Data Mining of MISR Aerosol Product using Spatial Statistics

Tao Shi

Department of Statistics

The Ohio State Univeriuty

Columbus, CA 43210

Email: taoshi@stat.ohio-state.edu

Noel Cressie

Department of Statistics

The Ohio State Univeriuty

Columbus, CA 43210

Email: ncressie@stat.ohio-state.edu

Abstract— In climate models, aerosol forcing is the major source of uncertainty in climate forcing, over the industrial period. To reduce this uncertainty, instruments on satellites have been put in place to collect global data. However, missing and noisy observations impose considerable difficulties for scientists researching global aerosol distribution, aerosol transportation, and comparisons between satellite observations and global-climate-model outputs. In this paper, we propose a Spatial Mixed Effects (SME) statistical model to predict the missing values, denoise the observed values, and quantify the spatial-prediction uncertainties. The computations associated with the SME model are linear scalable to the number of data points, which makes it feasible to process massive global satellite data. We apply our proposed methodology, which we call Fixed Rank Kriging (FRK), to the level-3 Aerosol Optical Depth dataset collected by NASA's Multi-angle Imaging SpectroRadiometer (MISR) instrument flying on the Terra satellite. Overall, our results were superior to those from nonstatistical methods and, importantly, FRK has an uncertainty measure associated with it.

I. INTRODUCTION

Estimates from the 2001 IPCC (Intergovernmental Panel on Climate Change) Report show that aerosol (liquid or solid airborne particulates having sizes ranging from more than 0.1 μm to less than 1 μm) forcing contains the major source of uncertainty in climate forcing, over the industrial period. Calculations of aerosol effects on climate require the use of chemical-transport, energy-radiation, and general-circulation models. The validity of these models needs to be evaluated against aerosol information retrieved from satellites and other data sources. Stratification of Aerosol Optical Depth (AOD) by particle type is a key component of this strategy.

The Terra satellite, launched on December 18, 1999, is part of the Earth Observing System (EOS) sponsored by the National Aeronautics and Space Administration (NASA), and one of the instruments on board is the Multi-angle Imaging SpectroRadiometer (MISR). The MISR instrument retrieves key aerosol properties, such as aerosol optical depth, as well as aerosol shape and size. The global aerosol information provided by MISR is vital to our understanding of geophysical processes governing global changes of Earth's climate, including scattering and absorption of solar radiation and other parameters (Diner et al. [5] and Kaufman et al. [13]).

MISR cameras cover a swath at the Earth's surface that is approximately 360 km wide and extends across the daylight side of the Earth from the Arctic down to Antarctica. There are 233 geographically distinct, but overlapping, swaths, which are also called paths. The MISR instrument collects data from all paths on a repeat cycle of 16 days; that is, it covers the exact same path every 16 days. The upper panel in Fig. 1 shows a global map of the locations where MISR had successful retrieval of AOD on one particular day (April 1, 2001). Each colored strip represents data retrieval from one MISR orbit. The gray areas are either where there was no coverage on that day, or where the aerosol algorithm failed to retrieve AOD. The spatial resolution of MISR level-2 aerosol data is $17.6km$ by $17.6km$, which are converted to level-3 data at a much lower spatial (0.5 $degree$ by 0.5 $degree$) and temporal resolution (1 day) by averaging those observations falling in lower-resolution pixels over a certain period of time. The upper panel in Fig. 1 shows the level-3 AOD data on April 1, 2006.

Level-3 data are not nearly as massive, but they are still sparse in different regions of the globe. For example, the middle panel in Fig. 1 shows a MISR level-3 AOD product (0.5 $degree$ by 0.5 $degree$) for the period of April 1, 2001 to April 16, 2001, which spans a complete MISR repeat-cycle. The plot is a global map of the averaged AOD values, where the averages are taken pixel-by-pixel over all level-2 data successfully retrieved in each 0.5 $degree$ by 0.5 $degree$ pixel during those 16 days. Even for this low spatial and temporal resolution, there are pixels with no data (gray pixels on the map) that cover the poles and leave holes over other parts of the mid-latitudes (e.g., over South America).

These missing data create great difficulties for those doing research on global aerosol distribution and transportation, on comparison between satellite observations and the global-climate-model outputs, and on other related climate studies. In this paper, our goal is to fill in the missing data and to denoise the existing data, at level 3, in a statistically optimal way. This involves an initial modeling of the spatial trend, followed by a spatial-prediction method called *kriging*.

Kriging, or spatial best linear unbiased prediction (spatial BLUP), has become very popular in the earth and environmental sciences, where it is sometimes known as optimum interpolation. Kriging methodology is able to produce maps of optimal predictions and associated prediction standard errors
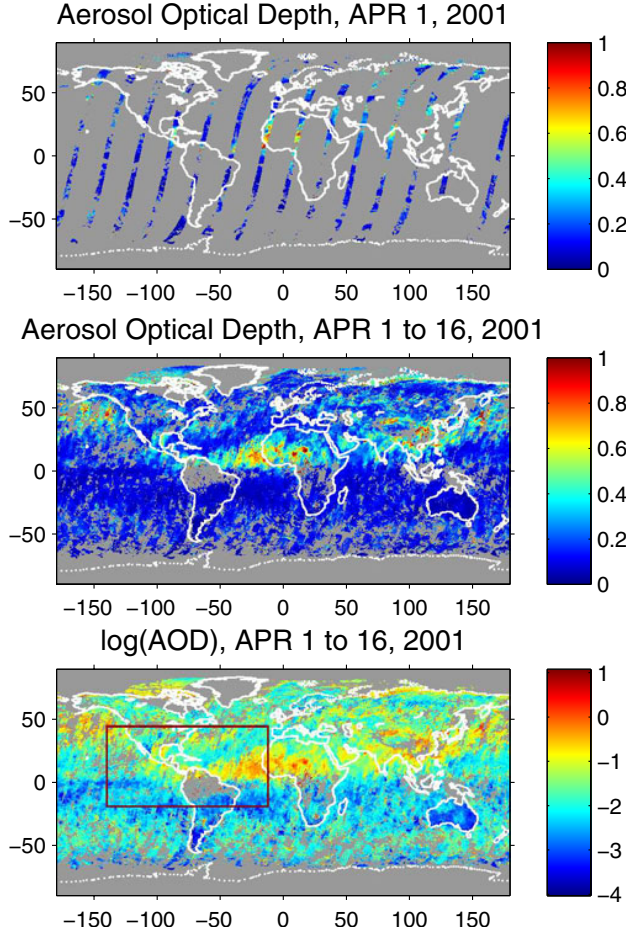
Fig. 1. MISR Aerosol Optical Depth observations with gray areas representing unobserved pixels. Upper panel: level-3 AOD data for April 1, 2001; middle panel: level-3 AOD data for April 1 through April 16, 2001; lower panel: level-3 log(AOD) data for April 1 through April 16, 2001.

from incomplete and noisy spatial data (e.g., Cressie [2], Ch. 3). However, solving the kriging equations directly involves inversion of an $n \times n$ covariance matrix Σ, where n data may require $O(n^3)$ computations to obtain Σ^{-1}. Under these circumstances, straightforward kriging on global MISR AOD data is impossible.

