

Fast Spatial Prediction of Global Processes from Satellite Data

Hsin-Cheng Huang
Institute of Statistical Science
Academia Sinica
Taipei 115
Taiwan

Noel Cressie
Department of Statistics
The Ohio State University
Columbus, OH 43210-1247
U.S.A.

John Gabrosek
Department of Mathematics and Statistics
Grand Valley State University
Allendale, MI 49401
U.S.A.

April 10, 2000

SUMMARY

Polar orbiting satellites remotely sense the earth and its atmosphere, producing data sets that give daily global coverage. For any given day, the data are many and spatially irregular. Our goal in this article is to predict values that are spatially regular at different resolutions; such values are often used as input to general circulation models (GCMs) and the like. Not only do we wish to predict optimally, but because data acquisition is relentless, our algorithm must also process the data very rapidly. This article presents a new statistical prediction methodology that preserves “mass balance” across resolutions and computes spatial predictions and prediction (co)variances extremely fast. Data from the Total Ozone Mapping Spectrometer (TOMS) instrument, on the Nimbus-7 satellite, are used for illustration.

Key words: Change-of-resolution Kalman filter, change of support, EM algorithm, heterogeneous variances, mass balance, multi-resolution, TOMS, tree-structured model.

1 INTRODUCTION

The general statistical problem posed in this article is that of fast, statistically optimal, spatial prediction of global processes based on spatially irregular data. Importantly, the spatial predictions are needed at different spatial resolutions; thus, one of the challenges is to make the predictions and the prediction (co)variances consistent across resolutions. By combining several small regions into a larger region and several larger regions into an even

larger region, and so forth, we build up a scheme for changing resolutions. Then, an acyclic directed graph can be constructed by drawing arrows from larger “parent” regions to smaller “child” regions, which provides a framework for a statistical model that is autoregressive in levels of resolution (Chou *et al.*, 1994; Huang and Cressie, 2000).

This article concentrates on the special case of autoregressive tree-structured models (e.g., Chou *et al.*, 1994; Luettgen and Willsky, 1995a, 1995b; Fieguth and Willsky, 1996), where the optimal spatial prediction procedures are shown to be extremely fast. Imagine a bathtub whose tap is always running. The tub has a capacity of one day of water and we need to keep emptying it so that no overflow occurs. The analogy between water and data is obvious. Using the usual stochastic methods of spatial prediction (e.g., Cressie, 1993, Sect. 5.9.1), is like trying to empty the bathtub with a small bucket.

The approach we take here for optimal spatial prediction is easily able to handle large-to-massive amounts of daily data, without overflow into the next day’s processing time. The problem was motivated by the need to process massive amounts of global data obtained from satellites remotely sensing the earth and its atmosphere. Further, the results are needed at different resolutions to accommodate the requirements of scientists studying regional and global processes. In this article, we develop a mass-balanced, change-of-resolution Kalman filter that is statistically optimal, very fast, and consistent across changes of resolution.

A basic physical requirement of the model is that of “mass balance”. For example, if total column ozone (TCO) is measured in Dobson units per unit volume, then the TCO of a parent region should be equal to the volume-weighted average of its children’s TCO. We explicitly incorporate this requirement into the state equation of our tree-structured model, from which we produce resolution-consistent optimal spatial predictions. Mass balance in one-dimensional nested models for count data is presented by Kolaczyk (1999).

In this article, we demonstrate that autoregressive tree-structured models can also handle missing data or different data sources that are themselves at different levels of spatial resolution (e.g., the problem of combining satellite data, ground-station data, balloon data, and so forth). Moreover, heterogeneous variances due to change of spatial support are ac-

counted for in the model. In the past, these models have been used in situations where only variables at the finest scale are of practical interest. However, being able to find statistically optimal spatial predictors at different levels of resolution is a particularly appealing feature for environmental data, since the required level of resolution often depends on whether the data are to be used in local, regional, or global calculations. For example, consider spatial prediction of TCO, where spatially regular “level-3” data are given at the finest scale of 1 degree latitude by 1.25 degree longitude. Although the predicted value of a larger area (e.g., a 4 degree latitude by 5 degree longitude cell) could be obtained by taking a volume-weighted average of the corresponding finest-scale predicted values, the prediction (co)variances of these larger regions do not follow in a likewise simple manner. The algorithm we present in this article gives both optimal predictions and prediction (co)variances, at differing spatial resolutions, rapidly.

In Section 2, we review general tree-structured models and the associated change-of-resolution Kalman-filter algorithm. The mass-balanced, tree-structured model is introduced in Section 3. Estimation of parameters based on maximum likelihood via the EM algorithm is also given. Section 4 presents an application to TCO satellite data collected from the Total Ozone Mapping Spectrometer (TOMS) instrument. A discussion of the important features of our model, of the associated spatial-prediction algorithm, and of future research problems is provided in Section 5.

2 TREE-STRUCTURED MODELS

One important use of the statistical methodology presented in this article is the rapid processing of satellite data on a global scale. The US National Aeronautic and Space Administration (NASA) is sponsoring a program called the Earth Observing System (EOS), which consists of a series of satellites to measure the chemical and physical processes of the earth and its atmosphere (NASA, 1992). An important part of this project is to produce regularly spaced spatio-temporal (level-3) data on grid cells at various resolutions, based on irregularly sampled (level-2) satellite data.

One simple way to produce daily level-3 data is to use the sample mean (or median) of those level-2 data that fall in a grid cell for that day. However, this approach fails to capture spatial dependencies inherent in the level-2 data; the closer together they are, the more alike they tend to be. In order to incorporate the spatial information, we may regard the level-2 data as being generated from a stochastic process. A general spatial model can be written as:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \eta(\mathbf{s}); \quad \mathbf{s} \in D, \quad (1)$$

where D is a spatial region of interest, $\mu(\cdot)$ is a deterministic mean process, and $\eta(\cdot)$ is a zero-mean spatially colored stochastic process. The goal is to obtain the optimal predictor and the prediction (co)variance of different aggregations of $Y(\cdot)$. At a given resolution and location D_u , where $|D_u| \equiv \int_{D_u} 1 d\mathbf{s} > 0$, we wish to predict:

$$y_u \equiv \frac{1}{|D_u|} \int_{D_u} Y(\mathbf{s}) d\mathbf{s} = \mu_u + \eta_u, \quad (2)$$

based on the noisy data $\mathbf{Z} \equiv (z_{u_1}, \dots, z_{u_N})'$, where μ_u and η_u are defined analogously to y_u ,

$$z_{u_i} \equiv \frac{1}{|D_{u_i}|} \int_{D_{u_i}} Y(\mathbf{s}) d\mathbf{s} + \varepsilon_{u_i},$$

ε_{u_i} represents measurement error, and $|D_{u_i}| > 0$; $i = 1, \dots, N$. If all the first moments and the second moments of the spatial variables are known, the optimal (linear) predictor of y_u is given by

$$\hat{y}_u = \mu_u + \text{cov}(\eta_u, \mathbf{Z}) (\text{var}(\mathbf{Z}))^{-1} (\mathbf{Z} - \boldsymbol{\mu}),$$

where $\boldsymbol{\mu} \equiv (\mu_{u_1}, \dots, \mu_{u_N})'$. When we have massive amounts of data, it may not be computationally feasible to compute \hat{y}_u directly, since we first have to estimate variance and covariance entries, and then we have to take the inverse of an extremely high-dimensional matrix. The difficulties are only exacerbated when many different prediction regions are considered.

There are several problems that are of considerable challenge here. First, the methodology to produce statistically optimal level-3 data should take spatial dependencies into

account. Second, since level-3 data will be used in regional, national, or even global monitoring programs, we have to produce optimal predictions and prediction variances so that no statistical contradictions appear after aggregating or disaggregating. That is, we require a type of “mass balance” based on the simple equation,

$$y_v = \sum_{j=1}^J \frac{|D_{v_j}|}{|D_v|} y_{v_j},$$

where $D_v = D_{v_1} \cup \dots \cup D_{v_J}$ and $\{D_{v_j}\}$ are disjoint. Third, and of equal importance, optimal predictions for massive mounts of data should be produced rapidly. All of these challenges will be dealt with using tree-structured models.

2.1 Hidden Tree-Structured Model

Consider a (multivariate) Gaussian random process $\{\mathbf{y}_u : u \in U\}$ indexed by the nodes of a directed tree (U, E) , where U is the set of nodes, and E is the set of directed edges. Let u_0 denote the root of the tree, and let $pa(u)$ denote the parent of a node $u \in U \setminus \{u_0\}$. We say that u is a leaf of the tree if it has no children. Then

$$E = \{(pa(u), u) : u \in U \setminus \{u_0\}\}.$$

A hidden multivariate Gaussian tree-structured process $\{\mathbf{y}_u : u \in U\}$ on the tree (U, E) is defined as follows. Assume that the Gaussian process evolves from parents to children in an autoregressive manner according to the following:

$$\begin{aligned} \mathbf{y}_u &= \mathbf{A}_u \mathbf{y}_{pa(u)} + \mathbf{w}_u; & u \in U \setminus \{u_0\}, \\ \mathbf{z}_u &= \mathbf{C}_u \mathbf{y}_u + \boldsymbol{\epsilon}_u; & u \in U, \end{aligned} \tag{3}$$

where $\{\mathbf{z}_u\}$ are (potential) observations, $\{\mathbf{y}_u\}$ are the hidden, zero-mean, Gaussian state vectors that are to be predicted, $\{\mathbf{A}_u\}$ and $\{\mathbf{C}_u\}$ are deterministic matrices, $\{\mathbf{w}_u\}$ and $\{\boldsymbol{\epsilon}_u\}$ are independent, zero-mean, Gaussian vectors with covariance matrices,

$$\begin{aligned} \mathbf{W}_u &\equiv \text{var}(\mathbf{w}_u); & u \in U \setminus \{u_0\}, \\ \boldsymbol{\Phi}_u &\equiv \text{var}(\boldsymbol{\epsilon}_u); & u \in U, \end{aligned}$$

$\{\mathbf{y}_u\}$ and $\{\boldsymbol{\epsilon}_u\}$ are independent, and $\mathbf{y}_{pa(u)}$ and \mathbf{w}_u are independent for $u \in U \setminus \{u_0\}$. The goal is to obtain statistically optimal predictors of the state vectors $\{\mathbf{y}_u\}$ based on available data $\{\mathbf{z}_{u_1}, \dots, \mathbf{z}_{u_N}\}$.

Note that the covariance between any two variables \mathbf{y}_u and $\mathbf{y}_{u'}$ on the tree can be computed recursively along the paths from their first common ancestor $an(u, u')$ to u and u' . That is,

$$cov(\mathbf{y}_u, \mathbf{y}_{u'}) = \mathbf{A}_{u_1} \cdots \mathbf{A}_u var(\mathbf{y}_{an(u, u')})(\mathbf{A}_{u'_1} \cdots \mathbf{A}_{u'})', \quad (4)$$

where $(an(u, u'), u_1, \dots, u)$ and $(an(u, u'), u'_1, \dots, u')$ are the paths from $an(u, u')$ to u and u' , respectively.

