UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2002

Fast, resolution-consistent spatial prediction of global processes from satellite data

Hsin-Cheng Huang
Institute Of Statistical Science, Academia Sinica, Nankang

Noel A. Cressie
University of Wollongong, ncressie@uow.edu.au

John Gabrosek
Grand Valley State University

Publication Details

Huang, H., Cressie, N. A. & Gabrosek, J. (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. Journal of Computational and Graphical Statistics, 11 (1), 63-88.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Fast, resolution-consistent spatial prediction of global processes from satellite data

Abstract

Polar orbiting satellites remotely sense the earth and its atmosphere, producing datasets that give daily global coverage. For any given day, the data are many and measured at spatially irregular locations. 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 applies a multiresolution autoregressive tree-structured model, and presents a new statistical prediction methodology that is resolution consistent (i.e., 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.

Keywords

global, processes, fast, resolution, consistent, satellite, spatial, data, prediction

Disciplines

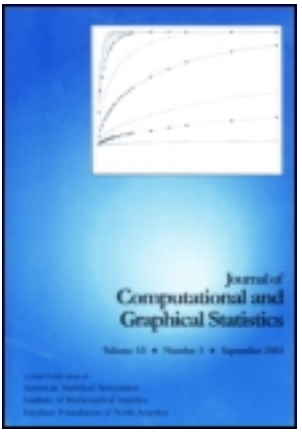
Physical Sciences and Mathematics

Publication Details

Huang, H., Cressie, N. A. & Gabrosek, J. (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. Journal of Computational and Graphical Statistics, 11 (1), 63-88.

This journal article is available at Research Online: http://ro.uow.edu.au/infopapers/2491

This article was downloaded by: [University of Wollongong]
On: 29 November 2012, At: 15:01
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Computational and Graphical Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ucgs20

Fast, Resolution-Consistent Spatial Prediction of Global Processes From Satellite Data

Hsin-Cheng Huang, Noel Cressie and John Gabrosek
Hsin-Cheng Huang is Assistant Research Fellow, Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan . Noel Cressie is Professor, Department of Statistics, TheOhioStateUniversity, Columbus,OH 43210-1247 . John Gabrosek is Assistant Professor, Department of Statistics, Grand Valley State University, Allendale, MI 49401-9403.
Version of record first published: 01 Jan 2012.

To cite this article: Hsin-Cheng Huang, Noel Cressie and John Gabrosek (2002): Fast, Resolution-Consistent Spatial Prediction of Global Processes From Satellite Data, Journal of Computational and Graphical Statistics, 11:1, 63-88

To link to this article: http://dx.doi.org/10.1198/106186002317375622

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Fast, Resolution-Consistent Spatial Prediction of Global Processes From Satellite Data

Hsin-Cheng HUANG, Noel CRESSIE, and John GABROSEK

Polar orbiting satellites remotely sense the earth and its atmosphere, producing datasets that give daily global coverage. For any given day, the data are many and measured at spatially irregular locations. 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 applies a multiresolution autoregressive tree-structured model, and presents a new statistical prediction methodology that is resolution consistent (i.e., 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; Covariance function; Mass balance; Multiresolution; Nonstationarity; Tree-structured model.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

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, Willsky, and Nikoukhah 1994; Huang and Cressie 1997, 2001).

This article concentrates on autoregressive tree-structured models, where optimal spatial prediction procedures have been shown to be extremely fast. Pioneers in this area have been A. S. Willsky and colleagues (e.g., Chou et al. 1994; Luettgen and Willsky 1995a,

Hsin-Cheng Huang is Assistant Research Fellow, Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan (E-mail: hchuang@stat.sinica.edu.tw). Noel Cressie is Professor, Department of Statistics, The Ohio State University, Columbus, OH 43210-1247 (E-mail: ncressie@stat.ohio-state.edu). John Gabrosek is Assistant Professor, Department of Statistics, Grand Valley State University, Allendale, MI 49401-9403 (E-mail: gabrosej@gvsu.edu).

©2002 *American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America*
Journal of Computational and Graphical Statistics, Volume 11, Number 1, Pages 63–88

1995b; Fieguth and Willsky 1996); in Daoudi, Frakt, and Willsky (1999), these models are referred to as multiscale autoregressive (MAR) models. In our application to global spatial prediction of total column ozone, we focus on a specific subclass of these models and introduce the natural requirement of "mass balance." A referee has pointed out to us that mass balance is a particular case of the mathematical property of "internality" introduced in a conference proceedings paper by Frakt and Willsky (1998), although here we motivate mass balance as an important physical property.

The approach we take here for optimal spatial prediction is easily able to handle large-to-massive amounts of daily data, without overow 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 apply a 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 was 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 and covariances are accounted 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 data are given at a fine scale of 1 degree latitude by 1.25 degree longitude. Although the predicted value of a larger area (e.g., a four-degree latitude by five-degree longitude cell) could be obtained by taking a volume-weighted average of the fine-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 autoregressive tree-structured models and the associated change-of-resolution Kalman-filter algorithm. The mass-balanced, tree-structured model is introduced in Section 3. Section 4 presents an application to TCO satellite data collected from the Total Ozone Mapping Spectrometer (TOMS) instrument using the proposed model (after appropriate detrending). We show how the model parameters are estimated after stratification to handle nonstationarities. 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.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Downloaded by [University of Wollongong] at 15:01 29 November 2012

2. AUTOREGRESSIVE 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 U.S. National Aeronautic and Space Administration (NASA) is sponsoring a program called the Earth Observing System (EOS), which consists of a series of satellites measuring the chemical and physical processes of the earth and its atmosphere (NASA 1992). There are different levels of satellite data classified by NASA, where in NASA's terminology, "level" refers to different stages of data processing. Level-1 data refer to raw radiances measured by satellites after some calibration. Level-2 data refer to geophysical parameters at the finest space-time resolution, which are processed from the corresponding level-1 data. These data are spatially and temporally irregular, so further processing of the data is required to yield level-3 data on regular grid cells. 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:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \eta(\mathbf{s}); \quad \mathbf{s} \in D, \tag{2.1}$$

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 statistical predictor and the prediction (co)variance of different aggregations of $Y(\cdot)$. Throughout the article, we refer to optimal prediction as the method of prediction that minimizes the mean-squared prediction error under the presumed model. At a given resolution and location D_u, where $|D_u| \equiv \int_{D_u} 1 d\mathbf{s} > 0$, we wish to predict:

$$y_u \equiv \frac{1}{|D_u|} \int_{D_u} Y(\mathbf{s}) d\mathbf{s} = \mu_u + \eta_u, \tag{2.2}$$

based on the noisy data $\mathbf{Z} \equiv \left(z_{u_1}, \ldots, z_{u_N}\right)'$, 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 $|D_{u_i}| > 0$; $i = 1, \ldots, N$. 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

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

where $\boldsymbol{\mu} \equiv \left(\mu_{u_1}, \ldots, \mu_{u_N}\right)'$. 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 challenging problems 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 $D_v = D_{v_1} \cup \cdots \cup D_{v_J}$ 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 autoregressive tree-structured models.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

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 $\mathrm{pa}(u)$ 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

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

Chou et al. (1994) introduced a hidden multivariate Gaussian tree-structured process $\{\mathbf{y}_u : u \in U\}$ on the tree (U, E), which is defined as follows. Assume that the Gaussian process evolves from parents to children in an autoregressive manner according to the following:

$$\begin{aligned} \mathbf{y}_u &= \mathbf{A}_u \mathbf{y}_{\mathrm{pa}(u)} + \mathbf{w}_u; \quad u \in U \setminus \{u_0\}, \\ \mathbf{z}_u &= \mathbf{C}_u \mathbf{y}_u + \boldsymbol{\epsilon}_u; \quad u \in U, \end{aligned} \tag{2.3}$$

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,

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

$\{\mathbf{y}_u\}$ and $\{\boldsymbol{\epsilon}_u\}$ are independent, and $\mathbf{y}_{\mathrm{pa}(u)}$ 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 $\{\mathbf{z}_{u_1}, \ldots, \mathbf{z}_{u_N}\}$. As motivation for (2.3), we think of children inheriting some of their behavior from their parent and the rest from other nonspecified sources. The AR(1) models in time series have a similar motivation.