In recent years, various methods have been proposed to approximate the kriging equations (Barry and Pace [1]; Nychka [17]; Nychka et al. [18]; Kammann and Wand [12]; Furrer et al. [7]). Suggestions include giving an equivalent representation in terms of orthogonal bases and truncating the bases, doing covariance tapering, using approximate iterative methods such as conjugate-gradient, or replacing the data locations with a smaller set of space-filling locations.

Instead of approximating the kriging equations, we cloud look for classes of covariance functions for which kriging can be done exactly, even though the spatial datasets are large (e.g., Huang et al. [10]; Tzeng et al. [19]; Johannesson et al. [11]). In this article, we use a spatial covariance function

based on what we call a Spatial Random Effects (SRE) model, which leads to a Spatial Mixed Effects (SME) model for the data process. In our application, we use multi-resolution basis functions to capture the spatial dependence in the data. The kriging computations that follow from the SME model can be carried out using Fixed Rank Kriging (FRK), proposed in Cressie and Johannesson [4]. It was shown there that FRK is linear scalable in the number of data, so it is capable of handling the very large datasets associated with the MISR instrument.

In Section 2, we discuss the SME model and review the FRK computations associated with it. In Section 3, we propose a procedure to select basis functions for the SME model, from a family of multi-resolution W-wavelets. The procedure is based on Huang and Cressie's [9] approach of decomposing deterministic and stochastic terms such that large absolute wavelet coefficients are selected for use as either deterministic or stochastic components. In Section 4, we apply the SME model, with selected W-wavelet basis functions, to the MISR level-3 global AOD data, and we compare the result to other commonly used non-statistical methods, such as Inverse Distance Weighting and Nearest Neighbors Smoothing. Section 5 concludes the paper with discussion about future research, including incorporating spatio-temporal variability into the optimal predictor.

II. SPATIAL MIXED EFFECTS (SME) MODEL

Consider a real-valued spatial process $\{Y(\mathbf{s}) \colon \mathbf{s} \in D \subset \mathbb{R}^d\}$, for which we are interested in making inference based on data from $Z(\cdot)$ that have measurement error incorporated:

$$Z(\mathbf{s}) \quad \equiv \quad Y(\mathbf{s}) + \varepsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (1)$$

where $\{\varepsilon(\mathbf{s}) \colon \mathbf{s} \in D\}$ is a spatial white-noise process with mean 0 and, for $\sigma^2 > 0$, $\mathrm{var}(\varepsilon(\mathbf{s})) = \sigma^2 v(\mathbf{s}) \in (0, \infty); \mathbf{s} \in D$.

The hidden process $Y(\cdot)$ is assumed to have a linear mean structure,

$$Y(\mathbf{s}) \quad = \quad \mathbf{T}(\mathbf{s})'\boldsymbol{\beta} + \nu(\mathbf{s}); \quad \mathbf{s} \in D, \quad (2)$$

where $\mathbf{T}(\cdot) \equiv (T_1(\cdot), \ldots, T_p(\cdot))'$ represents a vector process of known covariates; the coefficients $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_p)'$ are unknown; and the process $\nu(\cdot)$ has zero mean and a (generally nonstationary) spatial covariance function,

$$\mathrm{cov}(\nu(\mathbf{u}), \nu(\mathbf{v})) \quad \equiv \quad C(\mathbf{u}, \mathbf{v}); \quad \mathbf{u}, \mathbf{v} \in D. \quad (3)$$

Henceforth, we assume that $\nu(\cdot)$ follows a *Spatial Random Effects* (SRE) model, which is:

$$\nu(\mathbf{s}) \equiv \mathbf{S}(\mathbf{s})'\boldsymbol{\eta},$$

where $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \ldots, S_r(\cdot))'$ represents a set of r basis functions and $\boldsymbol{\eta} \equiv (\eta_1, \ldots, \eta_r)'$ is a zero-mean random vector with covariance matrix given by $\mathbf{K}_{r \times r}$. Then from (1) and (2), the data process $Z(\cdot)$ is a *Spatial Mixed Effects* (SME) model; that is,

$$Z(\mathbf{s}) \quad \equiv \quad \mathbf{T}(\mathbf{s})'\boldsymbol{\beta} + \mathbf{S}(\mathbf{s})'\boldsymbol{\eta} + \varepsilon(\mathbf{s}); \quad \mathbf{s} \in D, \quad (4)$$

in which $\mathbf{T}(\mathbf{s})'\boldsymbol{\beta}$ and $\mathbf{S}(\mathbf{s})'\boldsymbol{\eta}$ model the large- and small-scale spatial variations, respectively, and $\varepsilon(\mathbf{s})$ represents the measurement error in each potential observation.

In fact, the process $Z(\cdot)$ is known only at a finite number of spatial locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$; define the vector of available data to be $\mathbf{Z} \equiv (Z(\mathbf{s}_1), \cdots, Z(\mathbf{s}_n))'$. Our interest is in inference on the hidden Y-process, not the noisy Z-process, using the data $\mathbf{Z}$. In what follows, we consider point prediction, where we wish to predict the Y-process at a location $\mathbf{s}_0$; $\mathbf{s}_0 \in D$, regardless of whether $\mathbf{s}_0$ is or is not an observation location. Block prediction can be generalized straightforwardly for SRE models; see the discussion in Section 6.