2.2 Change-of-Resolution Kalman Filter

Chou *et al.* (1994) developed a fast, change-of-resolution Kalman-filter algorithm for tree-structured models; a Bayesian version of this is given below. The algorithm consists of two steps, the leaves-to-root filtering step, followed by the root-to-leaves smoothing step. In the leaves-to-root filtering step, the algorithm goes from the leaves of the tree, recursively computing the optimal predictor of the state vector \mathbf{y}_u based on the data observed at node u and its descendent nodes. Once the root u_0 is reached, the optimal predictor of \mathbf{y}_{u_0} is obtained, based on all the data. In the root-to-leaves smoothing step, the algorithm goes from the root of the tree, recursively computing the optimal predictor of the state vector \mathbf{y}_u at a node u based on all the data. These two steps, though more complicated, are analogous to the filtering step and the smoothing step of the standard Kalman filter in time series. The algorithm has also been extended to more general graphical Markov models by Huang and Cressie (2000).

First, we introduce some notation. Boldface will be used to denote either a vector, a set of vectors, or a matrix. Denote $u \prec u'$ if $u' = u$ or u' is a descendent of u . For $u \in U$, let

$$\begin{aligned} U_u &\equiv \{u' : u \prec u'\}, \\ \boldsymbol{\gamma}_u &\equiv \begin{cases} 1; & \mathbf{z}_u \text{ is observed at } u, \\ 0; & \text{otherwise,} \end{cases} \\ \mathbf{Z} &\equiv \{\mathbf{z}_u : \boldsymbol{\gamma}_u = 1\}, \end{aligned}$$

$$\begin{aligned}
\mathbf{Z}_u &\equiv \{z_{u'} : \gamma_{u'} = 1, u' \in U_u\}, \\
\mathbf{Z}_u^* &\equiv \{z_{u'} : \gamma_{u'} = 1, u' \in U_u \setminus \{u\}\}, \\
\hat{\mathbf{y}}_{u_1|u_2} &\equiv \mathbb{E}(\mathbf{y}_{u_1} | \mathbf{Z}_{u_2}), \\
\hat{\mathbf{y}}_{u_1|u_2}^* &\equiv \mathbb{E}(\mathbf{y}_{u_1} | \mathbf{Z}_{u_2}^*), \\
\hat{\mathbf{y}}_u &\equiv \mathbb{E}(\mathbf{y}_u | \mathbf{Z}), \\
\mathbf{V}_u &\equiv \text{var}(\mathbf{y}_u), \\
\mathbf{\Gamma}_{u_1|u_2} &\equiv \text{var}(\mathbf{y}_{u_1} | \mathbf{Z}_{u_2}), \\
\mathbf{\Gamma}_{u_1|u_2}^* &\equiv \text{var}(\mathbf{y}_{u_1} | \mathbf{Z}_{u_2}^*), \\
\mathbf{\Gamma}_u &\equiv \text{var}(\mathbf{y}_u | \mathbf{Z}), \\
\mathbf{\Gamma}_{u_1, u_2} &\equiv \text{cov}(\mathbf{y}_{u_1}, \mathbf{y}_{u_2} | \mathbf{Z}) = \text{cov}(\hat{\mathbf{y}}_{u_1} - \mathbf{y}_{u_1}, \hat{\mathbf{y}}_{u_2} - \mathbf{y}_{u_2}).
\end{aligned}$$

Also, for $u \in U \setminus \{u_0\}$, let

$$\begin{aligned}
\mathbf{B}_u &\equiv \mathbf{V}_{pa(u)} \mathbf{A}'_u \mathbf{V}_u^{-1}, \\
\mathbf{R}_u &\equiv \mathbf{V}_{pa(u)} - \mathbf{V}_{pa(u)} \mathbf{A}'_u \mathbf{V}_u^{-1} \mathbf{A}_u \mathbf{V}_{pa(u)}.
\end{aligned}$$

Note that $\{\mathbf{V}_u\}$ can be computed recursively; from (3), we have that

$$\mathbf{V}_u = \mathbf{A}_u \mathbf{V}_{pa(u)} \mathbf{A}'_u + \mathbf{W}_u; \quad u \in U \setminus \{u_0\}.$$

In the leaves-to-root filtering step, we start with the leaves and proceed towards the root of the tree, against the directions of the edges. At each node u , $\hat{\mathbf{y}}_{u|u}$ and $\mathbf{\Gamma}_{u|u}$ are obtained recursively. Using conditional distribution results for multivariate Gaussian processes, for a leaf $u \in U$, we have

$$\hat{\mathbf{y}}_{u|u} = \gamma_u \mathbf{V}_u \mathbf{C}'_u (\mathbf{C}_u \mathbf{V}_u \mathbf{C}'_u + \mathbf{\Phi}_u)^{-1} \mathbf{z}_u, \quad (5)$$

$$\mathbf{\Gamma}_{u|u} = \mathbf{V}_u - \gamma_u \mathbf{V}_u \mathbf{C}'_u (\mathbf{C}_u \mathbf{V}_u \mathbf{C}'_u + \mathbf{\Phi}_u)^{-1} \mathbf{C}_u \mathbf{V}_u. \quad (6)$$

For $u \in U$ that is not a leaf, let $\mathbf{ch}(u) \equiv (ch(u, 1), \dots, ch(u, n_u))'$ denote the vector of the children of u , where n_u is the number of children of the node u . We have, for $i = 1, \dots, n_u$,

the following recursions:

$$\hat{\mathbf{y}}_{u|ch(u,i)} = \mathbf{B}_{ch(u,i)} \hat{\mathbf{y}}_{ch(u,i)|ch(u,i)}, \quad (7)$$

$$\mathbf{\Gamma}_{u|ch(u,i)} = \mathbf{B}_{ch(u,i)} \mathbf{\Gamma}_{ch(u,i)|ch(u,i)} \mathbf{B}'_{ch(u,i)} + \mathbf{R}_{ch(u,i)}, \quad (8)$$

$$\hat{\mathbf{y}}_{u|u}^* = \mathbf{\Gamma}_{u|u}^* \left(\sum_{i=1}^{n_u} \mathbf{\Gamma}_{u|ch(u,i)}^{-1} \hat{\mathbf{y}}_{u|ch(u,i)} \right), \quad (9)$$

$$\mathbf{\Gamma}_{u|u}^* = \left\{ \mathbf{V}_u^{-1} + \sum_{i=1}^{n_u} \left(\mathbf{\Gamma}_{u|ch(u,i)}^{-1} - \mathbf{V}_u^{-1} \right) \right\}^{-1}, \quad (10)$$

$$\hat{\mathbf{y}}_{u|u} = \mathbf{\Gamma}_{u|u} \left(\gamma_u \mathbf{C}'_u \mathbf{V}_u^{-1} \mathbf{z}_u + (\mathbf{\Gamma}_{u|u}^*)^{-1} \hat{\mathbf{y}}_{u|u}^* \right), \quad (11)$$

$$\mathbf{\Gamma}_{u|u} = \mathbf{\Gamma}_{u|u}^* - \gamma_u \mathbf{\Gamma}_{u|u}^* \mathbf{C}'_u \left(\mathbf{C}_u \mathbf{\Gamma}_{u|u}^* \mathbf{C}'_u + \mathbf{\Phi}_u \right)^{-1} \mathbf{C}_u \mathbf{\Gamma}_{u|u}^*. \quad (12)$$

At the root u_0 , we have

$$\hat{\mathbf{y}}_{u_0} = \hat{\mathbf{y}}_{u_0|u_0},$$

$$\mathbf{\Gamma}_{u_0} = \mathbf{\Gamma}_{u_0|u_0}.$$

The root-to-leaves smoothing step moves from the root to the leaves in the direction of the edges, such that $\hat{\mathbf{y}}_u$ and $\mathbf{\Gamma}_u$ can be computed recursively, for $u \in U$:

$$\hat{\mathbf{y}}_u = \hat{\mathbf{y}}_{u|u} + \mathbf{\Gamma}_{u|u} \mathbf{B}'_u \mathbf{\Gamma}_{pa(u)|u}^{-1} \left(\hat{\mathbf{y}}_{pa(u)} - \hat{\mathbf{y}}_{pa(u)|u} \right), \quad (13)$$

$$\mathbf{\Gamma}_u = \mathbf{\Gamma}_{u|u} + \mathbf{\Gamma}_{u|u} \mathbf{B}'_u \mathbf{\Gamma}_{pa(u)|u}^{-1} \left(\mathbf{\Gamma}_{pa(u)} - \mathbf{\Gamma}_{pa(u)|u} \right) \mathbf{\Gamma}_{pa(u)|u}^{-1} \mathbf{B}_u \mathbf{\Gamma}_{u|u}. \quad (14)$$

Complete derivation of the algorithm can be found in Chou *et al.* (1994) and via a Bayesian approach by Huang and Cressie (2000).

Luetzgen and Willsky (1995a) show that the prediction errors $\hat{\mathbf{y}}_u - \mathbf{y}_u$; $u \in U$, also follow a multi-resolution tree-structured model. That is,

$$\hat{\mathbf{y}}_u - \mathbf{y}_u = \mathbf{G}_u \left(\hat{\mathbf{y}}_{pa(u)} - \mathbf{y}_{pa(u)} \right) + \boldsymbol{\xi}_u; \quad u \in U \setminus \{u_0\}, \quad (15)$$

where

$$\mathbf{G}_u \equiv \mathbf{\Gamma}_{u|u} \mathbf{B}'_u \mathbf{\Gamma}_{pa(u)|u}^{-1}; \quad u \in U \setminus \{u_0\},$$

$\hat{\mathbf{y}}_{pa(u)} - \mathbf{y}_{pa(u)}$ and $\{\boldsymbol{\xi}_u : u \in U \setminus U_0\}$ are independent, zero-mean, and Gaussian for $u \in U \setminus \{u_0\}$, and

$$\text{var}(\boldsymbol{\xi}_u) = \mathbf{R}_u \equiv \mathbf{V}_{pa(u)} - \mathbf{V}_{pa(u)} \mathbf{A}'_u \mathbf{V}_u^{-1} \mathbf{A}_u \mathbf{V}_{pa(u)}; \quad u \in U \setminus \{u_0\}.$$

Note that \mathbf{G}_u ; $u \in U$, can be computed in the leaves-to-root filtering step. From (4) and (15), for any $u, u' \in U$, we can obtain the prediction covariance between any two variables as:

$$\begin{aligned}\mathbf{\Gamma}_{u,u'} &= \text{cov}(\hat{\mathbf{y}}_u - \mathbf{y}_u, \hat{\mathbf{y}}_{u'} - \mathbf{y}_{u'}) \\ &= \mathbf{G}_{u_1} \cdots \mathbf{G}_u \text{var}(\hat{y}_{an(u,u')} - y_{an(u,u')}) (\mathbf{G}_{u'_1} \cdots \mathbf{G}_{u'})',\end{aligned}$$

where the notation is the same as in eq. (4). In particular, we have

$$\mathbf{\Gamma}_{u,pa(u)} = \mathbf{G}_u \mathbf{\Gamma}_{pa(u)}; \quad u \in U \setminus \{u_0\}. \quad (16)$$

These formulas are needed later in the EM algorithm that is used to obtain maximum likelihood estimators of the model parameters; see Section 3.3.

The change-of-resolution Kalman-filter algorithm presented above does *not* have mass balance. Therefore, although it represents a start, the algorithm does not guarantee aggregation consistency for predictions and prediction variances. Moreover, we have yet to develop a method for model-parameter estimation and a way to incorporate heterogeneous variances. Solutions to these problems are given in the next section.