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, $\text{an}(u, u')$, to u and u'. That is,

$$\text{cov}\left(\mathbf{y}_u, \mathbf{y}_{u'}\right) = \mathbf{A}_{u_1} \ldots \mathbf{A}_u \, \text{var}\left(\mathbf{y}_{\text{an}(u,u')}\right)\left(\mathbf{A}_{u'_1} \ldots \mathbf{A}_{u'}\right)', \tag{2.4}$$

where $\left(\text{an}(u, u'), u_1, \ldots, u\right)$ and $\left(\text{an}(u, u'), u'_1, \ldots, u'\right)$ are the paths from $\text{an}(u, u')$ 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 autoregressive tree-structured models. 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 can also be applied to more general graphical Markov models by grouping variables to form an autoregressive tree-structured model (see Huang and Cressie 1997; Daoudi et al. 1999). The resulting change-of-resolution Kalman filter can be viewed as a special case of the more general junction-tree algorithm (Lauritzen and Spiegelhalter 1988; Lauritzen 1992) developed for general graphical Markov models.

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

$$\begin{aligned}
U_u &\equiv \{u' : u \prec u'\}, \\
\gamma_u &\equiv \begin{cases} 1; & \mathbf{z}_u \text{ is observed at } u, \\ 0; & \text{otherwise}, \end{cases} \\
\mathbf{Z} &\equiv \{\mathbf{z}_u : \gamma_u = 1\}, \\
\mathbf{Z}_u &\equiv \{\mathbf{z}_{u'} : \gamma_{u'} = 1,\ u' \in U_u\}, \\
\mathbf{Z}^*_u &\equiv \{\mathbf{z}_{u'} : \gamma_{u'} = 1,\ u' \in U_u \setminus \{u\}\}, \\
\hat{\mathbf{y}}_{u_1|u_2} &\equiv \text{E}\left(\mathbf{y}_{u_1} \middle| \mathbf{Z}_{u_2}\right), \\
\hat{\mathbf{y}}^*_{u_1|u_2} &\equiv \text{E}\left(\mathbf{y}_{u_1} \middle| \mathbf{Z}^*_{u_2}\right), \\
\hat{\mathbf{y}}_u &\equiv \text{E}\left(\mathbf{y}_u \middle| \mathbf{Z}\right), \\
\mathbf{V}_u &\equiv \text{var}\left(\mathbf{y}_u\right), \\
\mathbf{\Gamma}_{u_1|u_2} &\equiv \text{var}\left(\mathbf{y}_{u_1} \middle| \mathbf{Z}_{u_2}\right), \\
\mathbf{\Gamma}^*_{u_1|u_2} &\equiv \text{var}\left(\mathbf{y}_{u_1} \middle| \mathbf{Z}^*_{u_2}\right), \\
\mathbf{\Gamma}_u &\equiv \text{var}\left(\mathbf{y}_u \middle| \mathbf{Z}\right), \\
\mathbf{\Gamma}_{u_1,u_2} &\equiv \text{cov}\left(\mathbf{y}_{u_1}, \mathbf{y}_{u_2} \middle| \mathbf{Z}\right) = \text{cov}\left(\hat{\mathbf{y}}_{u_1} - \mathbf{y}_{u_1},\ \hat{\mathbf{y}}_{u_2} - \mathbf{y}_{u_2}\right).
\end{aligned}$$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

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

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

Note that $\{\mathbf{V}_u\}$ can be computed recursively; from (2.3), we have that

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

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 Bayes' theorem for multivariate Gaussian prior and data processes, we have, for a leaf $u \in U$,

$$\hat{\mathbf{y}}_{u|u} = \gamma_u\mathbf{V}_u\mathbf{C}_u'\left(\mathbf{C}_u\mathbf{V}_u\mathbf{C}_u' + \mathbf{\Phi}_u\right)^{-1}\mathbf{z}_u, \tag{2.5}$$

$$\mathbf{\Gamma}_{u|u} = \mathbf{V}_u - \gamma_u\mathbf{V}_u\mathbf{C}_u'\left(\mathbf{C}_u\mathbf{V}_u\mathbf{C}_u' + \mathbf{\Phi}_u\right)^{-1}\mathbf{C}_u\mathbf{V}_u. \tag{2.6}$$

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

$$\hat{\mathbf{y}}_{u|\mathrm{ch}(u,i)} = \mathbf{B}_{\mathrm{ch}(u,i)}\hat{\mathbf{y}}_{\mathrm{ch}(u,i)|\mathrm{ch}(u,i)}, \tag{2.7}$$

$$\mathbf{\Gamma}_{u|\mathrm{ch}(u,i)} = \mathbf{B}_{\mathrm{ch}(u,i)}\mathbf{\Gamma}_{\mathrm{ch}(u,i)|\mathrm{ch}(u,i)}\mathbf{B}_{\mathrm{ch}(u,i)}' + \mathbf{R}_{\mathrm{ch}(u,i)}, \tag{2.8}$$

$$\hat{\mathbf{y}}_{u|u}^* = \mathbf{\Gamma}_{u|u}^*\left(\sum_{i=1}^{n_u}\mathbf{\Gamma}_{u|\mathrm{ch}(u,i)}^{-1}\hat{\mathbf{y}}_{u|\mathrm{ch}(u,i)}\right), \tag{2.9}$$

$$\mathbf{\Gamma}_{u|u}^* = \left\{\mathbf{V}_u^{-1} + \sum_{i=1}^{n_u}\left(\mathbf{\Gamma}_{u|\mathrm{ch}(u,i)}^{-1} - \mathbf{V}_u^{-1}\right)\right\}^{-1}, \tag{2.10}$$

$$\hat{\mathbf{y}}_{u|u} = \mathbf{\Gamma}_{u|u}\left(\gamma_u\mathbf{C}_u'\mathbf{V}_u^{-1}\mathbf{z}_u + (\mathbf{\Gamma}_{u|u}^*)^{-1}\hat{\mathbf{y}}_{u|u}^*\right), \tag{2.11}$$

$$\mathbf{\Gamma}_{u|u} = \mathbf{\Gamma}_{u|u}^* - \gamma_u\mathbf{\Gamma}_{u|u}^*\mathbf{C}_u'\left(\mathbf{C}_u\mathbf{\Gamma}_{u|u}^*\mathbf{C}_u' + \mathbf{\Phi}_u\right)^{-1}\mathbf{C}_u\mathbf{\Gamma}_{u|u}^*. \tag{2.12}$$

At the root u_0, we have

$$\begin{aligned} \hat{\mathbf{y}}_{u_0} &= \hat{\mathbf{y}}_{u_0|u_0}, \\ \mathbf{\Gamma}_{u_0} &= \mathbf{\Gamma}_{u_0|u_0}. \end{aligned}$$

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 $u \in U$:

$$\hat{\mathbf{y}}_u = \hat{\mathbf{y}}_{u|u} + \mathbf{\Gamma}_{u|u}\mathbf{B}_u'\mathbf{\Gamma}_{\mathrm{pa}(u)|u}^{-1}\left(\hat{\mathbf{y}}_{\mathrm{pa}(u)} - \hat{\mathbf{y}}_{\mathrm{pa}(u)|u}\right), \tag{2.13}$$

$$\mathbf{\Gamma}_u = \mathbf{\Gamma}_{u|u} + \mathbf{\Gamma}_{u|u}\mathbf{B}_u'\mathbf{\Gamma}_{\mathrm{pa}(u)|u}^{-1}\left(\mathbf{\Gamma}_{\mathrm{pa}(u)} - \mathbf{\Gamma}_{\mathrm{pa}(u)|u}\right)\mathbf{\Gamma}_{\mathrm{pa}(u)|u}^{-1}\mathbf{B}_u\mathbf{\Gamma}_{u|u}. \tag{2.14}$$

The algorithm is fast with computing time only proportional to the number of the nodes in U. The proportionality constant at a node increases as the cube of the dimension of the random

Downloaded by [University of Wollongong] at 15:01 29 November 2012

vector associated with that node, due to taking a matrix inverse. Complete derivation of the algorithm can be found in Chou et al. (1994) or, using Bayes' theorem directly, by Huang and Cressie (2001). In fact, it can be derived from a more general junction-tree algorithm for general graphical Markov models (Lauritzen and Spiegelhalter 1988; Lauritzen 1992).