A. Fixed Rank Kriging

For the SME model (4), we can write the $n \times n$ theoretical covariance matrix of $\mathbf{Y} \equiv (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))'$ as, $\mathbf{C} = \mathbf{SKS}'$, and hence the covariance matrix of $\mathbf{Z}$ is:

$$\boldsymbol{\Sigma} = \mathbf{SKS}' + \sigma^2\mathbf{V}, \tag{5}$$

where recall that $\mathbf{K}$ is the $r \times r$ covariance matrix of $\boldsymbol{\eta}$, $\mathbf{S}$ is the $n \times r$ matrix whose (i, ℓ) element is $S_\ell(\mathbf{s}_i)$, and recall that $\sigma^2\mathbf{V}$ is diagonal with entries given by the measurement-error variances.

Cressie and Johannesson [4] studied the kriging computations for this type of covariance model (5), and they named the resulting prediction method Fixed Rank Kriging (FRK). The authors showed that FRK computations only require inversion of $r \times r$ matrices or diagonal $n \times n$ matrices; specifically they showed that $\boldsymbol{\Sigma}^{-1}$ can be expressed as:

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S}\{\mathbf{K}^{-1} + \mathbf{S}'(\sigma^2\mathbf{V})^{-1}\mathbf{S}\}^{-1} \\ &\quad \mathbf{S}'(\sigma^2\mathbf{V})^{-1}. \end{aligned} \tag{6}$$

In general, the kriging predictor can be written as:

$$\widehat{Y}(\mathbf{s}_0) = \mathbf{t}(\mathbf{s}_0)'\widehat{\boldsymbol{\beta}} + \mathbf{c}(\mathbf{s}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{T}\widehat{\boldsymbol{\beta}}),$$

where $\widehat{\boldsymbol{\beta}} \equiv (\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}$ and $\mathbf{c}(\mathbf{s}_0) \equiv (C(\mathbf{s}_0, \mathbf{s}_1), \ldots, C(\mathbf{s}_0, \mathbf{s}_n))'$. For the SRE model, this results in the FRK predictor:

$$\widehat{Y}(\mathbf{s}_0) = \mathbf{t}(\mathbf{s}_0)'\widehat{\boldsymbol{\beta}} + \mathbf{S}(\mathbf{s}_0)'\mathbf{KS}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \mathbf{T}\widehat{\boldsymbol{\beta}}), \tag{7}$$

where $\boldsymbol{\Sigma}^{-1}$ is given by (6). The FRK standard error is:

$$\begin{aligned}\sigma_k(\mathbf{s}_0) &= \{\mathbf{S}(\mathbf{s}_0)'\mathbf{KS}(\mathbf{s}_0) - \mathbf{S}(\mathbf{s}_0)'\mathbf{KS}'\boldsymbol{\Sigma}^{-1}\mathbf{SKS}(\mathbf{s}_0) \\ &\quad + (\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{SKS}(\mathbf{s}_0))'(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1} \\ &\quad (\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{SKS}(\mathbf{s}_0))\}^{1/2}, \end{aligned} \tag{8}$$

where $\boldsymbol{\Sigma}^{-1}$ is again given by (6). As the prediction location $\mathbf{s}_0$ in (7) and (8) varies over D, a kriging-prediction map and a kriging-standard-error map, respectively, are generated.

Closer inspection of the kriging equations (7) and (8) reveals that for a fixed number of regressors p and a fixed rank r of the covariance model defined by (5), the computational burden is only linear in n. More interestingly, instead of storing the $n \times n$ matrix $\boldsymbol{\Sigma}^{-1}$ in computer memory, FRK only needs to deal with much smaller matrices, $\mathbf{T}'\boldsymbol{\Sigma}^{-1}$ ($p \times n$) and $\mathbf{S}'\boldsymbol{\Sigma}^{-1}$ ($r \times n$). Thus, FRK makes it feasible to construct maps of

kriging predictors and kriging standard errors based on very large spatial datasets.

B. Fitting the SME model by Weighted Least Squares

Given data $\mathbf{Z}$ at locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, and functions $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$, fitting the SME model (4) requires estimating $\boldsymbol{\beta}$, σ^2, and $\mathbf{K}$ from the data. We follow the estimation scheme proposed in Cressie and Johannesson [4], based on minimizing the Frobenius norm between an empirical covariance matrix and a theoretical covariance matrix derived from $\boldsymbol{\Sigma} = \mathbf{SKS}' + \sigma^2\mathbf{V}$. We briefly review the method here; interested readers can find more details in Cressie and Johannesson [4].

To estimate the covariance matrix $\mathbf{K}$, an empirical covariance is first computed, which is based on the method-of-moments. The ordinary-least-squares estimator of $\boldsymbol{\beta}$ is:

$$\overline{\boldsymbol{\beta}} \equiv (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Z}, \tag{9}$$

from which *detail residuals*, $\{Z(\mathbf{s}_i) - \mathbf{T}(\mathbf{s}_i)'\overline{\boldsymbol{\beta}}: i = 1, \ldots, n\}$ are defined.

As in classical geostatistics (e.g., Cressie [2], Ch. 2), "binned" data are used for computation of a method-of-moments estimator $\widehat{\boldsymbol{\Sigma}}_M$, which is the empirical covariance matrix of the *binned* detail residuals at bin centers $\{\mathbf{u}_j: j = 1, \ldots, M\}$, where $r \leq M < n$. In a similar way, the binned versions of $\mathbf{S}$ and $\mathbf{V}$ are obtained as $\overline{\mathbf{S}}$ and $\overline{\mathbf{V}}$ (see Cressie and Johannesson [4], Appendix). Finally, one chooses a positive-definite $\mathbf{K}$ and $\sigma^2 \in (0, \infty)$ such that $\overline{\boldsymbol{\Sigma}}_M(\mathbf{K}, \sigma^2) \equiv \overline{\mathbf{S}}\mathbf{K}\overline{\mathbf{S}}' + \sigma^2\overline{\mathbf{V}}$, is as "close" to $\widehat{\boldsymbol{\Sigma}}_M$ as possible. We measure the closeness of two matrices $\mathbf{A}$ and $\mathbf{B}$ via the square of the Frobenius norm:

$$\|\mathbf{A} - \mathbf{B}\|^2 \equiv \text{tr}((\mathbf{A} - \mathbf{B})'(\mathbf{A} - \mathbf{B})) = \sum_{j,k}(A_{jk} - B_{jk})^2,$$

which has also been used by Hastie [8] in deriving pseudosplines, and by Donoho et al. [6] in estimating covariances.