3 MULTI-RESOLUTION TREE-STRUCTURED MODELS

In the last decade, there has been a lot of research interest in multi-resolution methods, including multi-resolution representations of signals based on wavelet transforms (e.g., Daubechies, 1988; Mallat, 1989; Meyer, 1992), and multi-resolution stochastic models linking coarser-scale variables to finer-scale variables in an autoregressive manner via trees (e.g., Chou *et al.*, 1994; Luetngen and Willsky, 1995a, 1995b; Fieguth and Willsky, 1996). An advantage of using these methods is that many signals naturally have multi-scale features. Moreover, fast implementation algorithms can usually be developed. In this article, we develop a basic multi-resolution tree-structured model and then enhance it to include mass balance and heterogeneous variances. We also describe an EM-algorithm for estimating model parameters.

3.1 Basic Model

Consider a tree with J scales. Assume that there are N_1 nodes at the first scale (the coarsest scale), and each node at the j -th scale has n_j children; $j = 1, \dots, J - 1$. So, there are $N_j = N_1 n_1 \cdots n_{j-1}$ nodes at the j -th scale; $j = 2, \dots, J$. For example, if $N_1 = 1$ and $n_1 = \cdots = n_{J-1} = 4$, we obtain a quadtree; see Figure 1.

Figure 1 here

In all that is to follow, we consider the univariate case, where only one variable y is to be predicted at different locations and different resolutions. In (1) and (2), let the nodes of the tree be locations at the centroid of the corresponding disjoint subregions $\{D_{j,k}\}$, where $\cup_{k=0}^{N_1-1} D_{1,k} = \cdots = \cup_{k=0}^{N_J-1} D_{J,k} = D$ with $|D_{j,1}| = \cdots = |D_{j,N_j-1}| > 0$; $j = 1, \dots, J$. This assumption of equal number of children and equal areas within a resolution is made for simplicity and will be relaxed later, in Section 3.4. Consider a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$. For $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, let

$$y_{j,k} \equiv \frac{1}{|D_{j,k}|} \int_{D_{j,k}} Y(\mathbf{s}) d\mathbf{s}$$

be the hidden state variable we would like to predict at resolution j and location k and, without loss of generality, we denote the n_j children of $y_{j,k}$ by $y_{j+1, kn_j}, \dots, y_{j+1, (k+1)n_j-1}$. Thus, the parent of $y_{j,k}$ becomes $y_{j-1, [k/n_{j-1}]}$ for $j = 2, \dots, J$, $k = 0, \dots, N_j - 1$, where $[x]$ denotes the largest integer less than or equal to x . A multi-resolution, tree-structured model is given as:

$$y_{j,k} = y_{j-1, [k/n_{j-1}]} + w_{j,k}; \quad j = 2, \dots, J, \quad k = 0, \dots, N_j - 1, \quad (17)$$

$$z_{j,k} = y_{j,k} + \varepsilon_{j,k}; \quad j = 1, \dots, J, \quad k = 0, \dots, N_j - 1, \quad (18)$$

where $\{z_{j,k}\}$ are (potential) observations, $\varepsilon_{j,k} \sim N(0, \Phi_{j,k})$; $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, are independent, zero-mean, Gaussian random variables representing measurement errors,

$$y_{1,k} \sim N(0, \sigma_1^2); \quad k = 0, \dots, N_1 - 1, \text{ independently,}$$

$$w_{j,k} \sim N(0, \sigma_j^2); \quad j = 2, \dots, J, \quad k = 0, \dots, N_j - 1, \text{ independently,}$$

$y_{j,k}$ and $\varepsilon_{j,k}$ are independent for $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, and $y_{j-1,k}$ and $w_{j,k}$ are independent for $j = 2, \dots, J$, $k = 0, \dots, N_j - 1$. Note that, for $j = 1, \dots, J$,

$$\text{var}(\mathbf{Y}_j) = \sigma_j^2 \mathbf{I}_{N_j} + \sigma_{j-1}^2 \mathbf{I}_{N_{j-1}} \otimes (\mathbf{1}_{n_{j-1}} \mathbf{1}'_{n_{j-1}}) + \dots + \sigma_1^2 \mathbf{I}_{N_1} \otimes (\mathbf{1}_{n_1 \dots n_{j-1}} \mathbf{1}'_{n_1 \dots n_{j-1}}), \quad (19)$$

where $\mathbf{Y}_j \equiv (y_{j,0}, \dots, y_{j,N_j-1})'$, \mathbf{I}_m is the $m \times m$ identity matrix, and $\mathbf{1}_m$ is the $m \times 1$ vector whose entries are all 1.

Many environmental variables of interest are in per-unit-area or per-unit-volume units. Hence, for physical reasons, we should see the average of all the offspring variables of a parent node at the $(J - 1)$ -th scale to be equal to their parent variable. That is, we should see

$$y_{J-1,k} = \frac{1}{n_{J-1}} \sum_{l=0}^{n_{J-1}-1} y_{J,kn_{J-1}+l}; \quad k = 0, \dots, N_{J-1} - 1.$$

However, this mass-balance equation does not hold for the model (17), because the corresponding $\{w_{j,k}\}$ typically do not add to zero. Thus, in this basic model, only the finest-scale variables are meaningful. In what follows, we propose a simple way to ensure mass balance by imposing a linear constraint on the $\{w_{j,k}\}$.

3.2 Homogeneous, Mass-Balanced, Tree-Structured Models

Define

$$\begin{aligned} \mathbf{ych}_{(j,k)} &\equiv \left(y_{j+1, kn_j}, \dots, y_{j+1, (k+1)n_j-1} \right)'; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1, \\ \mathbf{wch}_{(j,k)} &\equiv \left(w_{j+1, kn_j}, \dots, w_{j+1, (k+1)n_j-1} \right)'; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1, \\ \mathbf{H}_{n_j} &\equiv \mathbf{I}_{n_j} - \frac{1}{n_j} \mathbf{1}_{n_j} \mathbf{1}'_{n_j}; \quad j = 2, \dots, J. \end{aligned}$$

That is, $\mathbf{ych}_{(j,k)}$ and $\mathbf{wch}_{(j,k)}$ consist of all the children of $y_{j,k}$ and $w_{j,k}$, respectively. Then, a multi-resolution, mass-balanced, tree-structured model is given as follows:

$$\mathbf{ych}_{(j,k)} = y_{j,k} \mathbf{1}_{n_j} + \mathbf{wch}_{(j,k)}; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1, \quad (20)$$

$$z_{j,k} = y_{j,k} + \varepsilon_{j,k}; \quad j = 1, \dots, J, \quad k = 0, \dots, N_j - 1, \quad (21)$$

where $\{z_{j,k}\}$ are (potential) observations, $\varepsilon_{j,k} \sim N(0, \Phi_{j,k})$; $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, are independent, zero-mean, Gaussian random variables representing measurement errors,

and

$$\mathbf{w}_{\mathbf{ch}(j,k)} \sim N(\mathbf{0}, \sigma_{j+1}^2 \mathbf{H}_{n_j}); \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1,$$

so that

$$\mathbf{1}'_{n_j} \mathbf{w}_{\mathbf{ch}(j,k)} = 0.$$

Comparing (20) and (21) with (17) and (18), the new feature of the model is that we now have constraints on $\{\mathbf{w}_{\mathbf{ch}(j,k)}\}$, which imply that, for $j = 1, \dots, J-1$, $k = 0, \dots, N_j - 1$,

$$\frac{1}{n_j} \sum_{l=0}^{n_j-1} y_{j, kn_j+l} = \frac{1}{n_j} \mathbf{1}'_{n_j} \mathbf{y}_{\mathbf{ch}(j,k)} = y_{j,k}. \quad (22)$$

This is precisely *mass balance*. Now, not only are the finest-scale variables meaningful, as in the basic model in Section 3.1, but so too are the variables at all the other scales. By treating each $\mathbf{y}_{\mathbf{ch}(j,k)}$ as a single node for $j = 1, \dots, J-1$, $k = 0, \dots, N_j - 1$, the mass-balanced, tree-structured model given by (20) and (21) can be regarded as a vector tree-structured model on tree (U, E) *without* constraints, where

$$\begin{aligned} U &= \{(1, k) : k = 0, \dots, N_1 - 1\} \cup \{\mathbf{ch}(j, k) : j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1\}, \\ E &= \{((1, k), \mathbf{ch}(1, k)) : k = 0, \dots, N_1 - 1\} \\ &\quad \cup \left\{ \left(\mathbf{ch}(j-1, [k/n_{j-1}]), \mathbf{ch}(j, k) \right) : j = 2, \dots, J-1, \quad k = 0, \dots, N_j - 1 \right\}. \end{aligned}$$

Therefore, the change-of-resolution Kalman filter given in Section 2.2 can be applied, and we can obtain meaningful optimal predictors and prediction variances at multi-resolutions simultaneously. It should be noted that the prediction algorithm based on the mass-balanced model sacrifices some computational efficiency, because some scalar divisions are replaced by small matrix inversions at each grid node.

Note that for $j = 1, \dots, J-1$, $k = 0, \dots, N_j - 1$,

$$\text{var}(\mathbf{y}_{\mathbf{ch}(j,k)}) = \text{var}(y_{j,k}) \mathbf{I}_{n_j} + \sigma_{j+1}^2 \mathbf{H}_{n_j}.$$

Therefore, for $j = 1, \dots, J - 1$, $k = 0, \dots, N_j - 1$,

$$\begin{aligned} \left\{ \text{var} \left(\mathbf{y} \mathbf{c} \mathbf{h}_{(j,k)} \right) \right\}^{-1} &= \frac{n_j - 1}{(n_j - 1) \text{var} (y_{j,k}) + n_j \sigma_{j+1}^2} \mathbf{I}_{n_j} \\ &+ \frac{\sigma_{j+1}^2}{\text{var} (y_{j,k}) \left((n_j - 1) \text{var} (y_{j,k}) + n_j \sigma_{j+1}^2 \right)} \mathbf{1}_{n_j} \mathbf{1}'_{n_j}, \end{aligned}$$

has a closed-form expression and can be easily computed as part of the multi-resolution Kalman-filter. Also note that, for $j = 1, \dots, J$, the variance of $\mathbf{Y}_j = (y_{j,0}, \dots, y_{j,N_j-1})'$ is given by

$$\begin{aligned} \text{var}(\mathbf{Y}_j) &= \sigma_j^2 \mathbf{I}_{N_j} + (\sigma_{j-1}^2 - \sigma_j^2/n_{j-1}) \mathbf{I}_{N_{j-1}} \otimes \left(\mathbf{1}_{n_{j-1}} \mathbf{1}'_{n_{j-1}} \right) + \dots \\ &+ (\sigma_1^2 - \sigma_2^2/n_1) \mathbf{I}_{N_1} \otimes \left(\mathbf{1}_{n_1 \dots n_{j-1}} \mathbf{1}'_{n_1 \dots n_{j-1}} \right). \end{aligned} \quad (23)$$

Comparing (19) and (23) for a fixed scale $j \in \{1, \dots, J\}$, it follows that the joint distribution of \mathbf{Y}_j has the same form as that for an unconstrained tree-structured model if $\sigma_{i-1}^2 > \sigma_i^2/n_{i-1}$; $i = 2, \dots, j - 1$. Therefore, if one starts with an unconstrained tree-structured model, there exists a mass-balanced, tree-structured model such that the covariance structures match at the finest resolution, $j = J$.