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

$$\hat{\mathbf{y}}_u - \mathbf{y}_u = \mathbf{G}_u \left(\hat{\mathbf{y}}_{\mathrm{pa}(u)} - \mathbf{y}_{\mathrm{pa}(u)} \right) + \xi_u; \quad u \in U \setminus \{u_0\}, \tag{2.15}$$

where

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

$\hat{\mathbf{y}}_{\mathrm{pa}(u)} - \mathbf{y}_{\mathrm{pa}(u)}$ and $\{\xi_u : u \in U \setminus U_u\}$ are independent, zero-mean, and Gaussian for $u \in U \setminus \{u_0\}$, and

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

Note that $\mathbf{G}_u$; $u \in U$, can be computed in the leaves-to-root filtering step. From (2.4) and (2.15), for any $u, u' \in U$, we can obtain the prediction covariance between any two variables as:

$$\begin{aligned} \mathbf{\Gamma}_{u,u'} &= \mathrm{cov}\left(\hat{\mathbf{y}}_u - \mathbf{y}_u,\ \hat{\mathbf{y}}_{u'} - \mathbf{y}_{u'}\right) \\ &= \mathbf{G}_{u_1} \ldots \mathbf{G}_u \,\mathrm{var}\left(\hat{y}_{\mathrm{an}(u,u')} - y_{\mathrm{an}(u,u')}\right) \left(\mathbf{G}_{u_1'} \ldots \mathbf{G}_{u'}\right)', \end{aligned}$$

where the notation is the same as in Equation (2.4).

The change-of-resolution Kalman-filter algorithm presented above yields predictors that do *not* preserve mass balance. That is, the algorithm does not guarantee aggregation consistency for predictions and prediction variances. Moreover, we need 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. MULTIRESOLUTION TREE-STRUCTURED SPATIAL MODELS

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

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Downloaded by [University of Wollongong] at 15:01 29 November 2012

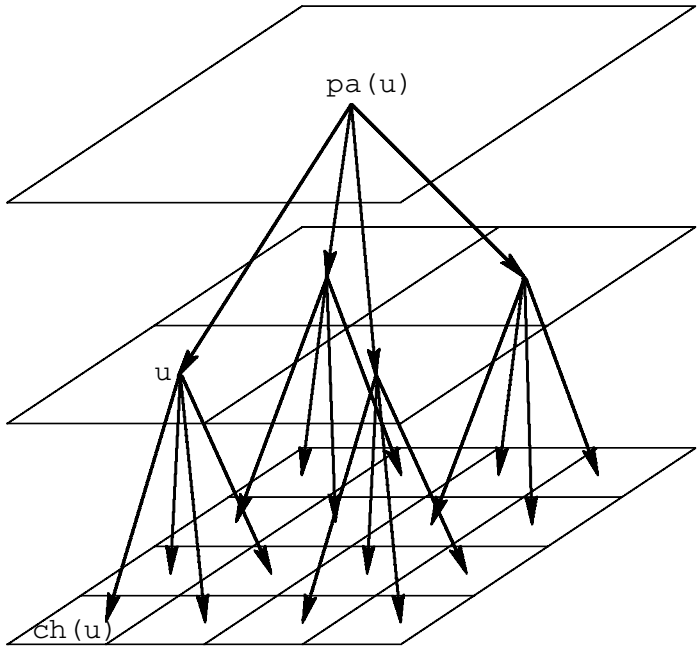


Figure 1. A quadtree.

3.1 BASIC SPATIAL MODEL

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

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 (2.1) and (2.2), let the nodes of a tree be locations at the centroid of the corresponding subregions $\{D_{j,k}\}$, where $\cup_{k=0}^{N_1-1} D_{1,k} = \cdots = \cup_{k=0}^{N_J-1} D_{J,k} = D$ and $\{D_{j,k} : k = 0, \ldots, N_{j-1}\}$ are disjoint with $|D_{j,0}| = \cdots = |D_{j,N_j-1}| > 0$; $j = 1, \ldots, J$. 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 $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J$, 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 that 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 $y_{j+1,\, kn_j}, \ldots, y_{j+1,\, (k+1)n_j - 1}$. Thus, $y_{\text{pa}(j,k)} = y_{j-1,\, [k/n_{j-1}]}$ for $k = 0, \ldots, N_j - 1$, $j = 2, \ldots, J$, where $[x]$ denotes the largest integer less than or equal to x. A multiresolution autoregres-

sive tree-structured model is given as follows:

$$y_{j,k} \quad = \quad y_{\mathrm{pa}(j,k)} + w_{j,k}; \quad k = 0, \ldots, N_j - 1,\ j = 2, \ldots, J, \tag{3.1}$$

is the state equation, and the measurement equation is given by

$$z_{j,k} \quad = \quad y_{j,k} + \varepsilon_{j,k}; \quad k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J, \tag{3.2}$$

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

$$\begin{aligned} y_{1,k} &\sim N(0, \sigma_1^2); \quad k = 0, \ldots, N_1 - 1, \quad \text{independently,} \\ w_{j,k} &\sim N(0, \sigma_j^2); \quad k = 0, \ldots, N_j - 1,\ j = 2, \ldots, J, \quad \text{independently,} \end{aligned}$$

$y_{j,k}$ and $\varepsilon_{j,k}$ are independent for $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J$, and $y_{j-1,k}$ and $w_{j,k}$ are independent for $k = 0, \ldots, N_j - 1$, $j = 2, \ldots, J$. Hence, for $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J$, $E(y_{j,k}) = 0$; and for $j = 1, \ldots, J$,

$$\begin{aligned} \mathrm{var}(\mathbf{Y}_j) = \sigma_j^2\, \mathbf{I}_{N_j} + \sigma_{j-1}^2 \mathbf{I}_{N_{j-1}} \otimes (\mathbf{1}_{n_{j-1}} \mathbf{1}'_{n_{j-1}}) \\ + \cdots + \sigma_1^2 \mathbf{I}_{N_1} \otimes (\mathbf{1}_{n_1 \ldots n_{j-1}} \mathbf{1}'_{n_1 \ldots n_{j-1}}), \end{aligned} \tag{3.3}$$

where $\mathbf{Y}_j \equiv \left(y_{j,0}, \ldots, y_{j,N_j-1}\right)'$, $\mathbf{I}_m$ is the $m \times m$ identity matrix, and $\mathbf{1}_m$ is the $m \times 1$ vector whose entries are all 1. Note that the model could allow for heterogeneous means. If we assume that the mean of the root node of the kth tree is μ_k (instead of zero), then all children in that tree would also have mean μ_k; $k = 0, \ldots, N_1 - 1$.

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

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

However, this mass-balance equation does not hold for the model (3.1), 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, and hence resolution consistency, by imposing a linear constraint on the $\{w_{j,k}\}$.

3.2 HOMOGENEOUS, MASS-BALANCED, TREE-STRUCTURED SPATIAL MODELS

Define

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

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Downloaded by [University of Wollongong] at 15:01 29 November 2012

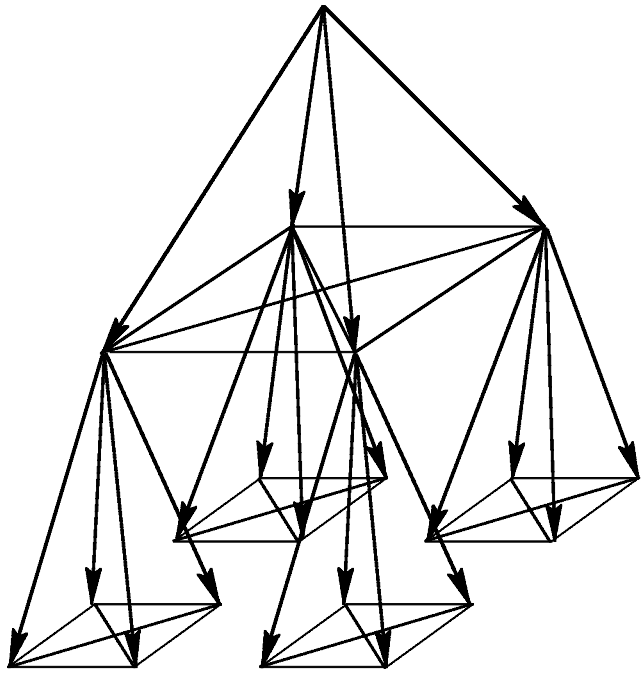


Figure 2. A vector tree formed by grouping variables associated with a quadtree.