Minimizing the Frobenius norm between $\overline{\boldsymbol{\Sigma}}_M(\mathbf{K}, \sigma^2)$ and $\widehat{\boldsymbol{\Sigma}}_M$ leads to estimators $\hat{\sigma}^2 \in (0, \infty)$ that minimize

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}}_M - \overline{\boldsymbol{\Sigma}}_M(\widehat{\mathbf{K}}, \sigma^2)\|^2 &= \sum_{j,k}\big((\widehat{\boldsymbol{\Sigma}}_M - \mathbf{P}(\widehat{\boldsymbol{\Sigma}}_M))_{jk} \\ &\quad - \sigma^2(\overline{\mathbf{V}} - \mathbf{P}(\overline{\mathbf{V}}))_{jk}\big)^2, \end{aligned}$$

where $\overline{\mathbf{S}} = \mathbf{QR}$ is the Q-R decomposition of $\overline{\mathbf{S}}$; $\mathbf{K}(\sigma^2) \equiv \mathbf{R}^{-1}\mathbf{Q}'(\widehat{\boldsymbol{\Sigma}}_M - \sigma^2\overline{\mathbf{V}})\mathbf{Q}(\mathbf{R}^{-1})'$; and $\mathbf{P}(\mathbf{A}) \equiv \mathbf{QQ}'\mathbf{AQQ}'$ for any $M \times M$ matrix $\mathbf{A}$. Notice that this is just a simple linear regression with slope σ^2 and zero intercept. Hence, the minimization, constrained so that $\sigma^2 > 0$, can be easily carried out. Finally, the $\mathbf{K}$ that minimizes the Frobenius norm is obtained from:

$$\widehat{\mathbf{K}} \equiv \mathbf{R}^{-1}\mathbf{Q}'(\widehat{\boldsymbol{\Sigma}}_M - \widehat{\sigma}^2\overline{\mathbf{V}})\mathbf{Q}(\mathbf{R}^{-1})', \tag{10}$$

where $\widehat{\sigma}^2$ needs to be adjusted so that $\widehat{\mathbf{K}}$ is positive-definite. This completes the model-fitting step. Using these estimated parameters substituted into (7) and (8), FRK prediction can be implemented.
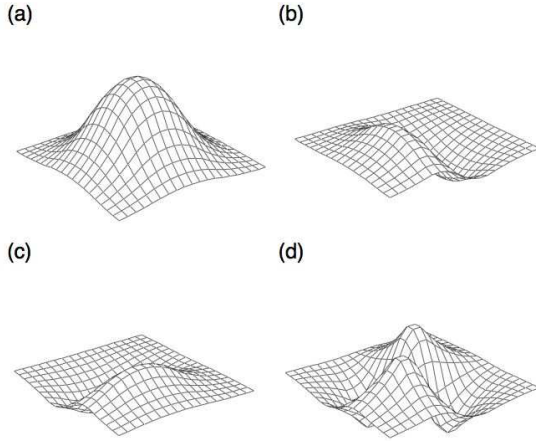
(a)   (b)

(c)   (d)

Fig. 2.   Two-dimensional W-Wavelets basis functions: (a) $\Psi^{(1)}$, (b) $\Psi^{(2)}$, (c) $\Psi^{(3)}$, and (d) $\Psi^{(4)}$.

III. MULTI-RESOLUTION BASIS SELECTION

To analyze the MISR AOD data, we propose to fit the model (4), where $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ are chosen from multi-resolution W-wavelet basis functions. We give a brief discussion of W-wavelets in Section III-A and then present our selection strategy for $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ in Section III-B.

A. W-wavelets

Multi-resolution methods using wavelet basis functions have received quite a lot of attention recently in the Statistics literature. The popularity of wavelets lies in their ability to provide estimates of functions that have discontinuities or varying degrees of smoothness over their domain. An additional advantage of wavelets is their efficient computation using Discrete Wavelet Transform (DWT) algorithms.

Wavelets are functions with varying scales and locations, obtained by dilating and translating a mother wavelet ψ. Associated with each ψ is a scaling function ϕ, also called a father wavelet. Readers are referred to the review paper by Nason and Silverman [16] for more details on wavelet bases. In this paper, we concentrate on a family of wavelets called W-wavelets, in particular the quadratic-spline W-wavelets (Kwong and Tang [14]). Nychka et al. [18] suggested using them as basis functions because they are able to approximate the shape of common covariance models and handle boundaries easily.

To model the spatial data $\mathbf{Z}$, a two-dimensional W-wavelet basis is constructed through translation and scaling of four template functions:

$$\Psi^{(1)}(x,y) = \phi(x)\phi(y), \qquad \Psi^{(2)}(x,y) = \psi(x)\phi(y),$$
$$\Psi^{(3)}(x,y) = \phi(x)\psi(y), \qquad \Psi^{(4)}(x,y) = \psi(x)\psi(y),$$

where $\Psi^{(1)}$ corresponds to the role of the father wavelet and $\Psi^{(2)}$, $\Psi^{(3)}$, and $\Psi^{(4)}$ to that of mother wavelets in the one-dimensional case. These template functions are the result of a tensor product of the one-dimensional father and mother W-wavelet functions (see details in Nychka et al. [18]). Fig. 2

shows the quadratic parents of this family, which are piecewise quadratic splines. Unlike other popular wavelet bases, the W-wavelets are not orthogonal, nor are they compactly supported. In order to model the MISR AOD data, we choose functions $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ in (4) from the two-dimensional quadratic-spline W-wavelets.

As is well known, missing values pose difficulties when computing the DWT, which requires a datum for every location on a regular grid. Matsuo et al. [15] proposed filling in any missing values using a Monte-Carlo Expectation-Maximization approach, but in our case the computational effort would be overwhelming. Our approach is to fill in the missing values on the regular grid using Mean Polishing (e.g., Cressie [2], pp. 184, 185), but note that the "complete" data are only used for the purpose of basis-function selection. Mean Polish is a simple, computationally fast method that puts the missing value at location $\mathbf{s}$ equal to $\bar{Z}_{lon}(\mathbf{s}) + \bar{Z}_{lat}(\mathbf{s}) - \bar{Z}$, where $\bar{Z}_{lon}(\mathbf{s})$, $\bar{Z}_{lat}(\mathbf{s})$, and $\bar{Z}$ are the means of data with the same longitude as $\mathbf{s}$, with the same latitude as $\mathbf{s}$, and all observed data, respectively. The DWT on the "complete" data is then used to accomplish the selection of $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ functions; see the next section.