3.3 Parameter Estimation

The vector of model parameters $\boldsymbol{\theta} \equiv (\sigma_1^2, \dots, \sigma_J^2)'$ can be estimated by maximizing the likelihood function assuming that the measurement-error variances $\{\Phi_{j,k}\}$ are known. In practice, this knowledge comes from information supplied with the measuring device or from independent experiments. Let

$$\begin{aligned} \mathbf{Y} &\equiv (\mathbf{Y}'_1, \dots, \mathbf{Y}'_J)', \\ \mathbf{Z} &\equiv \{z_{j,k} : \gamma_{j,k} = 1\}, \end{aligned}$$

where for $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$,

$$\gamma_{j,k} \equiv \begin{cases} 1, & \text{if } z_{j,k} \text{ is observed,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that \mathbf{Y} is actually determined by a proper subset,

$$\mathbf{Y}^\dagger \equiv \mathbf{Y} \setminus \{y_{j, kn_{j-1}} : j = 2, \dots, J, k = 0, \dots, N_{j-1} - 1\},$$

since we have the constraints given by (22).

From (20) and (21), the complete log-likelihood function based on \mathbf{Y}^\dagger and \mathbf{Z} can be written as

$$\begin{aligned} \log L(\boldsymbol{\theta}; \mathbf{Y}^\dagger, \mathbf{Z}) &= c - \frac{1}{2\Phi_{j,k}} \sum_{\gamma_k=1} (z_{j,k} - y_{j,k})^2 - \frac{N_1}{2} \log \sigma_1^2 - \frac{1}{2\sigma_1^2} \sum_{k=0}^{N_1-1} (y_{1,k})^2 \\ &\quad - \sum_{j=1}^{J-1} \sum_{k=0}^{N_j-1} \left\{ \frac{n_j-1}{2} \log \sigma_{j+1}^2 + \frac{1}{2\sigma_{j+1}^2} \left(\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger - y_{j,k} \mathbf{1}_{n_{j-1}} \right)' \right. \\ &\quad \left. \times \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \left(\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger - y_{j,k} \mathbf{1}_{n_{j-1}} \right) \right\}, \end{aligned} \quad (24)$$

where c is a constant,

$$\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger \equiv \left(y_{j+1, kn_{j+1}}, \dots, y_{j+1, (k+1)n_{j-1}} \right)'; \quad j = 1, \dots, J-1, k = 0, \dots, N_j - 1,$$

and

$$\mathbf{H}_{n_j}^\dagger \equiv \mathbf{I}_{n_{j-1}} - \frac{1}{n_j} \mathbf{1}_{n_{j-1}} \mathbf{1}'_{n_{j-1}}; \quad j = 2, \dots, J.$$

Although the “incomplete” log-likelihood based only on \mathbf{Z} is Gaussian, it has a complicated covariance matrix and is difficult to maximize directly. Instead, we apply the EM (Expectation-Maximization) algorithm (e.g., Dempster *et al.*, 1977) and treat \mathbf{Y}^\dagger as missing data. The EM algorithm is an iterative procedure starting with some initial estimator $\hat{\boldsymbol{\theta}}^{(0)}$. Each iteration consists of two steps, the expectation step (E-step) followed by the maximization step (M-step). At the i -th iteration, we evaluate

$$E_{i-1} \left(\log L(\boldsymbol{\theta}; \mathbf{Y}^\dagger, \mathbf{Z}) \mid \mathbf{Z} \right), \quad (25)$$

in the E-step, where E_{i-1} denotes the conditional expectation based on the parameter $\hat{\boldsymbol{\theta}}^{(i-1)}$ obtained from the $(i-1)$ -th iteration. We then find the $\hat{\boldsymbol{\theta}}^{(i)}$ that maximizes (25) in the M-step. The procedure is repeated until convergence. It has been shown that the likelihood

always increases at each iteration and the algorithm is guaranteed to converge under mild conditions (see Dempster *et al.*, 1977; Wu, 1983; Boyles, 1983).

For $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, let

$$\begin{aligned}\hat{y}_{j,k} &\equiv E(y_{j,k} | \mathbf{Z}), \\ \Gamma_{j,k} &\equiv \text{var}(y_{j,k} | \mathbf{Z}),\end{aligned}$$

and for $j = 1, \dots, J - 1$, $k = 0, \dots, N_j - 1$, let

$$\begin{aligned}\hat{\mathbf{y}}_{\mathbf{ch}(j,k)}^\dagger &\equiv E(\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger | \mathbf{Z}), \\ \mathbf{\Gamma}_{\mathbf{ch}(j,k)}^\dagger &\equiv \text{var}(\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger | \mathbf{Z}), \\ \mathbf{\Gamma}_{j+1,j,k}^\dagger &\equiv \text{cov}(\mathbf{y}_{\mathbf{ch}(j,k)}^\dagger, y_{j,k} | \mathbf{Z}).\end{aligned}$$

Upon taking conditional expectations in (25), conditional on the observed data \mathbf{Z} , we have in the E-step,

$$\begin{aligned}& E\left(\log L(\boldsymbol{\theta}; \mathbf{Y}^\dagger, \mathbf{Z}) | \mathbf{Z}\right) \\ &= c - \frac{1}{2\Phi_{j,k}} \sum_{\gamma_k=1} \left\{ (z_{J,k} - \hat{y}_{J,k})^2 + \Gamma_{J,k} \right\} - \frac{N_1}{2} \log \sigma_1^2 - \frac{1}{2\sigma_1^2} \sum_{k=0}^{N_1-1} \left\{ \Gamma_{1,k} + (\hat{y}_{1,k})^2 \right\} \\ &\quad - \sum_{j=1}^{J-1} \sum_{k=0}^{N_j-1} \left\{ \frac{(n_j - 1)}{2} \log \sigma_{j+1}^2 + \frac{1}{2\sigma_{j+1}^2} \left(\hat{\mathbf{y}}_{\mathbf{ch}(j,k)}^\dagger - \hat{y}_{j,k} \mathbf{1}_{n_j-1} \right)' \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \right. \\ &\quad \quad \times \left(\hat{\mathbf{y}}_{\mathbf{ch}(j,k)}^\dagger - \hat{y}_{j,k} \mathbf{1}_{n_j-1} \right) + \frac{1}{2\sigma_{j+1}^2} \text{tr} \left(\left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{\Gamma}_{\mathbf{ch}(j,k)}^\dagger \right) \\ &\quad \quad \left. - \frac{1}{\sigma_{j+1}^2} \mathbf{1}'_{n_j-1} \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{\Gamma}_{j+1,j,k}^\dagger + \frac{\Gamma_{j,k}}{2\sigma_{j+1}^2} \mathbf{1}'_{n_j-1} \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{1}_{n_j-1} \right\}.\end{aligned}$$

Note that at the i -th iteration, the expectations that are in the definitions of the terms $\{\hat{y}_{j,k}\}$, $\{\Gamma_{1,k}\}$, $\{\mathbf{\Gamma}_{\mathbf{ch}(j,k)}^\dagger\}$, and $\{\mathbf{\Gamma}_{j+1,j,k}^\dagger\}$ assume the parameter value $\hat{\boldsymbol{\theta}}^{(i-1)}$, and they are computed using the change-of-resolution Kalman-filter algorithm given by (5)-(14) and (16).

Now consider the M-step. It is not difficult to see that $E(\log L(\boldsymbol{\theta}; \mathbf{Y}^\dagger, \mathbf{Z}) | \mathbf{Z})$ achieves its maximum at

$$\hat{\sigma}_1^2 = \frac{1}{N_1} \sum_{k=0}^{N_1-1} \left\{ \Gamma_{1,k} + (\hat{y}_{1,k})^2 \right\},$$

$$\begin{aligned} \hat{\sigma}_{j+1}^2 = & \frac{1}{N_j(n_j - 1)} \sum_{k=0}^{N_j-1} \left\{ \left(\hat{\mathbf{y}}_{\mathbf{ch}(j,k)}^\dagger - \hat{y}_{j,k} \mathbf{1}_{n_j-1} \right)' \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \left(\hat{\mathbf{y}}_{\mathbf{ch}(j,k)}^\dagger - \hat{y}_{j,k} \mathbf{1}_{n_j-1} \right) \right. \\ & + \text{tr} \left(\left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{\Gamma}_{\mathbf{ch}(j,k)}^\dagger \right) - 2 \mathbf{1}'_{n_j-1} \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{\Gamma}_{j+1,j,k}^\dagger \\ & \left. + \mathbf{\Gamma}_{j,k} \mathbf{1}'_{n_j-1} \left(\mathbf{H}_{n_j}^\dagger \right)^{-1} \mathbf{1}_{n_j-1} \right\}; \quad j = 1, \dots, J-1. \end{aligned}$$

At the i -th iteration, this yields parameter estimate $\hat{\boldsymbol{\theta}}^{(i)}$. Now return to the E-step and repeat until convergence.

3.4 Heterogeneous, Mass-Balanced, Tree-Structured Models

In this section, we shall construct tree-structured models whose variances within a given resolution may be heterogeneous. This may happen because the number of children may differ from node to node or the children's spatial supports may differ. First, we introduce some definitions and notation. Let $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ be a spatial process defined on a spatial region of interest D , with $|D| > 0$.

Definition 1 *A collection of subsets $\{D_{j,k} \subset D : j = 1, \dots, J, k = 0, \dots, N_j - 1\}$ is called a nested partitioning on D , where $|D| > 0$, if the following conditions hold:*

- (i) $|D_{j,k}| > 0, \quad j = 1, \dots, J, k = 0, \dots, N_j - 1;$
- (ii) $\{D_{j,k} : k = 0, \dots, N_j - 1\}$ are disjoint, and $\bigcup_{k=0}^{N_j-1} D_{j,k} = D$, for each $j = 1, \dots, J;$
- (iii) Given any $D_{j,k}; j = 2, \dots, J, k = 0, \dots, N_j - 1$, there exists a $k' \in \{0, \dots, N_{j-1} - 1\}$ such that $D_{j,k} \subset D_{j-1,k'}$. We denote $(j-1, k') = pa((j, k))$.

Note that given a nested partition $\{D_{j,k} \subset D : j = 1, \dots, J, k = 0, \dots, N_j - 1\}$, one can define an associated tree with nodes $\{(j, k) : j = 1, \dots, J, k = 0, \dots, N_j - 1\}$, and directed edges

$$E \equiv \{(pa((j, k)), (j, k)) : j = 2, \dots, J, k = 0, \dots, N_j - 1\}.$$

For $j = 1, \dots, J, k = 0, \dots, N_j - 1$, let

$$y_{j,k} \equiv \frac{1}{a_{j,k}} \int_{D_{j,k}} Y(\mathbf{s}) d\mathbf{s},$$

where $a_{j,k} \equiv |D_{j,k}|$ denotes the cell areas, and $n_{j,k}$ is the number of children of $y_{j,k}$. Further, let

$$\mathbf{y}_{ch(j,k)} \equiv \left(y_{ch(j,k,1)}, \dots, y_{ch(j,k,n_{j,k})} \right)'; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1,$$

denote the vector of values associated with the children of $y_{j,k}$, let

$$\mathbf{a}_{ch(j,k)} \equiv \left(a_{ch(j,k,1)}, \dots, a_{ch(j,k,n_{j,k})} \right)'; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1,$$

denote the vector of the children's cell areas, define $V_{j,k} \equiv \text{var}(y_{j,k})$; $j = 1, \dots, J$, $k = 0, \dots, N_1 - 1$, and define $V_{ch(j,k,l)} \equiv \text{var}(y_{ch(j,k,l)})$; $j = 1, \dots, J-1$, $k = 0, \dots, N_1 - 1$, $l = 1, \dots, n_{j,k}$.

