Supplementary materials for this article are available online. Please click the JASA link at http://pubs.amstat.org.

An Ensemble Kalman Filter and Smoother for Satellite Data Assimilation

Jonathan R. STROUD, Michael L. STEIN, Barry M. LESHT, David J. SCHWAB, and Dmitry BELETSKY

This paper proposes a methodology for combining satellite images with advection-diffusion models for interpolation and prediction of environmental processes. We propose a dynamic state-space model and an ensemble Kalman filter and smoothing algorithm for on-line and retrospective state estimation. Our approach addresses the high dimensionality, measurement bias, and nonlinearities inherent in satellite data. We apply the method to a sequence of SeaWiFS satellite images in Lake Michigan from March 1998, when a large sediment plume was observed in the images following a major storm event. Using our approach, we combine the images with a sediment transport model to produce maps of sediment concentrations and uncertainties over space and time. We show that our approach improves out-of-sample RMSE by 20%–30% relative to standard approaches. This article has supplementary material online.

KEY WORDS: Circulant embedding; Covariance tapering; Gaussian random field; Nonlinear state-space model; Spatial statistics; Spatio-temporal model; Variogram.

1. INTRODUCTION

Satellites provide a valuable tool for environmental monitoring. They produce high-resolution images of geophysical variables such as stratospheric ozone, surface winds, and ocean chlorophyll. These images are used for many purposes, including estimation of temporal trends, seasonal cycles, spatio-temporal variability, and inputs to computer models. The amount of remote-sensing data has greatly increased in recent years. According to the Joint Center on Satellite Data Assimilation (*http://www.jcsda.noaa.gov*), there has been a "hundred-thousand fold increase in satellite data this decade from nearly fifty new instruments." However, despite their availability and vast potential, little statistical work has been done in this area. The goal of this paper is to provide a new statistical approach for analyzing spatio-temporal satellite data.

To motivate our approach, Figures 1 and 2 show a sequence of satellite images of Lake Michigan during the development of a large sediment plume. The plots highlight a number of features. First, the data are high dimensional with each image consisting of over 14,000 pixels. Second, the images are unequally spaced and have large amounts of missing data due to cloud cover. Third, there is a dominant transport effect, with the sediment plume moving westward and expanding over time. Finally, the background light intensity varies from image to image, indicating spatially correlated measurement errors. These features are common in many satellite datasets. The goal of our analysis is to use the images to estimate the sediment concentration field over space and time and provide estimates of uncertainty, while accounting for the features above.

The current literature on space–time analysis of satellite data is quite sparse. Niu and Tiao (1995) and Stein (2007) analyzed total column ozone at a single latitude using space–time ARMA models and variogram models, respectively. Johannesson, Cressie, and Huang (2007) analyzed total column ozone on a global scale using dynamic multiresolution models, processing nearly one million observations. Wikle et al. (2001) proposed a Bayesian hierarchical model to combine satellite-derived surface winds in the South Pacific with analyzed numerical model output. The last two papers relied on dimension-reduction approaches to deal with the large satellite images. The former assumed conditional independence across spatial resolutions while the latter relied on wavelet and spectral methods to incorporate dynamics in the underlying wind fields.

Other authors have proposed dynamic models that explicitly incorporate transport processes. Wikle, Berliner, and Cressie (1998) defined the state vector on a grid and assumed a vector autoregressive evolution equation with nearest-neighbor structure. Wikle and Cressie (1999) specified the evolution equation through a space–time interaction function (kernel). Brown et al. (2000) used a similar idea motivated by stochastic differential equations. Wikle (2002) and Higdon (2002) proposed space–time interaction kernels that vary smoothly over space, while Huang and Hsu (2004) defined the kernels based on available wind speed data. Wikle (2003) and Xu and Wikle (2007) considered diffusion and advection-diffusion models, whereas Gelpke and Künsch (2001) assumed an advection model and estimated velocity fields from a sequence of satellite images. All but the last paper assume the transport coefficients are constant over either space or time. However, this assumption may be unrealistic for many satellite data applications. Our approach removes this constraint and allows the coefficients to vary over both space and time.