B. Basis Selection

Recall that the deterministic term $\mathbf{T}(\cdot)'\boldsymbol{\beta}$ and the stochastic term $\mathbf{S}(\cdot)'\boldsymbol{\eta}$ in (4) model large- and small-scale spatial variation, respectively. This suggests generally that we choose $\mathbf{T}(\cdot)$ from wavelets at coarser scales and $\mathbf{S}(\cdot)$ from scales finer than those used for $\mathbf{T}(\cdot)$.

Our strategy for basis-function selection is based on methodology presented in Huang and Cressie [9], in which the authors proposed a method to decompose deterministic and stochastic terms in wavelet signal processes. Among the wavelet coefficients at certain scales, those with large absolute wavelets coefficients are selected for use as either deterministic or stochastic components of (4), as follows.

In the SME model (4), we choose $\mathbf{T}(\cdot)$ as all W-wavelets involved in scales $1, \ldots, J_0$, plus those with large absolute coefficients at scale $J_0 + 1$; in practice, J_0 is usually small and its choice depends on the desired level of large-scale spatial variability. To choose $\mathbf{S}(\cdot)$, we take those basis functions left unselected for $\mathbf{T}(\cdot)$ at scale $J_0 + 1$, plus wavelets at scale $J_0 + 2$ that have large absolute coefficients. The method used to determine which are the large absolute wavelet coefficients is defined below.

To identify the large absolute wavelet coefficients automatically at a given scale J, we adopt the method discussed in Huang and Cressie [9]. Let m be the number of wavelet functions at scale J, and $w_1, \cdots, w_m$ be the coefficients calculated from the DWT. Let $|w|^{(p)}$ be the p-quantile of $|w_1|, \cdots, |w_m|$, and $q^{(p)}$ be the p-quantile of the standard Gaussian distribution. We compute

$$\hat{\tau} = |w|^{(1-2\alpha)}/q^{(1-\alpha)},$$

where $0 < \alpha < 0.5$ is chosen to ensure that the Q-Q line with slope $\hat{\tau}$ fits the wavelet coefficients that are small or moderate
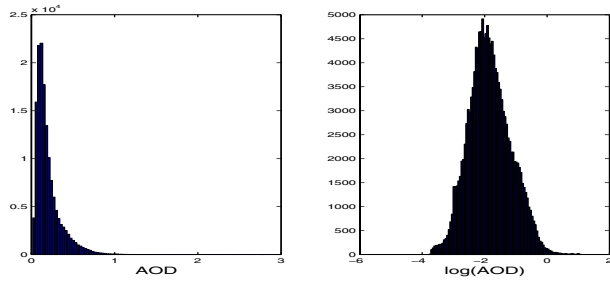
Fig. 3.   Left panel: Histogram of AOD; right panel: Histogram of log(AOD).

in absolute value. A large absolute wavelet coefficient at scale J is defined as a w_i that satisfies:

$$|w_i| > \hat{\tau} \times q_{\max},$$

where $q_{\max} \equiv \max\{q^{(1-p)} : |w|^{(1-2p)} < \hat{\tau}q^{(1-p)}, \; p = 1/2m, 2/2m, \cdots, 1/2\}$; see Huang and Cressie [9]. After identifying these coefficients, the corresponding wavelet functions are selected to be in either $\mathbf{T}(\cdot)$ or $\mathbf{S}(\cdot)$, according to the description given above.

IV. MISR Aerosol Optical Depth

We now carry out a spatial statistical analysis on the MISR global level-3 AOD data. The dataset used in this study are averaged level-2 AOD data, averaged within level-3 pixels, for the 16-day period of April 1, 2001 to April 16, 2001. Recall that this covers a MISR full repeat-cycle. The middle panel in Fig. 1 shows the global AOD data with gray areas corresponding to level-3 pixels with no data in them. In addition to the averaged AODs, the MISR level-3 product also provides the number of level-2 observations, $N(\mathbf{s})$, used in that average, for each pixel centered at $\mathbf{s}$ that is not gray.

Initial investigation shows a long-tailed distribution of AOD (the histogram is shown in the left panel of Fig. 3), so we use log(AOD) as the primary variable in our analysis. The histogram of log(AOD) is shown in the right panel of Fig. 3, clearly exhibiting a distributional shape that is close to symmetric. Therefore, we fit the SME model (4) to $\mathbf{Z}$, the log(AOD) data. Untransforming from log(AOD) to AOD is straightforward (e.g., Cressie [3]).

In Section 4.1, we apply FRK to data in a subregion deliberately chosen with missing data, and we compare its accuracy and computational cost to ad hoc, non-statistical spatial methods. In Section 4.2, we present FRK for (almost) global data. All computations were carried out in Matlab on a linux machine with a Pentium 4 dual core 3.0 Ghz processor and 2GB memory.

A. Spatial Analysis on a Subregion of the Globe

We apply the proposed FRK methodology to log(AOD) in the rectangular region between longitudes $-140°$ and $12°$ and between latitudes $-20°$ and $45°$, which covers a geographic region that includes North and South America, Africa, and
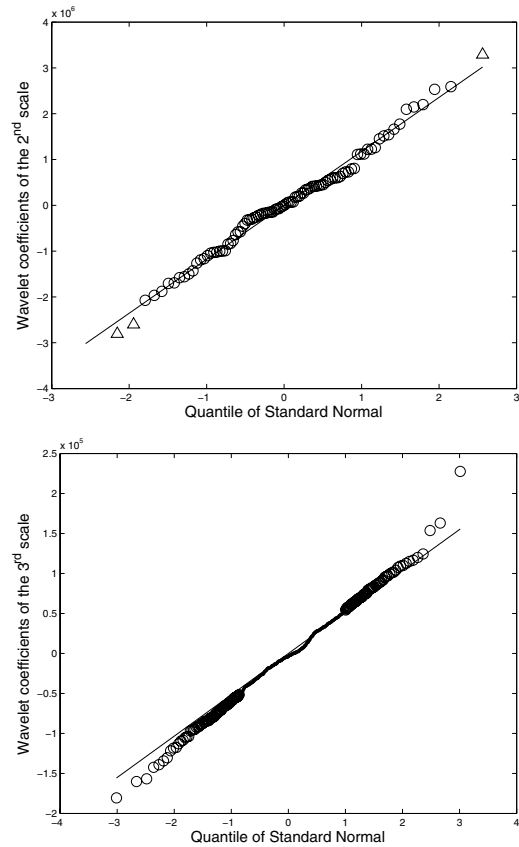

Fig. 4.   Wavelet coefficients for log(AOD) data in the subregion. Upper panel: $J = 2$; Lower panel: $J = 3$. The fitted Q-Q line is superimposed, and coefficients corresponding to wavelet functions in $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ are plotted as '$\triangle$' and 'O', respectively.

parts of the Atlantic and Pacific Oceans. This study region is delineated by the red box in the lower panel of Fig. 1.