A *heterogeneous*, mass-balanced, tree-structured model is defined as:

$$\mathbf{y}_{ch(j,k)} = y_{j,k} \mathbf{1}_{n_{j,k}} + \mathbf{w}_{ch(j,k)}; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1, \quad (26)$$

$$z_{j,k} = y_{j,k} + \varepsilon_{j,k}; \quad j = 1, \dots, J, \quad k = 0, \dots, N_j - 1, \quad (27)$$

where $\{z_{j,k}\}$ are (potential) observations, $\varepsilon_{j,k} \sim N(0, \Phi_{j,k})$; $j = 1, \dots, J$, $k = 0, \dots, N_j - 1$, are independent, zero-mean, Gaussian random variables representing measurement errors, and

$$\mathbf{w}_{ch(j,k)} \equiv (w_{ch(j,k,1)}, \dots, w_{ch(j,k,n_{j,k})})' \sim N(\mathbf{0}, \mathbf{W}_{ch(j,k)}),$$

with $\mathbf{W}_{ch(j,k)}$ obtained from eq. (31) in the Appendix, by substituting $n_{j,k}$, $\mathbf{a}_{ch(j,k)}$, and $(V_{ch(j,k,1)} - V_{j,k}, \dots, V_{ch(j,k,n_{j,k})} - V_{j,k})'$ for n , \mathbf{a} , and $(\sigma_1^2, \dots, \sigma_n^2)'$, respectively. Hence,

$$\mathbf{a}'_{ch(j,k)} \mathbf{w}_{ch(j,k)} = 0; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1. \quad (28)$$

From (26) and (28), we obtain the mass balance:

$$\mathbf{a}'_{ch(j,k)} \mathbf{y}_{ch(j,k)} = a_{j,k} y_{j,k}; \quad j = 1, \dots, J-1, \quad k = 0, \dots, N_j - 1.$$

That is, the whole is equal to the sum of its parts.

Note that for $j = 1, \dots, J$, if $n_j = n_{j,0} = \dots = n_{j,N_j-1}$, then the variance of $\mathbf{Y}_j = (y_{j,0}, \dots, y_{j,N_j-1})'$ is given by

$$\begin{aligned} \text{var}(\mathbf{Y}_j) &= \mathcal{W}_j + \mathcal{W}_{j-1} \otimes (\mathbf{1}_{n_j} \mathbf{1}'_{n_j}) + \dots + \mathcal{W}_1 \otimes (\mathbf{1}_{n_2 \dots n_{j-1}} \mathbf{1}'_{n_2 \dots n_{j-1}}) \\ &\quad + \sigma_1^2 \mathbf{I}_{N_1} \otimes (\mathbf{1}_{n_1 \dots n_{j-1}} \mathbf{1}'_{n_1 \dots n_{j-1}}), \end{aligned} \quad (29)$$

where $\mathcal{W}_1 \equiv \mathbf{W}_1$, and

$$\mathcal{W}_j \equiv \begin{pmatrix} \mathbf{W}_{ch(j-1,0)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{W}_{ch(j-1,N_{j-1}-1)} \end{pmatrix}; \quad j = 2, \dots, J-1.$$

Also note that it is not always possible to achieve mass balance with the statistical model (26) and (27) based on given variance parameters $\{V_{j,k}\}$. This is as it should be, since it is a warning that the parent-child relationship in (26) is not reasonable, given a large heterogeneity of variances. However, for multi-resolution models whose areas and numbers of children are homogeneous within each resolution, the heterogeneous, mass-balanced, tree-structured model given by (26) and (27) becomes a homogeneous, mass-balanced, tree-structured model given by (20) and (21), which is well defined, and (29) has the same form as (23).

4 TOTAL COLUMN OZONE OVER THE GLOBE

The problem of measuring total column ozone (TCO) has been of interest to scientists for decades. Ozone depletion results in an increased transmission of ultraviolet radiation (290-400 nm wavelength) through the atmosphere. This is mostly deleterious due to damage to DNA and cellular proteins that are involved in biochemical processes, affecting growth and reproduction.

Relatively few measurements of TCO were taken in the first quarter of the twentieth century; however, with the invention of the Dobson spectrophotometer, researchers gained the ability to measure efficiently and accurately TCO abundance (London, 1985). A system of ground-based stations has provided important TCO measurements for the past 40 years; however, the ground-based stations are relatively few in number and provide poor geographic

coverage of the earth. The advent of polar-orbiting satellites has dramatically enhanced the spatial coverage of measurements of TCO.

The Nimbus-7 polar-orbiting satellite was launched on October 24, 1978 with the Total Ozone Mapping Spectrometer (TOMS) instrument aboard. The TOMS instrument scans in three-degree steps to an extreme of 51 degrees on each side of nadir, in a direction perpendicular to the orbital plane (McPeters *et al.*, 1996). Each scan takes roughly eight seconds to complete, including one second for retrace (Madrid, 1978). The altitude of the satellite and scanning pattern of the TOMS instrument are such that consecutive orbits overlap, with the area of overlap depending on the latitude of the measurement. The TOMS instrument covers the entire globe in a 24-hour period. NASA receives the data, calibrates it (level 1), and pre-processes it to yield spatially and temporally irregular TCO measurements (level 2). The level-2 data are subsequently processed to yield a spatially and temporally uniform level-3 data product that is released widely to the scientific community. The level-3 data product uses 1 degree latitude by 1.25 degree longitude ($1^\circ \times 1.25^\circ$) equiangular grid cells.

There are several approaches that have been or can be used to handle large volumes of polar-orbiting satellite data. Fang and Stein (1998) use a moving average with seasonal dependence to investigate variations in zonal ozone levels for a fixed latitude. Niu and Tiao (1995) introduce a class of space-time regression models for analysis at a fixed latitude. Both papers use NASA's level-3 data product based on the TOMS instrument. Zeng and Levy (1995) propose a three-dimensional interpolation technique to fill in missing values for grid-cell locations at certain time points. Other possible approaches are geostatistical (e.g., Cressie, 1993, Ch. 3), although the disadvantage of kriging is that it does not handle large volumes of data well.

Level-2 TCO values and NASA's level-3 data product based on the TOMS instrument were obtained from the Ozone Processing Team of NASA/Goddard, Distributed Active Archive Center, and were stored in Hierarchical Data Format as developed by the National Center for Supercomputing Applications at the University of Illinois. Also, ground-station

data (Section 4.2) were obtained from the World Ozone Data Center, Downsview, Ontario, to provide a standard against which to compare different level-3 data products.

Dobson and co-workers showed the dependence of TCO on latitude almost 70 years ago (Dobson *et al.*, 1929). Figure 2 shows the latitude dependence of ozone zonal means using the TOMS data for October 1, 1988 (Gabrosek *et al.*, 1999). To obtain the figure, we computed the median of all level-2 data from that day, that fell within a $1^\circ \times 1.25^\circ$ grid cell, and we repeated the calculation for all 180×288 such grid cells. Then all grid-cell values at a given degree of latitude were averaged to produce a quantity we call the TCO zonal mean. Throughout the article, we use the convention that negative latitudes correspond to the Southern Hemisphere and positive latitudes to the Northern Hemisphere.

Figure 2 here

Recall that our goal is to produce a level-3 data product for all $1^\circ \times 1.25^\circ$ grid cells, on a daily basis, from the spatially irregular level-2 data referred to above. Based on the development in Section 2 and 3, we derive optimal spatial predictions of TCO using a heterogeneous, mass-balanced, tree-structured model; see Section 4.1. In Section 4.2, we apply our methodology to the TOMS data for October 1, 1988. Eventually, we shall apply this methodology to level-2 data from NASA’s Earth Observing System (EOS). Just one EOS instrument, the Multi-angle Imaging SpectroRadiometer (MISR) will generate roughly 80 gigabytes of data per day (Kahn, 1996), and its vehicle, the Terra satellite, has multiple instruments that will generate data equivalent to all the information stored in the library of Congress every seven weeks for at least six years (Kahn, 1998).

4.1 *Mass-Balanced, Tree-Structured Models for TCO*

Our spatial analysis of the TOMS data proceeds on the spatially irregular, zonal-mean-corrected TCO values:

$$\text{level-2 residual TCO} = \text{level-2 TCO} - \text{zonal mean}, \tag{30}$$

on a given day (here, October 1, 1988). This correction allows us to assume that (30) has zero mean, which is an important component of the tree-structured models given in Section 3.

Using the notation of Section 3, consider a multi-resolution tree structure with $J = 5$ resolutions. There are $N_1 = 40$ nodes at the first resolution, $n_1 = 9$ children of each of these nodes at the second resolution, $n_2 = 9$ children of each of these nodes at the third resolution, $n_3 = 4$ children of each of these nodes at the fourth resolution, and $n_4 = 4$ children of each of these nodes at the fifth and finest resolution. Thus, $N_2 = 360$, $N_3 = 3240$, $N_4 = 12960$, and $N_5 = 51840$. We use equiangular grid cells for all five scales. For each scale $j = 1, \dots, 5$, the grid cells $(j, 0), \dots, (j, N_j - 1)$ are defined according to the lexicographic order of longitude-latitude pairs. Specifically, grid cell (j, k) is defined to be between longitudes $i_{j,k}$ and $i_{j,k} + 45(N_j/N_1)^{-1/2}$ and between latitudes $l_{j,k}$ and $l_{j,k} + 36(N_j/N_1)^{-1/2}$, where

$$\begin{aligned} i_{j,k} &\equiv 45(N_j/N_1)^{1/2} \left[k / (5(N_j/N_1)^{1/2}) \right] - 180; & j = 1, \dots, 5, \quad k = 0, \dots, N_j - 1, \\ l_{j,k} &\equiv 36(N_j/N_1)^{1/2} \left(k - \left[k / (5(N_j/N_1)^{1/2}) \right] \right) - 90; & j = 1, \dots, 5, \quad k = 0, \dots, N_j - 1, \end{aligned}$$