That is, $\mathbf{y}_{\mathbf{ch}(j,k)}$ and $\mathbf{w}_{\mathbf{ch}(j,k)}$ consist of all the children of $y_{j,k}$ and $w_{j,k}$, respectively. Then, a multiresolution, mass-balanced, autoregressive tree-structured spatial model is given as follows:

$$\mathbf{y}_{\mathbf{ch}(j,k)} = y_{j,k}\mathbf{1}_{n_j} + \mathbf{w}_{\mathbf{ch}(j,k)}; \quad k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J-1, \tag{3.4}$$

is the state equation, and the measurement equation is given by

$$z_{j,k} = y_{j,k} + \varepsilon_{j,k}; \quad k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J, \tag{3.5}$$

where $\{z_{j,k}\}$ are (potential) observations, $\varepsilon_{j,k} \sim N(0, \phi_{j,k})$; $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J$, are independent, zero-mean, Gaussian random variables representing measurement errors, $y_{1,k} \sim N(0, \sigma_1^2)$; $k = 0, \ldots, N_1 - 1$, independently, and

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

Hence,

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

Comparing (3.4) and (3.5) with (3.1) and (3.2), the new feature of the model is that we now have constraints on $\left\{\mathbf{w}_{\mathbf{ch}(j,k)}\right\}$, which imply that, for $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J-1$,

$$\frac{1}{n_j}\sum_{l=0}^{n_j-1} y_{j,\,kn_j+l} = \frac{1}{n_j}\mathbf{1}'_{n_j}\mathbf{y}_{\mathbf{ch}(j,k)} = y_{j,k}\ . \tag{3.6}$$

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 $k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J-1$, the mass-balanced, tree-structured model given by (3.4) and (3.5) can be regarded as a *vector* tree-structured model on tree (U, E) *without* constraints, where

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

This new tree is obtained from the original tree by tying the original branches from an original node together to form a new node of the new tree (see Figure 2). The idea of grouping variables to form a tree can also be found in Huang and Cressie (1997) and Daoudi et al. (1999). 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. Note that for $k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J-1$,

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

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

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

has a closed-form expression and can be easily computed as part of the multiresolution Kalman filter. However, the prediction algorithm based on the mass-balanced model does sacrifice some computational efficiency, because some nontrivial but small matrix inversions are necessary at each grid node.

Also note that, for $j = 1, \ldots, J$, the variance of $\mathbf{Y}_j = \left(y_{j,0}, \ldots, y_{j,N_j - 1}\right)'$ is given by

$$\begin{aligned} \mathrm{var}(\mathbf{Y}_j) &= \sigma_j^2 \mathbf{I}_{N_j} + \left(\sigma_{j-1}^2 - \sigma_j^2 / n_{j-1}\right) \mathbf{I}_{N_{j-1}} \otimes \Big(\mathbf{1}_{n_{j-1}} \mathbf{1}_{n_{j-1}}'\Big) + \cdots \\ &\quad + \left(\sigma_1^2 - \sigma_2^2 / n_1\right) \mathbf{I}_{N_1} \otimes \Big(\mathbf{1}_{n_1 \ldots n_{j-1}} \mathbf{1}_{n_1 \ldots n_{j-1}}'\Big). \end{aligned} \tag{3.7}$$

Comparing (3.3) and (3.7) for a fixed scale $j \in \{1, \ldots, J\}$, it follows that the joint distribution of $\mathbf{Y}_j$ has the same form as that for an unconstrained tree-structured model if $\sigma_{i-1}^2 > \sigma_i^2 / n_{i-1}$; $i = 2, \ldots, j-1$. 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, $j = J$. Furthermore, estimation of the parameters $\{\sigma_1^2, \ldots, \sigma_J^2\}$ can be obtained using the EM algorithm (Dempster, Laird, and Rubin 1977) by treating the unobserved state variables $\{y_{j,k}\}$ as missing data. The algorithm

Downloaded by [University of Wollongong] at 15:01 29 November 2012

requires running the change-of-resolution Kalman-filter algorithm in the E-step, followed by simple closed-form computation in the M-step, and iterating until convergence. Variance parameters can also be estimated by exploiting the spatial relations between data and predictor; these are discussed in Section 3.3 and implemented on the TCO data in Section 4.

3.3 HETEROGENEOUS, MASS-BALANCED, TREE-STRUCTURED SPATIAL 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* $\{D_{j,k} \subset D : k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J\}$ *is called a nested partitioning on* D*, where* $|D| > 0$*, if the following conditions hold:*

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

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

$$E \equiv \left\{ \big(\mathrm{pa}\big((j,k)\big), (j,k)\big) : k = 0, \ldots, N_j - 1,\ j = 2, \ldots, J \right\}.$$

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

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

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

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

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

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

$$\mathbf{y}_{\mathbf{ch}(j,k)} = y_{j,k}\mathbf{1}_{n_{j,k}} + \mathbf{w}_{\mathbf{ch}(j,k)}; \quad k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J-1, \quad (3.8)$$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

is the state equation, and the measurement equation is given by

$$z_{j,k} \quad = \quad y_{j,k} + \varepsilon_{j,k}; \quad k = 0, \ldots, N_j - 1, \; j = 1, \ldots, J, \tag{3.9}$$

where $\{z_{j,k}\}$ are (potential) observations, $\varepsilon_{j,k} \sim N(0, \phi_{j,k})$; $k = 0, \ldots, N_j - 1$, $j = 1, \ldots, J$, are independent, zero-mean, Gaussian random variables representing measurement errors, $y_{1,k} \sim N(0, \sigma_1^2)$; $k = 0, \ldots, N_1 - 1$, independently, and

$$\mathbf{w}_{\mathbf{ch}(j,k)} \equiv \left(w_{\mathrm{ch}(j,k,1)}, \ldots, w_{\mathrm{ch}(j,k,\, n_{j,k})}\right)' \sim N\left(\mathbf{0}, \mathbf{W}_{\mathbf{ch}(j,k)}\right), \tag{3.10}$$

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

$$\mathbf{a}'_{\mathbf{ch}(j,k)} \mathbf{w}_{\mathbf{ch}(j,k)} = 0; \quad k = 0, \ldots, N_j - 1, \; j = 1, \ldots, J - 1. \tag{3.11}$$

From (3.8) and (3.11), we obtain the mass balance:

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

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

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

$$\begin{aligned} \mathrm{var}(\mathbf{Y}_j) = \mathcal{W}_j + \mathcal{W}_{j-1} \otimes \left(\mathbf{1}_{n_j} \mathbf{1}'_{n_j}\right) + \cdots + \mathcal{W}_1 \otimes \left(\mathbf{1}_{n_2 \ldots n_{j-1}} \mathbf{1}'_{n_2 \ldots n_{j-1}}\right) \\ + \sigma_1^2 \mathbf{I}_{N_1} \otimes \left(\mathbf{1}_{n_1 \ldots n_{j-1}} \mathbf{1}'_{n_1 \ldots n_{j-1}}\right), \end{aligned} \tag{3.12}$$

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

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

Also note that it is not always possible to achieve mass balance with the statistical model (3.8) and (3.9) 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 (3.8) is not reasonable, given a large heterogeneity of variances. Furthermore, there are other choices of covariance matrices $\left\{\mathbf{W}_{\mathbf{ch}(j,k)}\right\}$ that can be used to produce mass balance. The one we choose for a heterogeneous, mass-balanced, tree-structured model given by (3.8) and (3.9) reduces to that for a homogeneous, mass-balanced, tree-structured model given by (3.4) and (3.5), which is an attractive property, and (3.12) has the same form as (3.7). Obtaining the best choice for these covariance matrices is a topic for future research.

In general, the variance parameters in $\mathbf{W}_{\mathbf{ch}(j,k)}$,

$$V_{j,k} \equiv \frac{1}{|D_{j,k}|^2} \int_{D_{j,k}} \int_{D_{j,k}} C(\mathbf{s}, \mathbf{s}') d\mathbf{s} d\mathbf{s}'; \quad k = 0, \ldots, N_j - 1, \; j = 1, \ldots, J, \tag{3.13}$$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

are given in terms of $C(\mathbf{s}, \mathbf{s}') \equiv \text{cov}(Y(\mathbf{s}), Y(\mathbf{s}'))$. Maximum likelihood estimation of $\{V_{j,k}\}$ can be obtained by assuming a parametric model for the $\{V_{j,k}\}$, but it is usually computationally difficult (e.g., the M-step of the EM algorithm does not generally have a closed form, except for homogeneous, mass-balanced, tree-structured models). In Section 4, we propose an approach for estimating $\{V_{j,k}\}$ by first estimating $C(\cdot, \cdot)$ using covariograms on smaller disjoint regions, then "plugging in" the estimated function $\hat{C}(\cdot, \cdot)$ into (3.13).

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. Subsequently, 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 data product that is released widely to the scientific community ("level 3"). The level-3 data product uses 1 degree latitude by 1.25 degree longitude ($1^\circ \times 1.25^\circ$) equiangular grid cells (McPeters et al. 1996, p. 44).

There are several approaches that have been or can be used to handle large volumes of polar-orbiting satellite data. Fang and Stein (1998) used a moving average with seasonal dependence to investigate variations in zonal ozone levels for a fixed latitude. Niu and Tiao (1995) introduced 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) proposed 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, chap. 3), although the disadvantage of the geostatistical method known as kriging is that it does not handle large volumes of data well.