The study region is a test-bed for validation experiments. It was selected for three reasons. First, the observed log(AOD) exhibits strong spatial dependence. Second, there are $128 \times 256 = 32,768$ pixels ($0.5 \; degree$ by $0.5 \; degree$) in the region, and this fairly large amount of data presents a good case with which to compare the computational efficiencies of different methods. Third, 6,971 (21% of total) pixels are unobserved in this region, especially over areas of South America and the Pacific Ocean, so it is a good case for testing the accuracy of different methods for filling in missing values.

1) Basis Selection: Here we apply the basis-selection method proposed in Section 3 to the data in the subregion. We first compute the wavelet coefficients using the DWT on log(AOD), with the missing values initially filled in by Mean Polishing. Among the set of all coefficients, there are $4 \times 2 \times 4 = 32$ at the first scale. The second scale involves three sets of 4×8 basis functions (96 total), being scaled translates of $\Psi^{(2)}$, $\Psi^{(3)}$, and $\Psi^{(4)}$. Similarly, the third scale has $3 \times 8 \times 16 = 384$ basis functions.

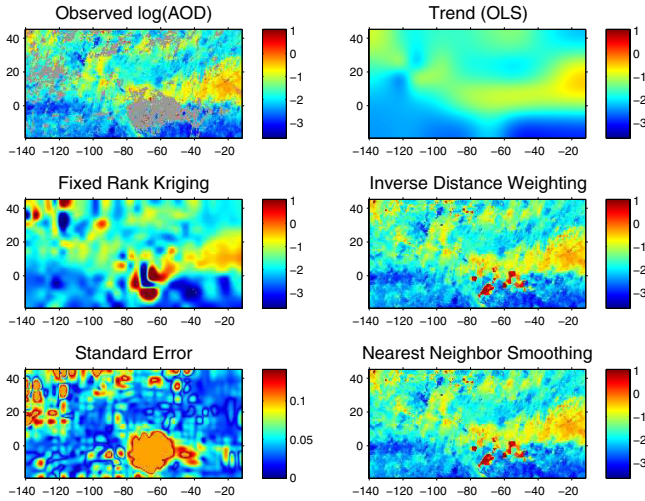We then follow the basis-selection procedure given in Sec-

Fig. 5. Upper-left panel: Observed log(AOD); middle-left panel: FRK predictions; lower-left panel: FRK standard errors; upper-right panel: OLS fitted trend $\mathbf{T}\overline{\beta}$; middle-right panel: Inverse Distance Weighting; lower-right panel: Nearest Neighbor Smoothing.

tion 3 with selection parameters $J_0 = 1$, $\alpha = 0.2$ at the second scale, and $\alpha = 0.35$ at the third scale. This procedure results in 35 basis functions for $\mathbf{T}(\cdot)$, 32 from the first scale and 3 from the second scale. Turning to $\mathbf{S}(\cdot)$, the procedure selects 229 basis functions with $96 - 3 = 93$ from the second scale and 136 from the third scale. The Q-Q plots of w's at the second scale and at the third scale are shown in Fig. 4, where the coefficients corresponding to wavelets in $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ are plotted as symbols '△' and 'O', respectively. From the figure, the wavelet coefficients within the two scales fit well to a Gaussian distribution, except for those with large absolute values.

2) SME Model Fitting: We now fit the SME model (4) to $\mathbf{Z}$ (i.e., observed log(AOD)) in the study region (also shown in the upper-left panel of Fig. 5). The Ordinary Least Squares (OLS) fit using basis functions $\mathbf{T}(\cdot)$ on the observed data leads to OLS coefficients $\overline{\beta}$, the fitted values $\hat{Z}^{OLS}(\mathbf{s}_i) \equiv \mathbf{T}(\mathbf{s}_i)'\overline{\beta}$, and detail residuals, $\{D(\mathbf{s}_i) \equiv Z(\mathbf{s}_i) - \hat{Z}^{OLS}(\mathbf{s}_i) : i = 1, \ldots, n\}$. The OLS fit is shown in the upper-right panel of Fig. 5. Now, to estimate $\mathbf{K}$ and σ^2 in the SME model, we bin the detail residuals $\{D(\mathbf{s}_i)\}$ into $M = 1253$ bins, which result from using a 5×5 window. Since each pixel is 0.5 *degrees* by 0.5 *degrees*, the bins are 2.5 *degrees* by 2.5 *degrees*.

Recall that the diagonal matrix $\overline{\mathbf{V}}$ is the binned version of $\mathbf{V}$; we assume $\overline{\mathbf{V}} = diag(1/N(\mathbf{s}))$. The binned empirical covariance matrix $\widehat{\Sigma}_M$ is calculated, and hence we obtain estimates $\hat{\sigma}^2$ and $\hat{\mathbf{K}}$ using the fitting procedure given in Section 2.2. Then the optimal spatial predictor, $\hat{Y}(\mathbf{s}_0)$, and the prediction standard error, $\sigma_k(\mathbf{s}_0)$, are calculated using the FRK equations (7) and (8), where $\mathbf{s}_0$ is allowed to range over all level-3 pixels in the study region. The middle-left and lower-left panels of Fig. 5 show the FRK predictions and the FRK standard errors, respectively.

	FRK	OLS	IDW	NNS
mean(MSPE)	0.346	0.4304	0.2186	0.2228
sd(MSPE)	0.0102	0.0086	0.0074	0.0074

The maps show that FRK captures the spatial variation well, and its standard error behaves as expected, showing high values at locations with many unobserved values. Especially, the hot spot between longitudes $-80°$ and $-40°$ and between latitudes $-10°$ and $10°$ indicates FRK prediction at those locations has uncertainty. We also observe that the uncertainty map exhibits regular "grid" pattern due to local supports of the wavelets basis functions. The computation of FRK in this subregion of 32,768 pixels took 33.7 seconds to fit the model and 84.1 seconds to compute all $32,768$ predicted values and their prediction standard errors.