Jonathan R. Stroud is Associate Professor, Department of Statistics, George Washington University, Washington, DC 20052 (E-mail: *stroud@gwu.edu*). Michael L. Stein is Ralph and Mary Otis Isham Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. Barry M. Lesht is Adjunct Professor, Department of Earth and Environmental Sciences, University of Illinois at Chicago, Chicago, IL 60607. David J. Schwab is Physical Oceanographer, Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI 48108. Dmitry Beletsky is Associate Research Scientist, Cooperative Institute for Limnology and Ecosystems Research, School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109. Funding was provided by the U.S. Environmental Protection Agency (EPA) through STAR Cooperative Agreement R-82940201 to the University of Chicago. However, this research has not been subjected to the EPA's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The authors are grateful to NASA for providing the SeaWiFS data and the SeaDAS processing software, and to NOAA for providing the in situ data through the EEGLE Project and the Coastal Ocean Program. The authors thank the editor, associate editor, and two referees for their suggestions which greatly improved the manuscript, and Chris Wikle, Bruno Sansó, and Lurdes Inoue for helpful comments.

Journal of the American Statistical Association
September 2010, Vol. 105, No. 491, Applications and Case Studies
DOI: 10.1198/jasa.2010.ap07636

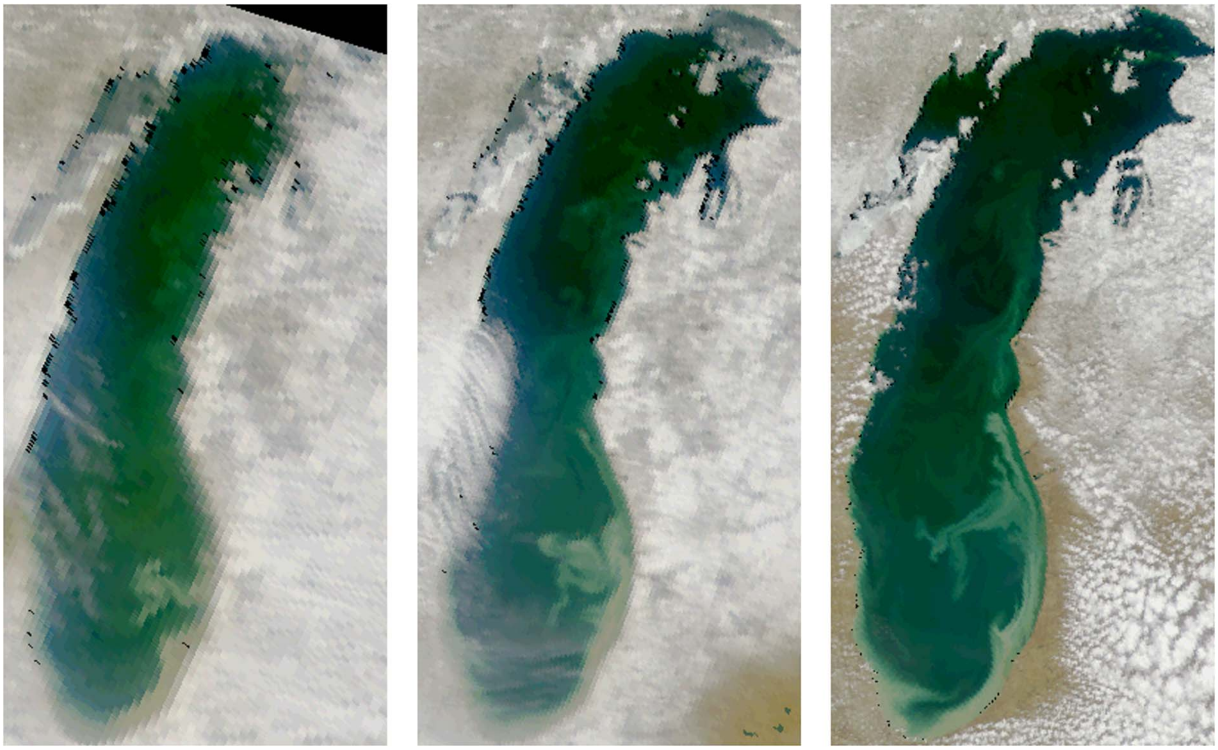

Figure 1. SeaWiFS satellite "True-Color" images of Lake Michigan on March 12, 16, and 24, 1998. The light brown areas in the lake indicate suspended sediment.