Level-2 TCO and the level-3 data product released by NASA were obtained from the

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Downloaded by [University of Wollongong] at 15:01 29 November 2012

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.

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 Sections 2 and 3, we derive optimal spatial predictions of TCO using a heterogeneous, mass-balanced, tree-structured model; see Section 4.1 and Section 4.2. In Section 4.3, 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). The amount of EOS data to be processed is massive. For example, 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

Using the notation of Section 3, consider a multiresolution tree structure with $N_1 = 40$ nodes at the first resolution and $J = 5$ resolutions. There are $n_1 = 9$ children of each of the 40 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, $N_2 = 360$, $N_3 = 3{,}240$, $N_4 = 12{,}960$, and $N_5 = 51{,}840$. We use equiangular grid cells based on latitude and longitude, for all five resolutions. In all that is to follow, we use the convention that negative latitudes correspond to the Southern Hemisphere and positive latitudes to the Northern Hemisphere.

For each resolution $j = 1, \ldots, 5$, the grid cells $(j, 0), \ldots, (j, N_j - 1)$ 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 (in units of degrees)

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

$[x]$ 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 $j = 1, \ldots, 5$. Therefore, for resolution $j = 1, \ldots, 5$, a consecutive sequence of grid cells starting from the South Pole and finishing at the North Pole is given by $(j, 0), \ldots, \left(j, 5(N_j/N_1)^{1/2} - 1\right)$. 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 ozone data falling in that grid cell, and let $m_{5,k} \equiv |\mathbf{Z}_{5,k}|$ be the dimension of $\mathbf{Z}_{5,k}$; $k = 0, \ldots, 51{,}839$. To start the change-of-resolution

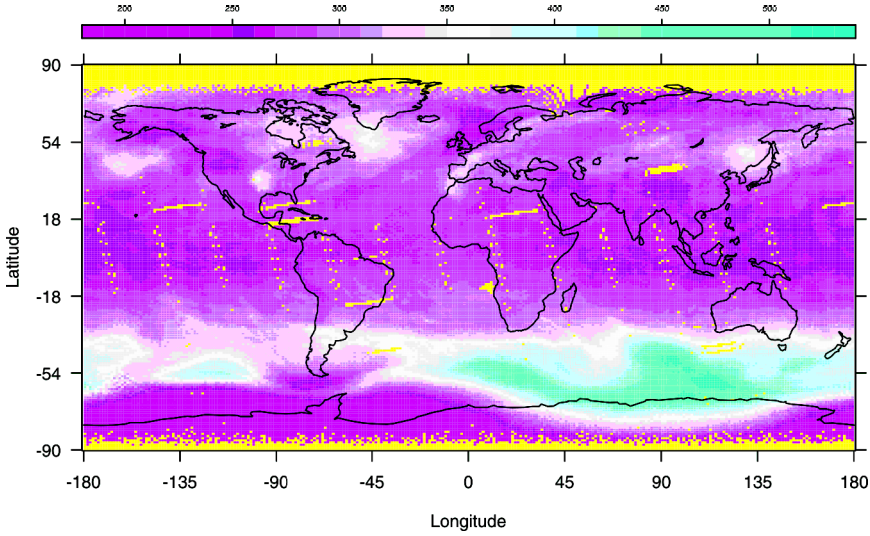


Figure 3. Spatially regular preliminary TCO data.

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 $\mathrm{cor}(\mathbf{Z}_{5,k})$, is determined by the proportions of areal overlaps among the level-2 observations within each grid cell $(5,k)$; $k = 0, \ldots, 51{,}839$. Hence, the spatially regular preliminary data are,

$$z_{5,k} \equiv \frac{\mathbf{1}'_{m_{5,k}} \left(\mathrm{cor}(\mathbf{Z}_{5,k})\right)^{-1} \mathbf{Z}_{5,k}}{\mathbf{1}'_{m_{5,k}} \left(\mathrm{cor}(\mathbf{Z}_{5,k})\right)^{-1} \mathbf{1}_{m_{5,k}}}; \quad m_{5,k} > 0\,,\ k = 0, \ldots, 51839. \tag{4.1}$$

Figure 3 shows the spatially regular preliminary data $\{z_{5,k}\}$, based on a Mercator projection of the globe. Notice that there are missing values for these data, corresponding to grid cells within which no level-2 observation fell on that day (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).

Before applying the change-of-resolution Kalman filter, the data $\{z_{5,k}\}$ have to be detrended to satisfy an assumption of zero mean. We use natural splines and periodic splines to model the trend (see Schumaker 1981). A set of natural cubic spline basis functions defined on $[-90, 90]$ with knots at $\{0, \pm 10, \pm 20, \pm 30, \pm 40, \pm 45, \pm 50, \pm 55, \pm 60, \pm 70, \pm 80, \pm 90\}$ has dimension 23, and is denoted by $\{f_1(\cdot), \ldots, f_{23}(\cdot)\}$. Note that more knots are taken in mid latitudes, where spatial variation is higher. Further, a set of cubic periodic spline basis functions (see Schumaker 1981, sec. 8.1) defined on $[-180, 180]$ with knots at $\{0, \pm 20, \pm 40, \pm 60, \pm 80, \pm 100, \pm 120, \pm 140, \pm 160, \pm 180\}$ has dimension 18, and is denoted by $\{g_1(\cdot), \ldots, g_{18}(\cdot)\}$. The trend function is assumed to be of the following form:

$$\mu(x,y) = \sum_{i=1}^{23} \beta_i f_i(y) + \sum_{j=1}^{18} \sum_{k=1}^{23} \beta_{j,k}\, g_j(x) f_k(y);$$

$$x \in [-180, 180],\ y \in [-90, 90],$$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

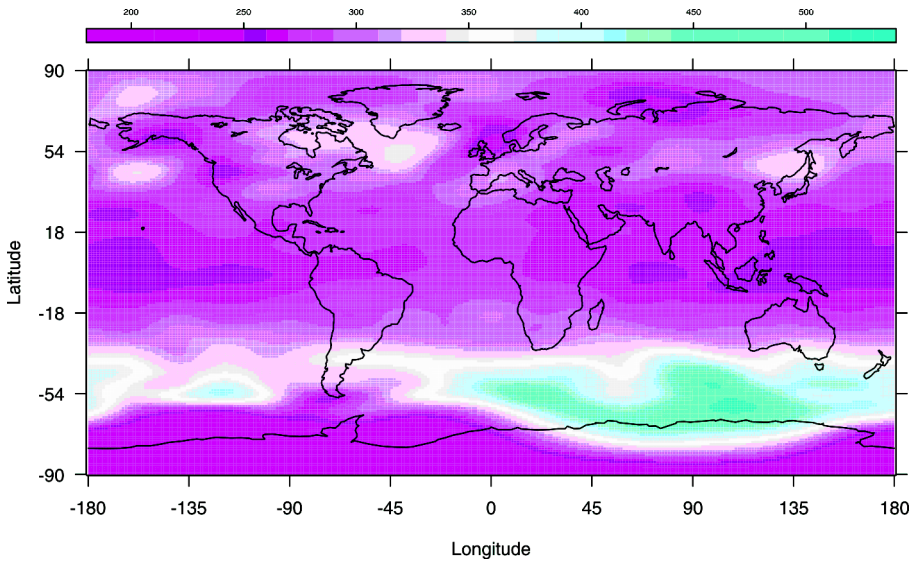


Figure 4. Trend of the TCO data.

with the constraints that $\sum_{j=1}^{18} \beta_{j,k} f_k(\pm 90) = 0$, for $k = 1, \ldots, 23$, and $\sum_{k=1}^{23} \beta_{j,k}\, g_j(180) = 0$, for $j = 1, \ldots, 18$. Note that $g_j(180) = g_j(-180)$, for $j = 1, \ldots, 18$. Therefore, $\mu(\cdot, \cdot)$ has the property that $\mu(x, 90)$ and $\mu(x, -90)$ do not depend on x, and $\mu(-180, y) = \mu(180, y)$, for all $y \in [-90, 90]$. Note that there are $23 + 17 * 21 = 380$ parameters to estimate in $\mu(x, y)$.