3) Comparison with OLS, IDW and NNS: Besides visual inspection of the FRK results, we can also compare, via cross-validation experiments, the accuracy of FRK with other computationally fast nonstatistical spatial-prediction methods, such as the Inverse Distance Weighting (IDW) method and the Nearest Neighbor Smoothing (NNS) method. These two methods are commonly used to address the missing-value problem in geosciences; the results of IDW and NNS are shown in the middle-right and lower-right panels of Fig. 5, respectively. We also include OLS in our comparison, which is another possible method for predicting missing values for large datasets.

Case 1: Cross-Validation by Random Sampling. For the first comparison, we randomly selected 10% of the 25,897 observed values and treated them as testing data, denoted as $\{Z(\mathbf{s}_i^{test})\}$. Then the remaining 90% of the data are used to predict them, based on the prediction methods FRK, OLS, IDW, and NNS. In this experiment, the IDW and NNS methods are computed for the 10 nearest observed values in a moving, variable-size window, and OLS uses the functions $\mathbf{T}(\cdot)$ that were selected.

The prediction accuracy of each method is evaluated by the Mean Squared Prediction Error ($MSPE$) over the testing data; that is,

$$MSPE = \frac{1}{n^{test}} \sum_i (Z(\mathbf{s}_i^{test}) - \hat{Y}(\mathbf{s}_i^{test}))^2,$$

where n^{test} as the size of the testing dataset. This cross-validation procedure is repeated 50 times (50 different 10% samples), and the mean and standard deviation of the 50 $MSPE$'s are reported in Table 1.

Overall, FRK outperforms OLS, but its prediction is not as close to $\{Z(\mathbf{s}_i^{test})\}$ as IDW and NNS. FRK only attempts to reproduce the smooth underlying process $\mathbf{Y}$, not the noisy measured values $\mathbf{Z}$. When the prediction location is surrounded by enough observed neighborhoods, we expect that IDW and NNS work well. However, these two neighborhood methods exhibit strong nonsmooth artifacts in areas where

TABLE II

MSPE OF FRK, IDW AND NNS FOR CASE 2.

	FRK	IDW	NNS
MSPE	0.4046	1.0717	1.0716

a majority of data are missing (e.g., a large area in South America, between longitudes $-80°$ and $-40°$ and between latitudes $-10°$ and $10°$). This effect will be investigated in the second cross-validation experiment. More importantly, IDW and NNS methods do not provide uncertainty measures, which are essential for comparison of AOD over different regions or at different time points.

Case 2: Contiguous Region Set Aside. In this comparison, we concentrate on evaluating the performance of FRK, IDW, and NNS in regions where a majority of the data are missing. To mimic this in our experiment, we set aside a vertical strip of observed values whose longitudes are between $-110°$ and $-90°$, and treat them as testing data. The remaining data are used to predict the testing data, based on the prediction methods FRK, IDW, and NNS. As reported in Table 2, MSPE of FRK is 0.4046, which is much lower than 1.0717 of IDW and 1.0716 of NNS, respectively. Therefore, FRK clearly outperforms IDW and NNS in this case. From these studies, we conclude that FRK is more reliable for analyzing massive MISR global AOD with its type of missing values.

B. Spatial Analysis on a "Global" Dataset

We present the results on an "almost global" dataset to show the potential of our method to process massive MISR datasets. We implement FRK using equations (7) and (8) on $\log(\text{AOD})$ data in the rectangular region between longitudes $-130°$ and $126°$ and between latitudes $-66°$ and $62°$. This almost global region contains $512 \times 256 = 131,072$ pixels of size $0.5 \ degree$ by $0.5 \ degree$. The region is delineated by the red box in the upper panel of Fig. 6.

In this experiment, the model-selection step is carried out with $J_0 = 2$, $\alpha = 0.1$ at the third scale, and $\alpha = 0.05$ at the fourth scale. The resulting functions $\mathbf{T}(\cdot)$ and $\mathbf{S}(\cdot)$ contain 163 and 429 wavelet functions, respectively. In the spatial-model-fitting step, there are $M = 2614$ bin centers resulting from bins that are $2.5 \ degrees$ by $2.5 \ degrees$. The middle and lower panels of Fig. 6 show the FRK predictions and the FRK standard errors given by (7) and (8), respectively. As for this subregion, we see that FRK prediction captures the spatial variation and the FRK standard errors behave appropriately (i.e., they are larger where there are fewer data). The uncertainty hot spots observed in some regions correspond to the areas where majority observations are missing. We also point out that the regular "grid" pattern observed in the uncertainty maps is due to local support of the wavelets basis functions used in fitting the covariance matrix $\mathbf{SKS}'$.

For this almost-global dataset with 131,072 pixels, the computation times for fitting and prediction are 156.7 seconds and 1477.0 seconds, respectively. It is worth pointing out that the map we produce here is at a $0.5 \ degree$ by $0.5 \ degree$
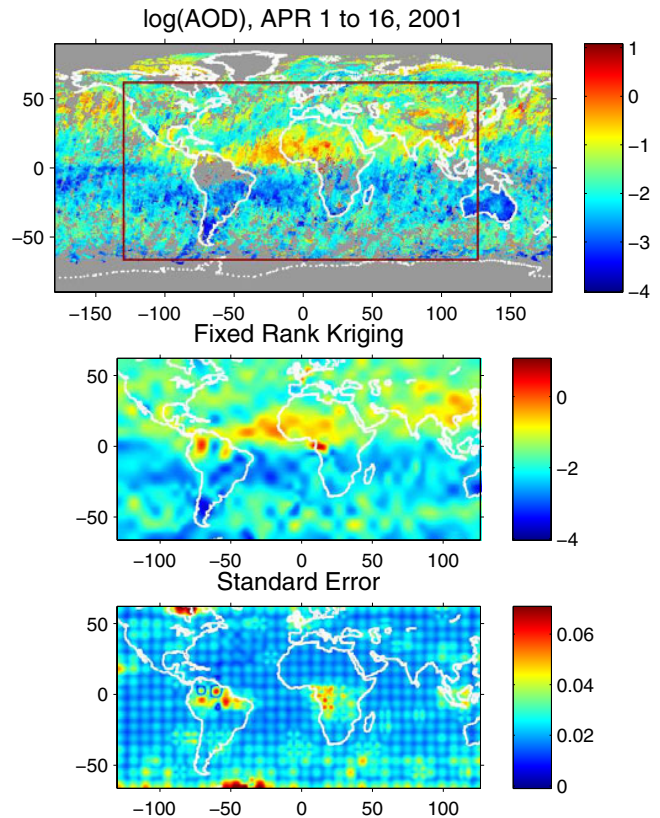