$[x]$ denotes the largest integer less than or equal to x , and note that $5(N_j/N_1)^{1/2}$ is the number of grid cells for a given longitude at scale $j = 1, \dots, 5$. Therefore, for scale $j = 1, \dots, 5$, a consecutive sequence of grid cells starting from the south pole and finishing at the north pole is given by $(j, 0), \dots, (j, 5(N_j/N_1)^{1/2} - 1)$. In particular, $(j, [5(N_j/N_1)^{1/2}/2])$ is a cell closest to the equator at the j -th scale; $j = 1, \dots, 5$. First, the level-2 residual ozone values are computed based on (30). For each $1^\circ \times 1.25^\circ$ grid cell $(5, k)$, let $\mathbf{Z}_{5,k}$ denote the vector of all the level-2 residual ozone data falling in that grid cell, and let $m_{5,k} \equiv |\mathbf{Z}_{5,k}|$ be the dimension of $\mathbf{Z}_{5,k}$; $k = 0, \dots, 51839$. To start the change-of-resolution Kalman filter, we need preliminary data at the finest resolution. These are obtained from generalized-least-squares estimators. The correlation matrix of $\mathbf{Z}_{5,k}$, denoted by $cor(\mathbf{Z}_{5,k})$, is determined by the proportions of areal overlaps among the level-2 observations within each grid cell $(5, k)$; $k = 0, \dots, 51839$. Hence, the preliminary data are,

$$z_{5,k} \equiv \frac{\mathbf{1}'_{m_{5,k}} (cor(\mathbf{Z}_{5,k}))^{-1} \mathbf{Z}_{5,k}}{\mathbf{1}'_{m_{5,k}} (cor(\mathbf{Z}_{5,k}))^{-1} \mathbf{1}_{m_{5,k}}}; \quad k = 0, \dots, 51839.$$

An estimated variance of the measurement error for a level-2 residual ozone datum falling in the grid cell $(5, k)$ is given by,

$$\phi_{5,k} = \frac{(\mathbf{Z}_{5,k} - z_{5,k} \mathbf{1}_{m_{5,k}})' (\text{cor}(\mathbf{Z}_{5,k}))^{-1} (\mathbf{Z}_{5,k} - z_{5,k} \mathbf{1}_{m_{5,k}})}{m_{5,k} - 1}; \quad k = 0, \dots, 51839, \quad m_{5,k} > 1.$$

Figure 3 shows the preliminary data $\{z_{5,k}\}$, based on a Mercator projection of the globe. Notice that there are missing values for those data, corresponding to grid cells within which no level-2 observation fell on that day. Also, $\{\phi_{5,k}\}$ are not defined for grid cells within which no or one level-2 observation falls (e.g., for latitudes within 10 degrees of a pole, there are frequently very few observations, because the TOMS instrument requires sunlight to take readings). Given the zonal dependence of TCO, we pool $\{\phi_{5,k}\}$ for each grid cell within a 1 degree latitude band (from latitude l to latitude $l + 1$) according to the weights $\{m_{5,k} - 1\}$, yielding a latitude-band pooled variance estimate ν_l ; $l = -90, \dots, 89$. As a result, the variance of the measurement error for $z_{5,k}$ can be estimated by

$$\hat{\Phi}_{5,k} = \nu_{l_{5,k}} \left(\mathbf{1}'_{m_{5,k}} \text{cor}(\mathbf{Z}_{5,k})^{-1} \mathbf{1}_{m_{5,k}} \right)^{-1}; \quad k = 0, \dots, 51839.$$

Figure 3 here

For optimal spatial predictions, we shall apply the heterogeneous, mass-balanced, tree-structured model of Section 3.4. For parameter estimation, we shall fit first a homogeneous, mass-balanced, tree-structured model using only the data from latitude -18 to latitude 18, since we have almost equal-area partitions in this region; notice that the ratio of the smallest grid-cell area (corresponding to latitudes $\pm 18^\circ$) to the largest cell area (corresponding to latitudes $\pm 1^\circ$) at the finest resolution is 0.954. The vector of parameters $\boldsymbol{\theta} \equiv (\sigma_1^2, \dots, \sigma_5^2)'$ is estimated by maximum likelihood using the EM-algorithm as described in Section 3.3. The resulting estimate is $\hat{\boldsymbol{\theta}} \equiv (\hat{\sigma}_1^2, \dots, \hat{\sigma}_5^2)' = (37.47, 9.48, 8.21, 3.78, 1.48)'$. Now, from (23), for each scale $j = 1, \dots, 5$, we can obtain the variance of the spatial variables closest to the equator as:

$$\text{var}(y_{j,k}) \equiv V_{j,k} = \sigma_1^2 + \left(1 - \frac{1}{n_1}\right) \sigma_2^2 + \dots + \left(1 - \frac{1}{n_{j-1}}\right) \sigma_j^2; \quad k = [5(N_j/N_1)^{1/2}/2],$$

where $(j, \lceil 5(N_j/N_1)^{1/2}/2 \rceil)$ is a grid cell closest to the equator at the j -th scale for $j = 1, \dots, 5$. Thus $\mathbf{V} \equiv (V_{1,2}, V_{2,7}, V_{3,22}, V_{4,45}, V_{5,90})'$ can be estimated with $\hat{\mathbf{V}}$, obtained by substituting $\hat{\sigma}_1^2, \dots, \hat{\sigma}_5^2$ into the expression above. Then the areas of these grid cells (5, 90), (4, 45), (3, 22), (2, 7), and (1, 2), are used as knots in a spline estimate of all such variances, $\{V_{j,k} : j = 1, \dots, 5, k = 0, \dots, N_j - 1\}$. Specifically, we model $\{V_{j,k}\}$ as a first-order polynomial spline f on $[0, \infty)$ taking values $\hat{V}_{5,90} > \hat{V}_{4,45} > \hat{V}_{3,22} > \hat{V}_{2,7} > \hat{V}_{1,2}$ at knots $a_{5,90} < a_{4,45} < a_{3,22} < a_{2,7} < a_{1,2}$, respectively (see Figure 4). Thus f is nonincreasing and we obtain the estimate,

$$\hat{V}_{j,k} = f(a_{j,k}); \quad j = 1, \dots, 5, \quad k = 0, \dots, N_j - 1.$$

Figure 4 here

4.2 *Optimal Spatial Prediction of TCO*

We can now apply the heterogeneous, mass-balanced, tree-structured model given by (26) and (27) to $\{z_{5,k}\}$ with measurement errors $\{\Phi_{5,k}\}$, hidden state variances $\{\mathbf{W}_{\mathbf{ch}(j,k)}\}$, and estimates substituted for unknown parameters. The predicted residual TCO values based on the mass-balanced change-of-resolution Kalman filter are shown in Figure 5 with the corresponding prediction standard errors shown in Figure 6, based on a Mercator projection of the globe.

Figures 5-6 here

The final predicted value (i.e., level-3 datum) is calculated as:

$$\text{level-3 (Kalman filter) TCO value} = \text{zonal mean} + \text{predicted residual TCO value.}$$

Figures 7 through 11 show the level-3 (Kalman filter) TCO values from resolution 1 (the coarsest resolution) through resolution 5 (the finest resolution), respectively, again based

on a Mercator projection of the globe. Note that the optimal predictions for all five resolutions are obtained simultaneously in one pass using the change-of-resolution Kalman-filter algorithm.

Figures 7-11 here

We shall now compare the level-3 predictions to 80 ground-station observations recorded on the same day. Let L3 denote a generic level-3 data product given at all $1^\circ \times 1.25^\circ$ grid cells for October 1, 1988. Then the mean squared error (MSE) for L3 is calculated as:

$$MSE = \frac{1}{80} \sum_{l=1}^{80} (GS(l) - L3(l))^2,$$

where $GS(l)$ represents the TCO reading for the l -th ground station for October 1, 1988. The NASA level-3 data product achieves a MSE of 146.08, compared to 135.95 for the level-3 data product from the heterogeneous, mass-balanced, tree-structured model. A small reduction (6.9%) in the MSE should be noted, although one should not read too much into this as the ground stations have very spotty global coverage. The important advantages of the methodology based on mass-balanced, tree-structured models is that it provides optimal predictions at multiple resolutions, and associated prediction standard errors.

5 DISCUSSION

We have presented a new methodology for fast spatial prediction that allows us to handle massive amounts of satellite data efficiently, even when they are sampled irregularly. It is based on a spatial model that is autoregressive in scale and for which “mass balance” is preserved. The important advantages of our methodology are first that it provides optimal spatial predictions at multiple resolutions, and associated prediction standard errors. Second, the mass balance guarantees consistent predictors and prediction variances as resolution requirements change, according to whether predictions are to be used in local,

regional, or global calculations. This property also allows us to incorporate data at different levels of resolution. Third, and by no means least, our spatial-prediction algorithms are extremely fast.

One drawback of Kalman filtering on trees is that the implied spatial covariance function is piecewise constant and nonstationary (see eq. (23) and eq. (29)), which can lead to predictions that are not shift invariant. A possible solution to this problem is to compute the spatial predictor as an average over a number of mass-balanced, tree-structured models with different tree branches that represent children shifted to have different parents. Of course, the prediction variances and covariances will be considerably more complicated and the computational complexity will increase with the number of trees used.

In practice, the number of resolutions and the number of children for a given parent have to be specified in advance. Such choices will depend on applications and will in general lead to different small-scale structure of the parametric covariance functions. However, for multi-resolution models like those used for mapping total column ozone over the globe, the overall covariance shape is approximately stationary and exponential, regardless of what the small-scale structure might be. This stability of larger-scale dependence, combined with the quantity of data typically available, lead to spatial predictions that should be quite robust to these choices; capturing the key parameters in the spatial covariance function is most important.

The methodology we have developed should be extendible to incorporate temporal dependence. We propose to model the temporal dependence at resolution 1 as a multivariate autoregressive process in time and to retain the tree structure at each time point, which as a whole again yields a tree structure. We could then run time backwards, from the resolution-1 nodes at time t to the corresponding nodes at time $t - 1$, in the leaves-to-root filtering step. This model, and the possibility of optimal spatio-temporal prediction using a similar Kalman-filter methodology, will be investigated in the future.

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research under grant N00014-99-1-0214 and the Environmental Protection Agency under grant R827257-01-0. Computing was partially supported by an NSF SCREMS grant DMS-9707740 awarded to the Department of Statistics, Iowa State University.

APPENDIX: CONSTRUCTION OF A COVARIANCE MATRIX FOR MASS BALANCE

We prove a technical result that can be used to ensure (28) for heterogeneous, mass-balanced, tree-structured models. For $\mathbf{a} \in \mathbb{R}^n$, denote $\text{diag}(\mathbf{a})$ to be the diagonal matrix with diagonal entries given by \mathbf{a} . Let

$$\begin{aligned}\mathbf{F}_n &\equiv \frac{1}{n-1} (n\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n), \\ \mathbf{G}_n &\equiv \left(1 - \frac{1}{(n-1)^2}\right) \mathbf{I}_n + \frac{1}{(n-1)^2} \mathbf{1}_n\mathbf{1}'_n.\end{aligned}$$

It is noted that $\mathbf{F}_n\mathbf{1}_n = 0$ and

$$\mathbf{G}_n^{-1} = \frac{n-1}{n^2(n-2)} \{n(n-1)\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n\}.$$

Proposition 1 *Suppose that $a_i > 0$, $\sigma_i^2 > 0$; $i = 1, \dots, n$, where $n > 2$. Let $\mathbf{a} \equiv (a_1, \dots, a_n)'$ and $\mathbf{c} \equiv \mathbf{G}_n^{-1}(a_1^2\sigma_1^2, \dots, a_n^2\sigma_n^2)'$. If*

$$\mathbf{U} \equiv (u_{i,j})_{n \times n} \equiv (\text{diag}(\mathbf{a}))^{-1} \mathbf{F}_n \text{diag}(\mathbf{c}) \mathbf{F}_n (\text{diag}(\mathbf{a}))^{-1}, \quad (31)$$

then the following statements hold:

- (i) $\mathbf{a}'\mathbf{U}\mathbf{a} = 0$,
- (ii) $u_{i,i} = \sigma_i^2$; $i = 1, \dots, n$,
- (iii) \mathbf{U} is non-negative definite if and only if

$$\min \{a_1^2\sigma_1^2, \dots, a_n^2\sigma_n^2\} \geq \frac{1}{n(n-1)} \sum_{j=1}^n a_j^2\sigma_j^2.$$

Proof. (i) $\mathbf{a}'\mathbf{U}\mathbf{a} = \mathbf{1}'_n\mathbf{F}_n\mathit{diag}(\mathbf{c})\mathbf{F}_n\mathbf{1}_n = 0$.