We estimate $\mu(\cdot, \cdot)$ using the aggregated data $\{z_{3,0}, \ldots, z_{3,3239}\}$ at resolution 3 by minimizing the weighted least squares:

$$\sum_{k=0}^{3239} \frac{1}{c_{3,k}} (z_{3,k} - \mu(x_{3,k}, y_{3,k}))^2,$$

where

$$z_{3,k} \equiv \frac{\displaystyle\sum_{k' : D_{5,k'} \subset D_{3,k}, m_{5,k'} \geq 1} a_{5,k'} z_{5,k'}}{\displaystyle\sum_{k' : D_{5,k'} \subset D_{3,k}, m_{5,k'} \geq 1} a_{5,k'}},$$

$$c_{3,k} \equiv \frac{\displaystyle\sum_{k' : D_{5,k'} \subset D_{3,k}, m_{5,k'} \geq 1} \left(\mathbf{1}'_{m_{5,k'}} \left(\mathrm{cor}(\mathbf{Z}_{5,k'}) \right)^{-1} \mathbf{1}_{m_{5,k'}} \right)^{-1} a_{5,k'}^2}{\left(\displaystyle\sum_{k' : D_{5,k'} \subset D_{3,k}, m_{5,k'} \geq 1} a_{5,k'} \right)^2},$$

and $(x_{3,k}, y_{3,k}) \equiv (i_{3,k} + 2.5,\ l_{3,k} + 2)$ is the centroid of the grid cell between longitudes $i_{3,k}$ and $i_{3,k} + 5$ and between latitudes $l_{3,k}$ and $l_{3,k} + 4$. The resulting estimate is denoted by $\hat{\mu}(\cdot, \cdot)$; see Figure 4.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

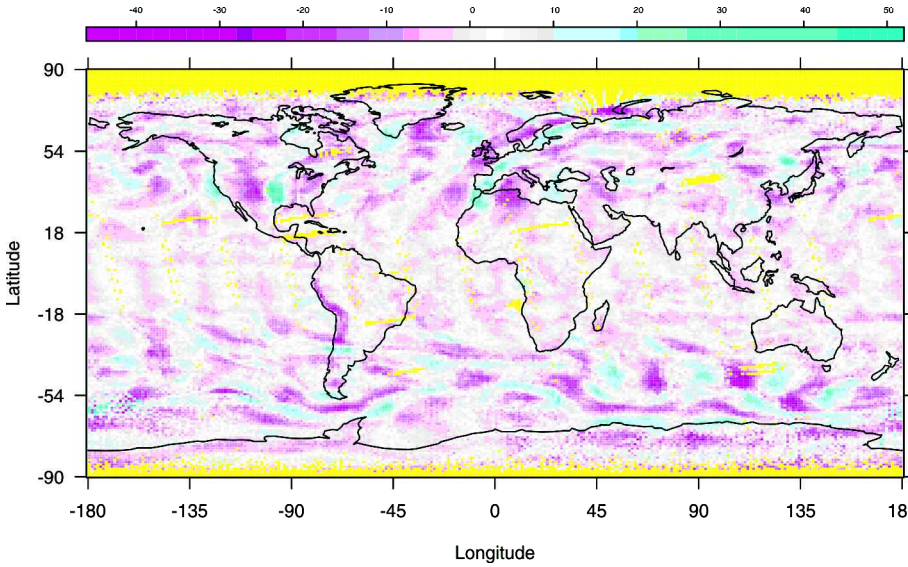


Figure 5. Spatially regular residual TCO defined by (4.2).

Our spatial analysis of the TOMS data proceeds on the spatially regular, detrended residual TCO values (see Figure 5):

$$z^*_{5,k} = z_{5,k} - \hat{\mu}(x_{5,k}, y_{5,k}), \tag{4.2}$$

on a given day (here, October 1, 1988), where $(x_{5,k}, y_{5,k}) \equiv (i_{5,k} + .625,\ l_{5,k} + .5)$. This trend correction allows us to assume that (4.2) has zero mean, which is an important assumption of the tree-structured models given in Section 3.

For optimal spatial predictions, we shall apply the heterogeneous, mass-balanced, tree-structured model of Section 3.4, where the variance parameters $\{V_{j,k}\}$ and $\{\phi_{j,k}\}$ for the hidden state variables and the measurement errors, respectively, are estimated using an approach described in Section 4.2.

4.2 Estimation of Variance Parameters

Under the heterogeneous, mass-balanced, tree-structured model given by (3.8) and (3.9), the parameters to be estimated are $\{\phi_{5,k}\}$ and $\{V_{j,k}\}$, where $V_{j,k}$ is given by (3.13). Now $\{V_{j,k}\}$ can be estimated by plugging into (3.13) an estimated spatial covariance function of $C(\cdot,\cdot)$, where recall that $C(\mathbf{s},\mathbf{s}') \equiv \text{cov}(Y(\mathbf{s}), Y(\mathbf{s}'))$.

The spatial covariance function $C(\cdot,\cdot)$ is expected to be nonstationary, but is approximately stationary in each of the 40 resolution-1 regions $\{D_{1,0}, \ldots, D_{1,39}\}$. Assume that the covariance functions for regions $D_{1,l}$; $l = 0, \ldots, 39$, are members of a family of anisotropic exponential covariograms. That is, for $\mathbf{s}, \mathbf{s}' \in D_{1,l}$; $l = 0, \ldots, 39$,

$$C(\mathbf{s},\mathbf{s}') = C_l(\mathbf{s}-\mathbf{s}') \equiv \beta_l \exp(-\|\mathbf{\Lambda}_l(\mathbf{s}-\mathbf{s}')\|),$$

where $\mathbf{\Lambda}_l$ is a 2×2 diagonal matrix with diagonal elements $\lambda_{l,1}$ and $\lambda_{l,2}$. The corresponding

Downloaded by [University of Wollongong] at 15:01 29 November 2012

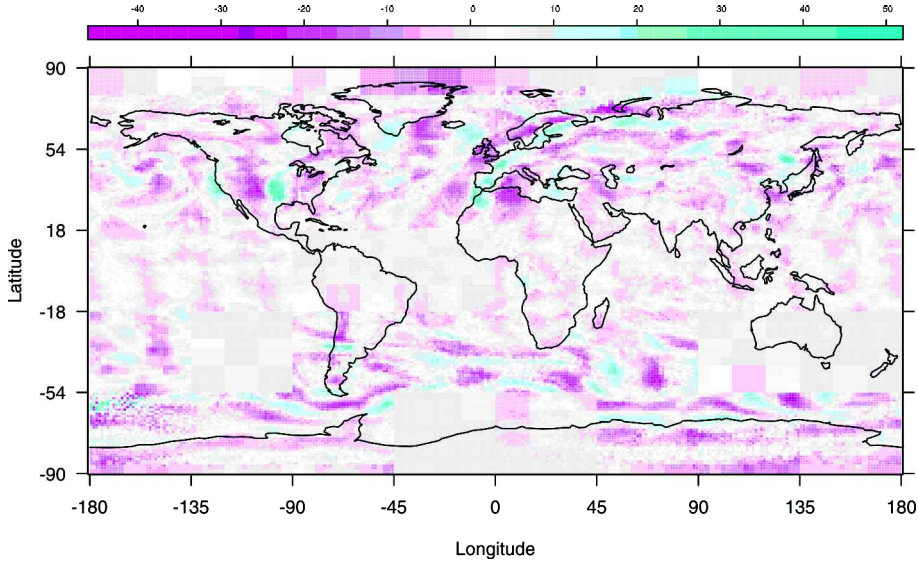


Figure 6. Kalman-filter predicted residual TCO at resolution 5.

variogram of $Z(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s})$, for $\mathbf{s} \in D_{1,l}$, is given by

$$2\gamma_l(\mathbf{h}) \equiv 2\,\alpha_l I(\mathbf{h} \neq \mathbf{0}) + 2\,\beta_l\left(1 - \exp(-\|\mathbf{\Lambda}_l\mathbf{h}\|)\right); \quad \mathbf{h} \in \mathbb{R}^2,\ l = 0, \ldots, 39,$$

where, for each $l = 0, \ldots, 39$, $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in D_{1,l}\}$ is a white-noise process with variance α_l, representing measurement errors. For each $l = 0, \ldots, 39$, we estimate $\{\alpha_l, \beta_l, \lambda_{l,1}, \lambda_{l,2}\}$ by fitting the variogram $2\gamma_l(\cdot)$, based on a large sample of the detrended level-2 data observed in region $D_{1,l}$. To ensure that both small-scale and large-scale information is incorporated, a stratified systematic sampling method is applied. First, each region $D_{1,l}$ is divided into 6×6 equiangular subregions (strata), each of which consists of 3×3 resolution-4 grid cells. Then all the data observed in each of the center cells of the 36 subregions define the sample for the region. The variogram was fitted using the method of weighted-least-squares (Cressie 1985), yielding estimators $\{\hat{\alpha}_l, \hat{\beta}_l, \hat{\lambda}_{l,1}, \hat{\lambda}_{l,2}\}$.