Fig. 6.   Upper panel: Observed $\log(\text{AOD})$; middle panel: FRK predictions; lower panel: FRK standard errors.

resolution, so the number of pixels we deal with here *exceeds* that of a global map at $1 \ degree$ by $1 \ degree$ resolution (where there are $360 \times 180 = 64,800$ pixels). It is clear that the accuracy (Section 4.1) and computation times (Section 4.2) resulting from our experiments show that FRK is capable of efficiently (statistically and computationally) modeling and analyzing massive global datasets.

V. DISCUSSION

In this paper, we propose a Spatial Random Effects (SRE) model using multi-resolution basis functions as part of a spatial model we fit to massive global satellite datasets. Through a methodology called Fixed Rank Kriging (Cressie and Johannesson [4]), we are able to use spatial statistics to predict the missing values, denoise the observed values, and provide a measure of uncertainty. Importantly, FRK is linearly scalable to the number of data points, which makes it suitable for massive global satellite data processing. We applied the proposed method to the MISR level-3 AOD data, and overall our results were superior to nonstatistical methods. FRK has the crucial advantage that it comes with a measure of uncertainty (the FRK standard error).

When we wish to predict processes at one resolution, but data are observed at another, the SRE model can be seen to

be particularly advantageous. Recall that the SRE model has covariance function, $\text{cov}(\nu(\mathbf{u}), \nu(\mathbf{v})) = \mathbf{S}(\mathbf{u})'\mathbf{KS}(\mathbf{v})$. Then it is easy to see that $\text{cov}(\nu(B_1), \nu(B_2)) = \mathbf{S}(B_1)'\mathbf{KS}(B_2)$, where $\nu(B) \equiv (1/|B|) \int_B \nu(\mathbf{u})d\mathbf{u}$ and $\mathbf{S}(B) \equiv (1/|B|) \int_B S(\mathbf{u})d\mathbf{u}$. That is, the *form* of the SRE model is invariant under aggregation.

The spatial model used in this paper may be easily generalized to a space-time model, $Z(\mathbf{s}, t) = \mathbf{T}(\mathbf{s})'\boldsymbol{\beta}(t) + \mathbf{S}(\mathbf{s})'\boldsymbol{\eta}(t) + \varepsilon(\mathbf{s}, t)$; $\mathbf{s} \in D$, $t \in T$. This model is not only flexible enough to capture the spatial and temporal variation of the hidden process, but it also shares the computational advantages of the spatial model. In future work, we shall also consider basis functions from spherical wavelets, in order to work directly on the globe.

REFERENCES

[1] R. P. Barry and R. K. Pace. Kriging with large data sets using sparse matrix techniques. *Communications in Statistics: Simulation and Computation*, 26:619–629, 1997.

[2] N. Cressie. *Statistics for Spatial Data, rev. edn.* John Wiley & Sons, New York, 1993.

[3] N. Cressie. Block kriging for lognormal spatial processes. *Mathematical Geology*, 38:forthcoming, 2006.

[4] N. Cressie and G. Johannesson. Fixed rank kriging for large spatial datasets. Technical Report No. 780, Department of Statistics, The Ohio State University, Columbus, OH, 2006.

[5] D. J. Diner, G. P. Asner, R. Davies, Y. Knyazikhin, J. Muller, A. W. Nolin, B. Pinty, C. B. Schaaf, and J. Stroeve. New directions in earth observing scientific applications of mulitangle remote sensing. *Bulletin of American Meteorological Society*, 80:2209–2228, 1999.

[6] D. L. Donoho, S. Mallet, and R. von Sachs. Estimating covariances of locally stationary processes: Rates of convergence of best basis methods. Technical Report No. 517, Department of Statistics, Stanford University, Stanford, CA, 1998.

[7] R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15:502–523, 2006.

[8] T. Hastie. Pseudosplines. *Journal of the Royal Statistical Society B*, 58:379–396, 1996.

[9] H.-C. Huang and N. Cressie. Deterministic/stochastic wavelet decomposition for recovery of signal from noisy data. *Technometrics*, 42:262–276, 2000.

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

[11] G. Johannesson, N. Cressie, and H.-C. Huang. Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics*, 2007, forthcoming.

[12] E. E. Kammann and M. P. Wand. Geoadditive models. *Applied Statistics*, 52:1–18, 2003.

[13] Y. J. Kaufman, B.N. Holben, D. Tanre, I. Slutsker, and A. Smirnov. Will aerosol measurements from Terra and Aqua polar orbiting satellites represent the daily aerosol abundance and properties? *Geophysical Research Letters*, 27:3861–3864, 2000.

[14] M. K. Kwong and P. T. P. Tang. W-matrices, nonorthogonal multiresolution analysis, and finite signals of arbitrary length. Technical Report MCS-P449-0794, Argonne National Laboratory, 1994.

[15] T. Matsuo, D. Nychka, and D. Paul. Multi-resolution (wavelet) based nonstationary covariance modeling for incomplete data: Smoothed Monte-Carlo approach. Technical Report The National Center for Atmospheric Research, 2006.

[16] G. P. Nason and B. W. Silverman. The discrete wavelet transform. *Journal of Computational and Graphical Statistics*, 3:163–191, 1994.

[17] D. Nychka. Spatial-process estimates as smoothers. In M. G. Schimek, editor, *Smoothing and Regression: Approaches, Computation, and Application*, pages 393–424. John Wiley & Sons, New York, 2000.

[18] D. Nychka, C. Wikle, and J. A. Royle. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modeling*, 2:315–331, 2002.

[19] S. Tzeng, H.-C. Huang, and N. Cressie. A fast, optimlal spatial-prediction method for massive datasets. *Journal of the American Statistican*, 100:1343–1357, 2005.