(ii) We have

$$\begin{aligned}
(a_1^2\sigma_1^2, \dots, a_n^2\sigma_n^2)' &= \mathbf{G}_n\mathbf{c} \\
&= \left(1 - \frac{1}{(n-1)^2}\right)\mathbf{c} + \frac{1}{(n-1)^2}\mathbf{1}_n\mathbf{1}'_n\mathbf{c} \\
&= \text{the vector of the diagonal elements of } \mathbf{F}_n\mathit{diag}(\mathbf{c})\mathbf{F}_n \\
&= \text{the vector of the diagonal elements of } \mathit{diag}(\mathbf{a})\mathbf{U}\mathit{diag}(\mathbf{a}) \\
&= (a_1^2u_{1,1}, \dots, a_n^2u_{n,n})',
\end{aligned}$$

where the third equality follows by direct calculation, and the fourth equality follows from (31). Hence $u_{i,i} = \sigma_i^2$; $i = 1, \dots, n$.

(iii) From (31), we know that \mathbf{U} is non-negative definite if and only if $c_i \geq 0$; $i = 1, \dots, n$. Now for each $i = 1, \dots, n$, by direct calculation, we have $c_i \geq 0$ if and only if $n(n-1)a_i^2\sigma_i^2 \geq \sum_{j=1}^n a_j^2\sigma_j^2$. This gives the desired result. \square

REFERENCES

- Boyles, R. A. (1983). On the convergence of the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **45**, 47-50.
- Chou, K. C., Willsky, A. S., and Nikoukhah, R. (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. Wiley, New York.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- Dobson, G. M. B., Harrison, D. C., and Lawrence, J. (1929). Measurement of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions, part III. *Royal Society (London) Proceedings, Series A*, **122**, 456-486.

- Fang, D. and Stein, M. (1998). Some statistical models and methods for analyzing the TOMS data. *Journal of Geophysical Research*, **103**, 26165-26182.
- Fieguth, P. W., and Willsky, A. S. (1996). Fractal estimation using models on multiscale trees. *IEEE Transactions on Signal Processing*, **44**, 1297-1300.
- Gabrosek, J. G., Cressie, N., and Huang, H.-C. (1999). Spatio-temporal prediction of level 3 data for NASA's Earth Observing System. In *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, K. Lowell ed. Ann Arbor Press, Chelsea, MI, 331-337.
- Huang, H.-C. and Cressie, N. (2000). Multiscale graphical modeling in space: Applications to command and control. *Spatial Statistics and Applications*, M. Moore ed. Springer Lecture Notes in Statistics. Springer, New York, forthcoming.
- Kahn, R. (1996). What shall we do with the data we are expecting in 1998? In *Massive Data Sets. Proceedings of a Workshop*. National Academy of Sciences, Washington, D.C., 15-21.
- Kahn, R. (1998). Why do we need discrete global grids for satellite remote sensing? (Abstract). *Computing Science and Statistics*, **30**, 285.
- Kolaczyk, E. D. (1999). Bayesian multi-scale models for Poisson processes. *Journal of the American Statistical Association*, **94**, 920-933.
- London, J. (1985). The observed distribution of atmospheric ozone and its variations. *Ozone in the Free Atmosphere*, R. C. Whitten and S. S. Prasad, eds. Van Nostrand Reinhold Company, New York, 11-80.
- Luetzgen, M. R. and Willsky, A. S. (1995a). Multiscale smoothing error models. *IEEE Transactions on Automatic Control*, **40**, 173-175.
- Luetzgen, M. R. and Willsky, A. S. (1995b). Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination. *IEEE Transactions on*

Image Processing, **4**, 194-207.

Madrid, C. R. (1978). The Nimbus-7 User's Guide. *Goddard Space Flight Center*. NASA, Greenbelt, Maryland.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674-693.

McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Schlesinger, B. M., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Taylor, S. L., Swissler, T., Torres, O., Labow, G., Byerly, W., and Cebula, R. P. (1996). The Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide. *NASA Reference Publication 1384*. NASA, Greenbelt, Maryland.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.

National Aeronautics and Space Administration (1992). *Report to Congress on the Restructuring of the Earth Observing System* (March, 9, 1992). NASA, Washington, D.C.

Niu, X. and Tiao, G. (1995). Modelling satellite ozone data. *Journal of the American Statistical Association*, **90**, 969-983.

Wu, C. F. (1983). On the convergence of the EM algorithm. *Annals of Statistics*, **11**, 95-103.

Zeng, L. and Levy, G. (1995). Space and time aliasing structure in monthly mean polar-orbiting satellite data. *Journal of Geophysical Research*, **100**, 5133-5142.

CAPTIONS FOR FIGURES

Figure 1. A quadtree.

Figure 2. TCO zonal mean versus latitude, for the TOMS data on October 1, 1988

Figure 3. Preliminary level-3 TCO data.

Figure 4. Spline fitting for the variances of TCO variables with respect to their grid-cell areas.
Units on the horizontal axis are in kilometers squared.

Figure 5. Predicted residual TCO.

Figure 6. Prediction standard errors for residual TCO.

Figure 7. Predicted TCO for resolution 1.

Figure 8. Predicted TCO for resolution 2.

Figure 9. Predicted TCO for resolution 3.

Figure 10. Predicted TCO for resolution 4.

Figure 11. Predicted TCO for the finest resolution (resolution 5).

