.

References

Chou, K. C., A. S. Willsky, and R. Nikoukhah: 1994, 'Multiscale systems, Kalman filters, and Riccati equations'. *IEEE Transactions on Automatic Control* **39**, 479–492.

Cressie, N.: 1993, *Statistics for Spatial Data (Revised Edition)*. New York: Wiley.

Cressie, N.: 2002, 'Variogram estimation'. In: A. H. El-Shaarawi and W. W. Piegorsch (eds.): *Encyclopedia of Environmetrics*, Vol. 4. New York: Wiley, pp. 2316–2321.

Huang, H.-C. and N. Cressie: 2001, 'Multiscale graphical modeling in space: application to command and control'. In: M. Moore (ed.): *Spatial Statistics: Methodological Aspects and Some Applications*, Vol. 159 of *Springer Lecture Notes in Statistics*. New York: Springer, pp. 83–113.

Huang, H.-C., N. Cressie, and J. Gabrosek: 2002, 'Fast, resolution-consistent spatial prediction of global processes from satellite data'. *Journal of Computational and Graphical Statistics* **11**, 63–88.

Ihaka, R. and R. Gentleman: 1996, 'R: A Language for Data Analysis and Graphics'. *Journal of Computational and Graphical Statistics* **5**, 299–314.

Johannesson, G.: 2003, 'Multi-resolution statistical modeling in space and time with application to remote sensing of the environment'. Ph.D. thesis, The Ohio State University.

Johannesson, G. and N. Cressie: 2003, 'Finding large-scale spatial trends in massive, global, environmental datasets'. Technical report, Department of Statistics, The Ohio State University.

Kolaczyk, E. D. and H. Huang: 2001, 'Multiscale statistical models for hierarchical spatial aggregation'. *Geographical Analysis* **33**, 95–118.

McCulloch, C. E. and S. R. Searle: 2001, *Generalized, Linear, and Mixed Models*. New York: Wiley.