This results in a covariogram estimate

$$\hat{C}_l(\mathbf{h}) = \hat{\beta}_l \exp(-\|\hat{\mathbf{\Lambda}}_l\mathbf{h}\|); \quad \mathbf{h} \in \mathbb{R}^2,$$

where $\hat{\mathbf{\Lambda}}_l$ is a 2×2 diagonal matrix with diagonal elements $\hat{\lambda}_{l,1}$ and $\hat{\lambda}_{l,2}$. Therefore, $V_{j,k}$ can be estimated by

$$\hat{V}_{j,k} = \frac{1}{|D_{j,k}|^2}\int_{D_{j,k}}\int_{D_{j,k}} \hat{C}_{l(j,k)}(\mathbf{s} - \mathbf{s}')d\mathbf{s}d\mathbf{s}'; \quad k = 0, \ldots, N_j - 1,\ j = 1, \ldots, J,$$

where $l(j,k) \in \{0, \ldots, 39\}$ is such that $D_{j,k} \subset D_{1,l(j,k)}$. From (4.1), the measurement-error variances $\{\phi_{5,k}\}$ can be estimated by

$$\hat{\phi}_{5,k} = \hat{\alpha}_{l(5,k)}\left(\mathbf{1}'_{m_{5,k}}\left(\mathrm{cor}(\mathbf{Z}_{5,k})\right)^{-1}\mathbf{1}_{m_{5,k}}\right)^{-1}; \quad k = 0, \ldots, 51839.$$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

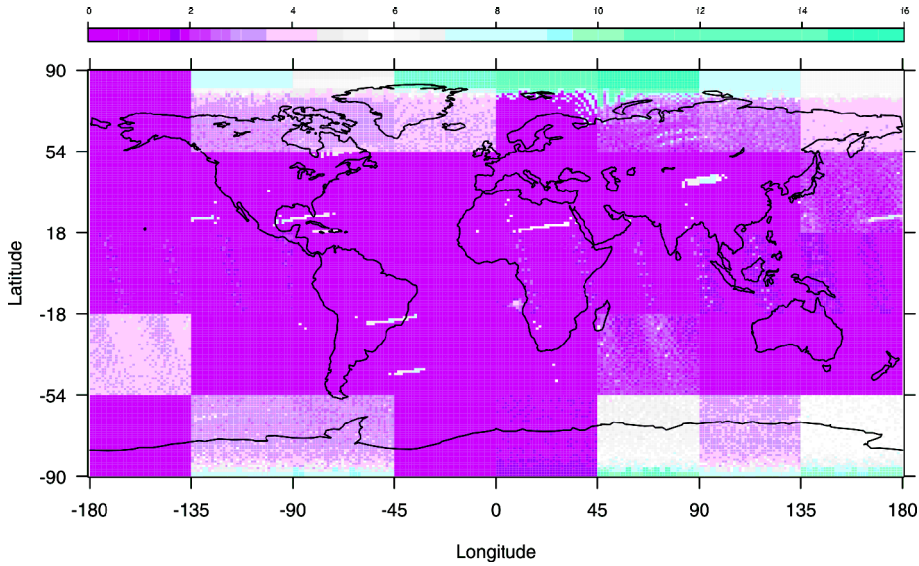


Figure 7. Prediction standard errors for predicted residual TCO shown in Figure 5.

4.3 Optimal Spatial Prediction of TCO

We can now apply the heterogeneous, mass-balanced, tree-structured model given by (3.8) and (3.9) to residuals $\{z^*_{5,k}\}$ given by (4.2), using the estimated measurement-error variances $\{\hat{\phi}_{5,k}\}$ and the estimated state-process variances $\{\hat{\mathbf{W}}_{\mathbf{ch}(j,k)}\}$ (obtained from $\{\hat{V}_{j,k}\}$ using (3.10) and (A.1)). Optimal spatial predictions can now be obtained by recursively applying the leaves-to-root Kalman-filtering algorithm (Equation (2.5) through Equation (2.12)) and, after the roots have been reached, by recursively applying the root-to-leaves Kalman-smoothing algorithm (Equation (2.13) and Equation (2.14)). The predicted residual TCO values based on the change-of-resolution Kalman filter are shown in Figure 6 with the corresponding prediction standard errors shown in Figure 7, based on a Mercator projection of the globe.

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

$$\text{level-3 TCO value} = \text{trend} + \text{predicted residual TCO value}\,,$$

where the trend is the same $\hat{\mu}(\cdot,\cdot)$ used in (4.2). Figure 8 shows the Kalman-filter level-3 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, and it takes about 45 minutes to implement the prediction algorithm using an S-Plus program on a Pentium-III-800MHz personal computer. Aggregations up to the coarser resolutions are obtained by area-weighted averaging of the level-3 TCO values, thus preserving mass balance of the predictors shown in Figure 8.

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 $1^\circ \times 1.25^\circ$

Downloaded by [University of Wollongong] at 15:01 29 November 2012

Downloaded by [University of Wollongong] at 15:01 29 November 2012

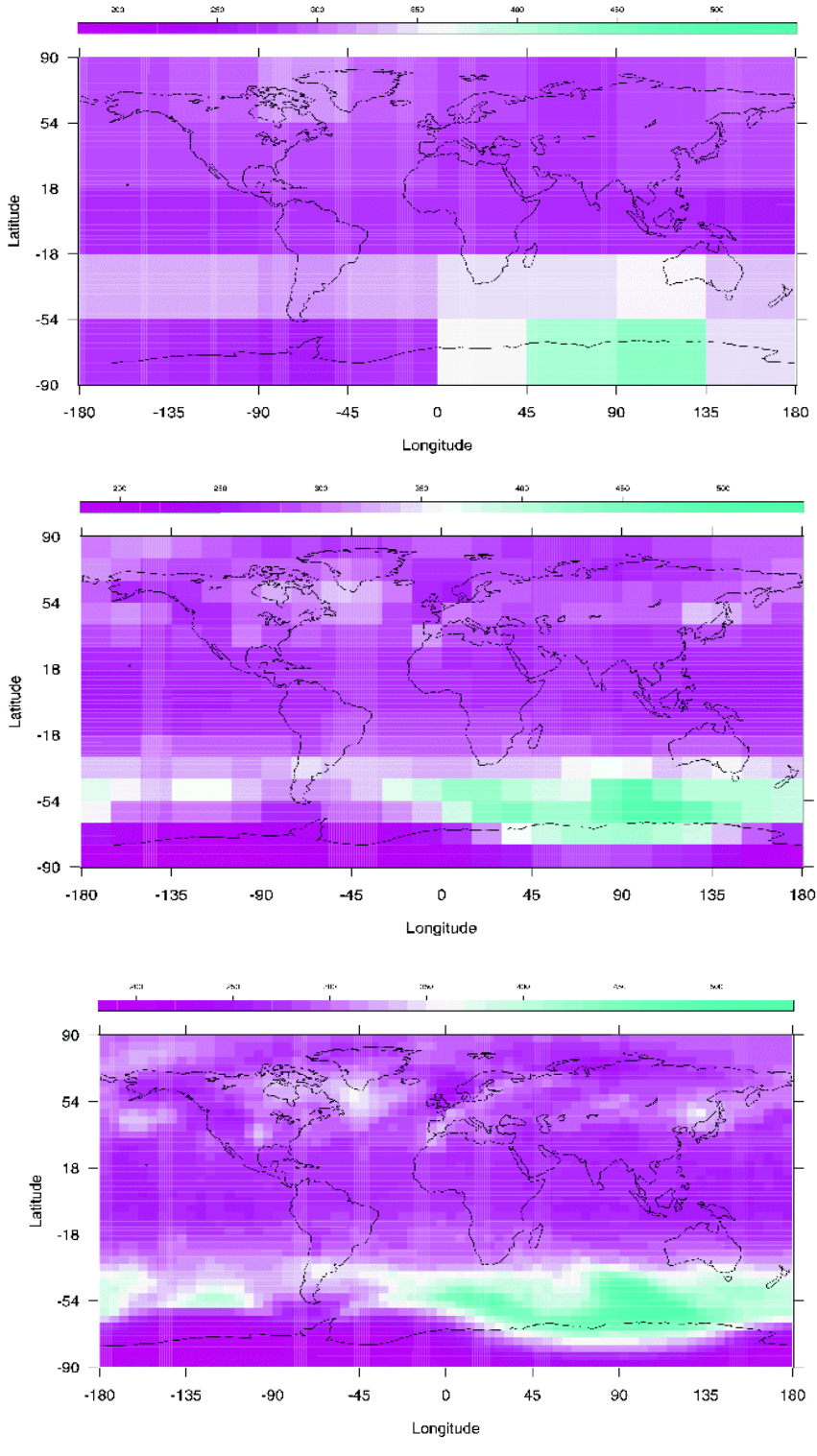


Figure 8a. Predicted TCO for resolutions 1 (top) to 3 (bottom).