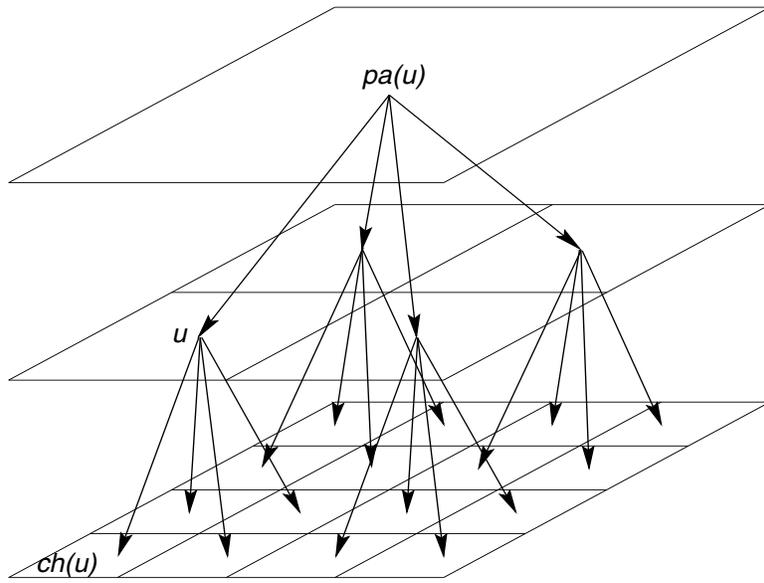


Figure 1

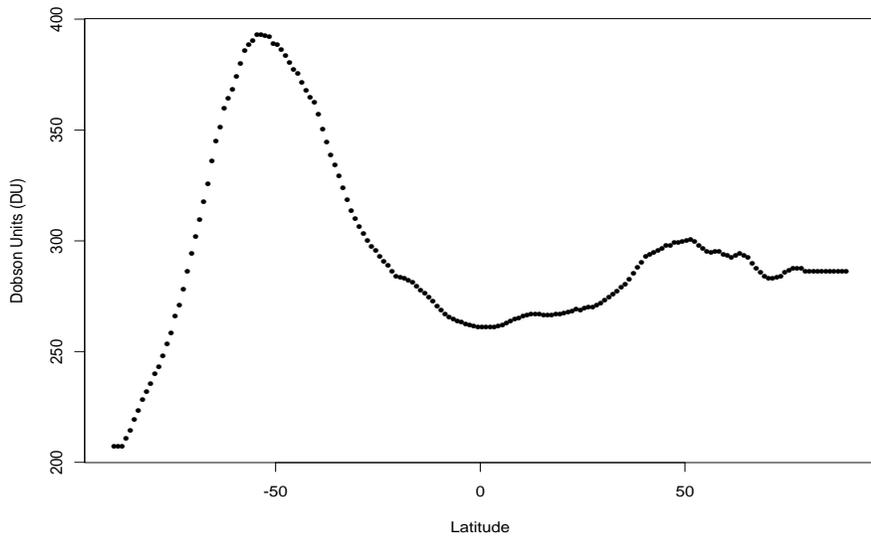


Figure 2

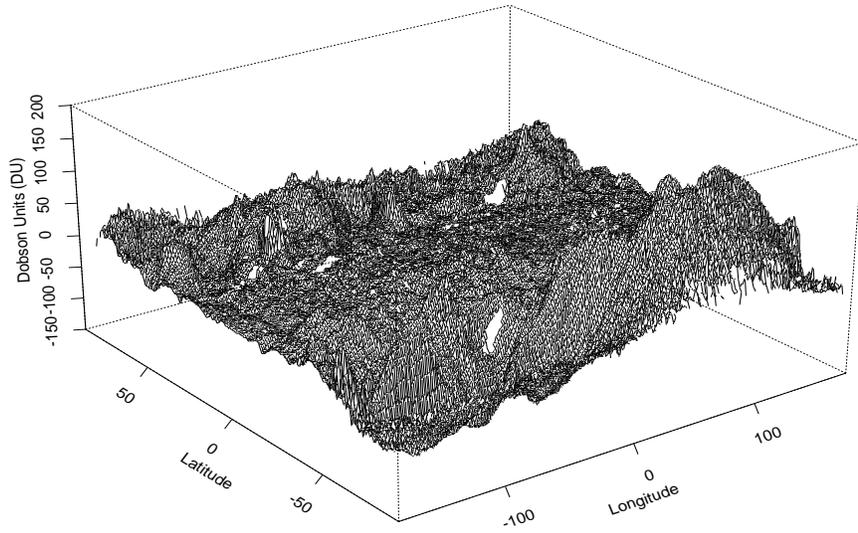


Figure 3

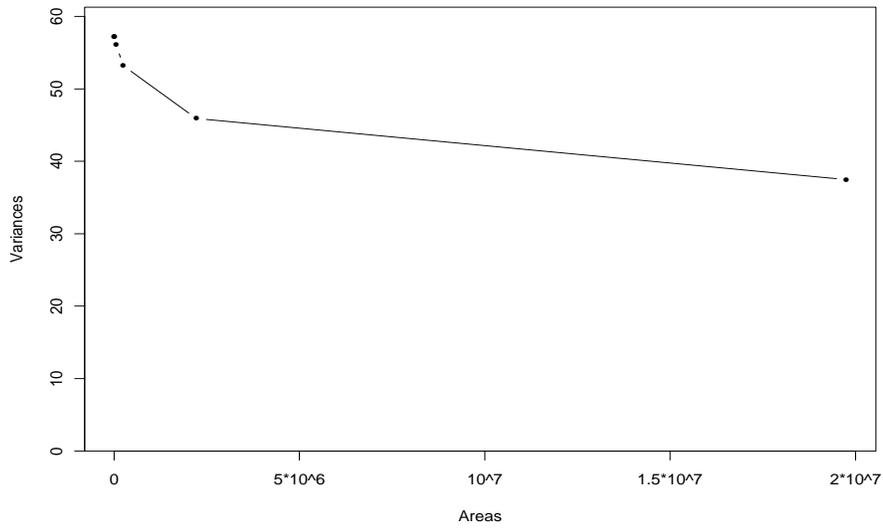


Figure 4

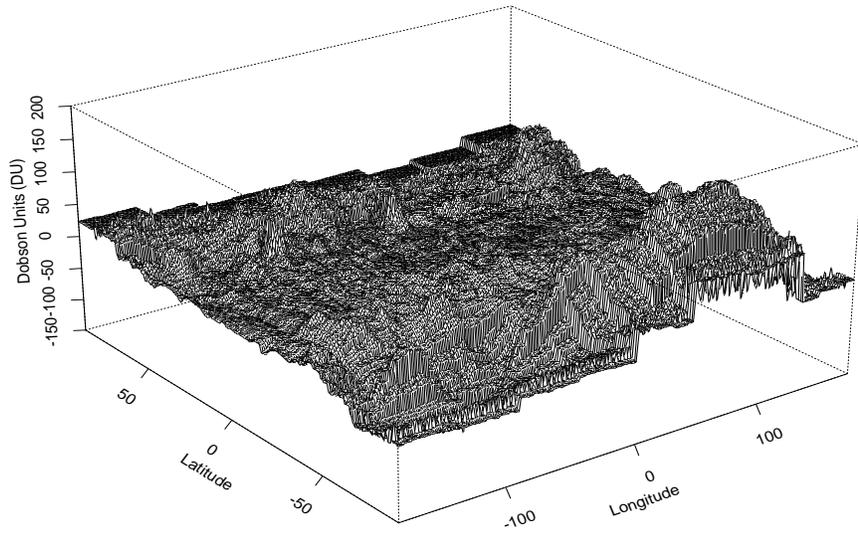


Figure 5

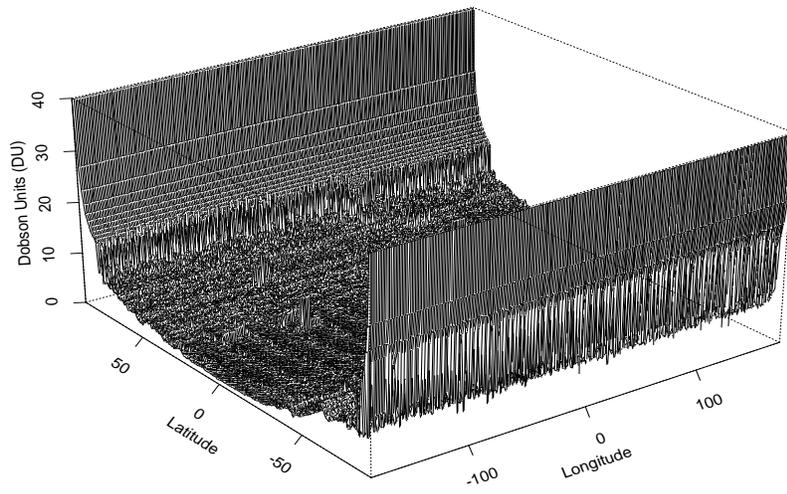


Figure 6

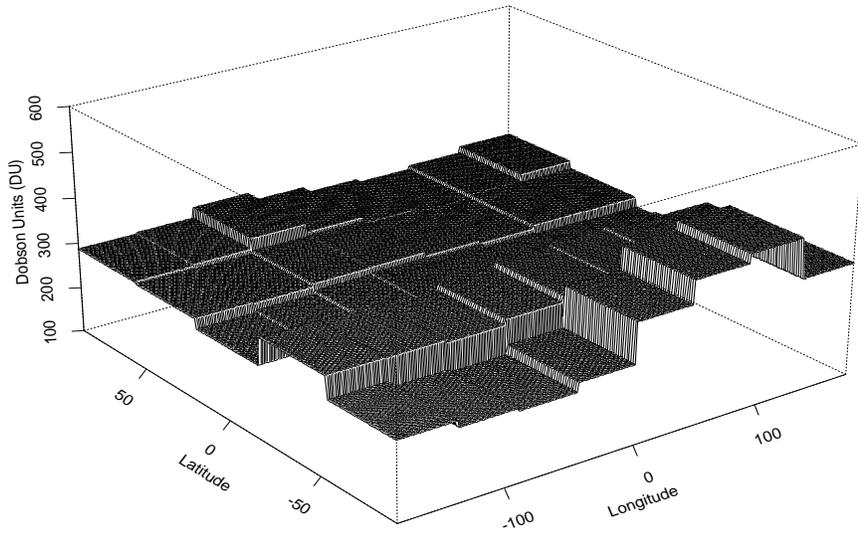


Figure 7

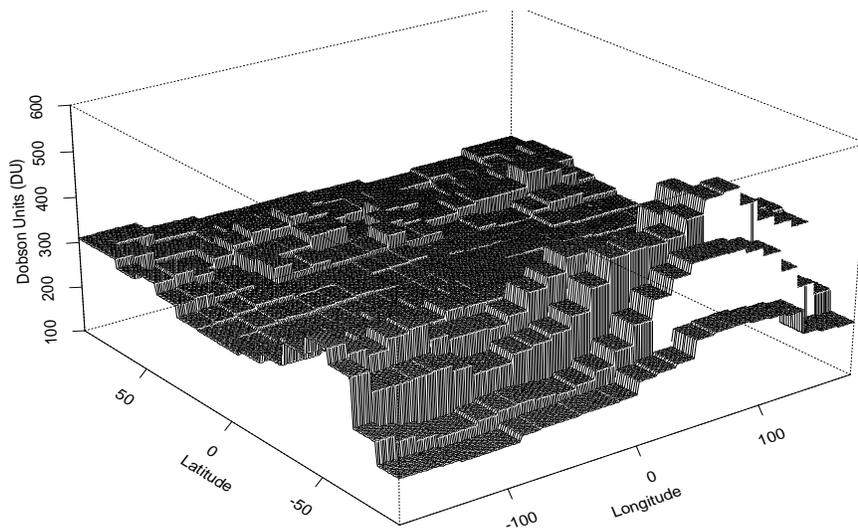


Figure 8

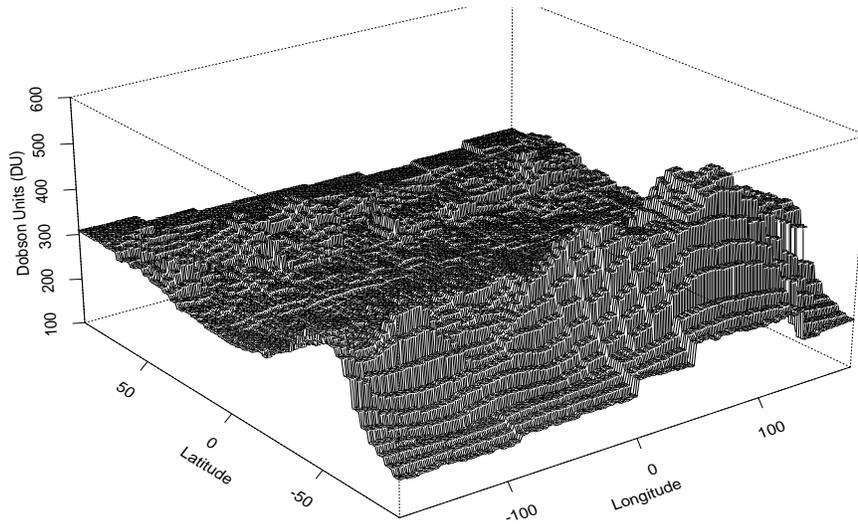


Figure 9

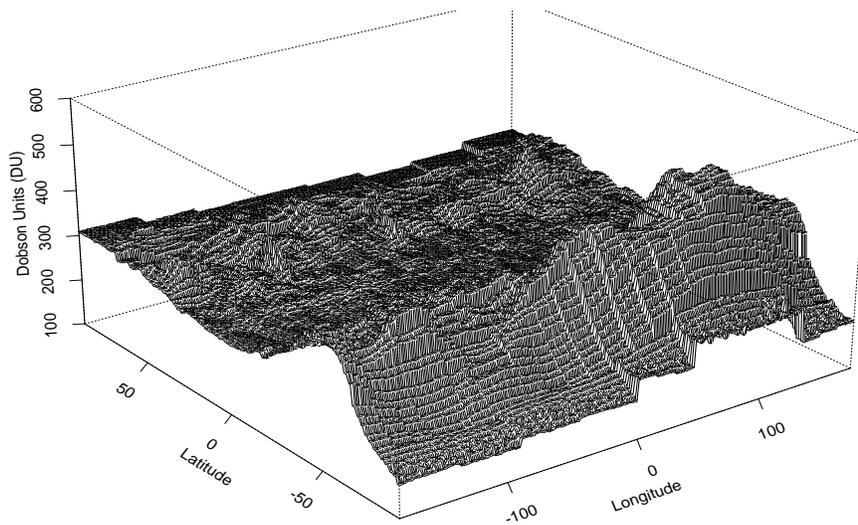


Figure 10

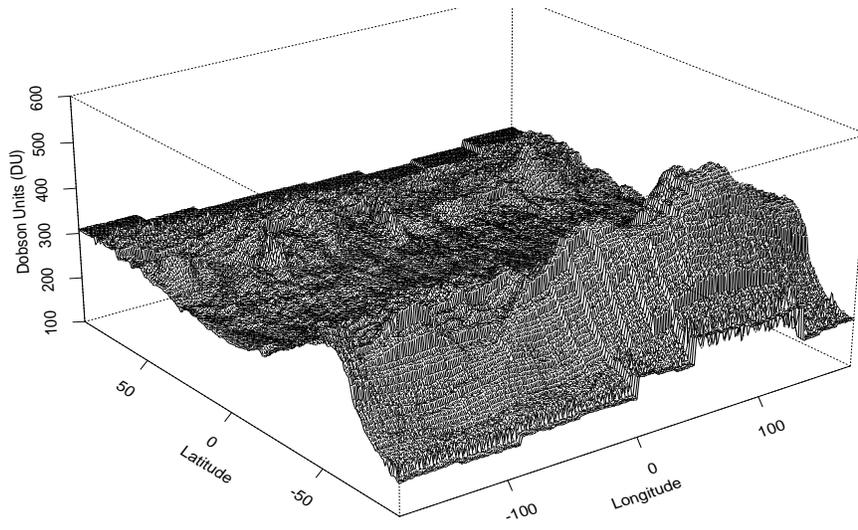


Figure 11