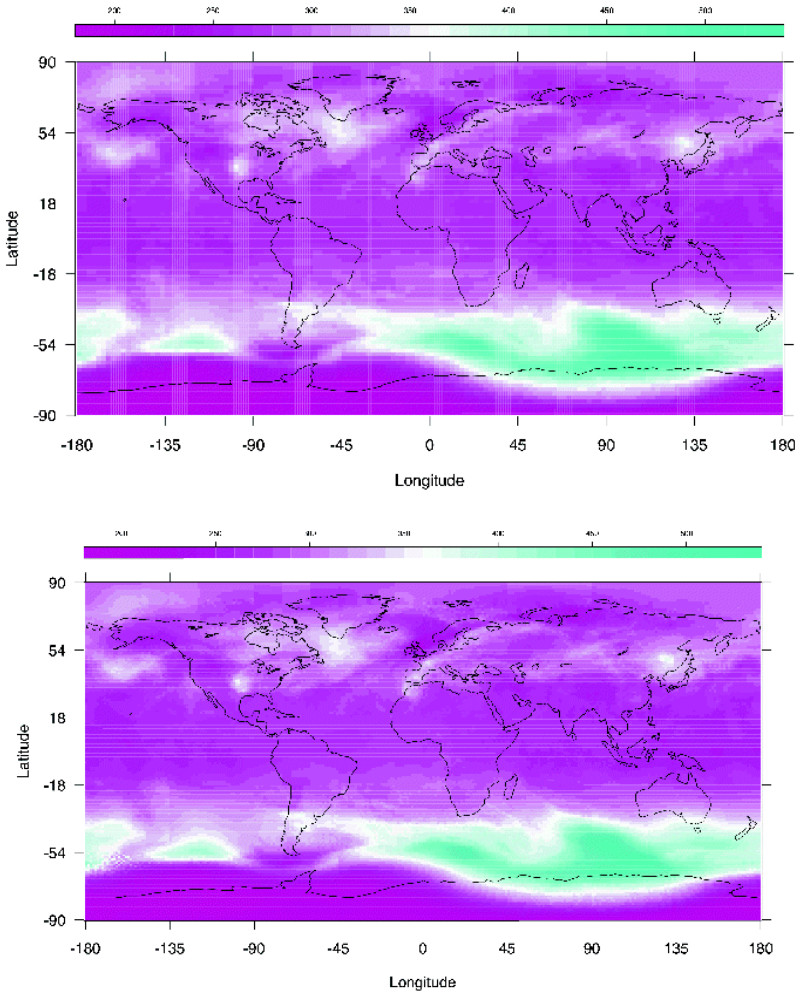


Figure 8b. Predicted TCO for resolutions 4 (top) to 5 (bottom).

Downloaded by [University of Wollongong] at 15:01 29 November 2012

grid cells for October 1, 1988. Then the mean squared error (MSE) for L3 is calculated as:

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

where $\mathrm{GS}(l)$ represents the TCO reading for the lth ground station for October 1, 1988. The NASA level-3 data product achieves a MSE of 146.08, compared to 140.08 for the level-3 data product from the heterogeneous, mass-balanced, tree-structured model. A small reduction of 3.79% 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.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

5. DISCUSSION

We have presented a new methodology that is extended from previous work on multiresolution autoregressive tree-structured models for fast spatial prediction. It is based on a new model for which the state process is resolution-consistent ("mass balance"), an important physical property for many environmental variables. The advantages of this 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, allowing us to handle massive amounts of possibly irregularly sampled data.

One drawback of Kalman filtering on trees is that the implied spatial covariance function is piecewise constant and nonstationary (see Equation (3.7) and Equation (3.12)), 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 multiresolution models like those used for mapping total column ozone over the globe, the overall covariance shape is approximately stationary and exponential within a resolution-1 region, 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 a coarse resolution, as a multivariate autoregressive process in time and to retain the tree structure at each time point. This result is again a tree. We could then run time backwards, from the coarse resolution 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.

APPENDIX: CONSTRUCTION OF A COVARIANCE MATRIX FOR MASS BALANCE

We prove a technical result that can be used to ensure (3.11) for heterogeneous, mass-balanced, tree-structured models. For $\mathbf{a} \in \mathbb{R}^n$, denote $\mathrm{diag}(\mathbf{a})$ to be the diagonal matrix

with diagonal entries given by $\mathbf{a}$. Let

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

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

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

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

$$\mathbf{U} \equiv \left(u_{i,j}\right)_{n\times n} \equiv \left(diag(\mathbf{a})\right)^{-1}\mathbf{F}_n diag(\mathbf{c})\mathbf{F}_n\left(diag(\mathbf{a})\right)^{-1}, \tag{A.1}$$

then the following statements hold:

1. $\mathbf{a}'\mathbf{U}\mathbf{a} = 0$,
2. $u_{i,i} = \sigma_i^2; \quad i = 1,\ldots,n$,
3. $\mathbf{U}$ *is non-negative definite if and only if*

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

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

2. We have

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

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

3. From (A.1), we know that $\mathbf{U}$ is non-negative definite if and only if $c_i \geq 0$; $i = 1,\ldots,n$. Now for each $i = 1,\ldots,n$, 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.

□

ACKNOWLEDGMENTS

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. The authors thank Ralph Kahn

Downloaded by [University of Wollongong] at 15:01 29 November 2012

for help in obtaining the TOMS data, and Kahn and Amy Braverman for helpful discussions regarding spatial prediction from satellite data.

[Received July 2000. Revised November 2001.]

REFERENCES

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. (1985), "Fitting Variogram Models by Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, 17, 563–586.

——— (1993), *Statistics for Spatial Data*, revised edition, New York: Wiley.

Daoudi, K., Frakt, A. B., and Willsky, A. S. (1999), "Multiscale Autoregressive Models and Wavelets," *IEEE Transactions on Information Theory*, 45, 828–845.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: SIAM.

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.

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.

Frakt, A., and Willsky, A. (1998), "Multiscale Autoregressive Models and the Stochastic Realization Problem," *Asilomar Conference on Signals, Systems, and Computers*, Asilomar, CA.

Huang, H.-C., and Cressie, N. (1997), "Multiscale Spatial Modeling," in *1997 Proceedings of the Section on Statistics and the Environment*, Alexandria, VA: American Statistical Association, pp. 49–54.

——— (2001), "Multiscale Graphical Modeling in Space: Applications to Command and Control," in *Spatial Statistics: Methodological Aspects and Some Applications*, ed. M. Moore, Springer Lecture Notes in Statistics, vol 159, New York: Springer, pp. 83–113.

Kahn, R. (1996), "What Shall We Do With the Data We Are Expecting in 1998?" in *Massive Data Sets. Proceedings of a Workshop*, Washington, DC: National Academy of Sciences, pp. 15–21.

——— (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.

Lauritzen, S. L. (1992), "Propagation of Probabilities, Means and Variances in Mixed Graphical Association Models," *Journal of the American Statistical Association*, 87, 1098–1108.

Lauritzen, S. L., and Spiegelhalter, D. J. (1988), "Local Computations With Probabilities on Graphical Structures and Their Application to Expert System" (with discussion), *Journal of the Royal Statistical Society*, Series B, 50, 157–224.

London, J. (1985), "The Observed Distribution of Atmospheric Ozone and its Variations," *Ozone in the Free Atmosphere*, eds. R. C. Whitten and S. S. Prasad, New York: Van Nostrand Reinhold Company, 11–80.

Luettgen, M. R., and Willsky, A. S. (1995a), "Multiscale Smoothing Error Models," *IEEE Transactions on Automatic Control*, 40, 173–175.

——— (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*, Greenbelt, MD: NASA.

Downloaded by [University of Wollongong] at 15:01 29 November 2012

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*, Greenbelt, MD: NASA.

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

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

Niu, X., and Tiao, G. (1995), "Modelling Satellite Ozone Data," *Journal of the American Statistical Association*, 90, 969–983.

Schumaker, L. L. (1981), *Spline Functions: Basic Theory*, New York: Wiley.

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.

Downloaded by [University of Wollongong] at 15:01 29 November 2012