In this paper, we propose a dynamic state-space model for high-dimensional satellite data. The model explicitly incorporates motion in the geophysical variable by defining the state evolution through an advection-diffusion model. The state vector is defined on a spatial grid and the partial differential equations are solved using finite-difference methods. The discrete-

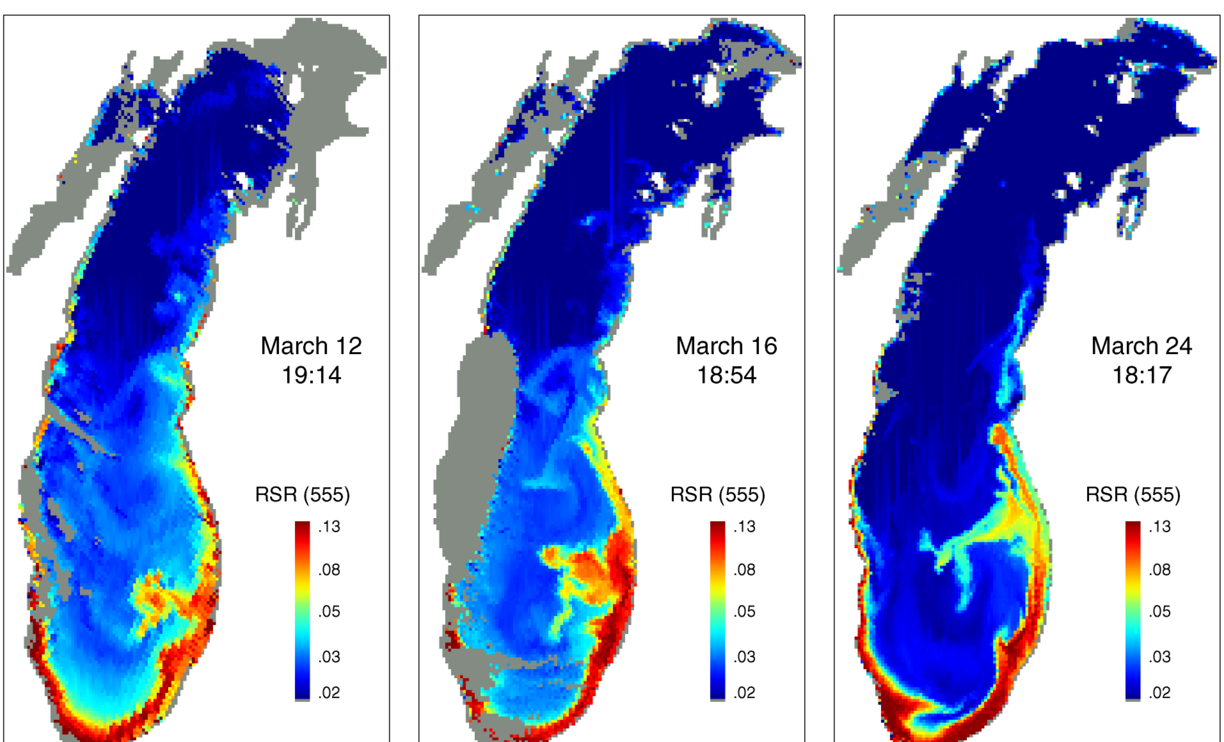

Figure 2. SeaWiFS remote sensing reflectance at 555 nanometers, on March 12, 16, and 24, 1998. Gray pixels indicate cloud cover, as identified by a screening algorithm. All times are GMT.

time evolution equation can then be written as a vector autoregression with a sparse time-varying transition matrix. While state-space models driven by advection-diffusion models are not new (Wikle 2003; Xu and Wikle 2007), to our knowledge this is the first application to satellite data in the statistics literature. Our approach also accommodates massive datasets, allows for missing values, and incorporates correlated errors and nonlinear measurement models.

To deal with the high-dimensional satellite data, we rely on a sequential Monte Carlo method known as the ensemble Kalman filter (Evensen 1994). This approach is essential because standard state-space methods such as Kalman filters and particle filters do not scale to high dimensions. Iterative approaches such as expectation–maximization (EM) and Markov chain Monte Carlo (MCMC) are computationally infeasible in this context. Furthermore, standard dimension-reduction techniques based on spectral and wavelet methods often remove important small-scale features in the images. Our approach retains this small-scale information while allowing fast computation through the idea of *covariance tapering*. We also propose two novel extensions of the ensemble Kalman filter. In particular, we develop a variational updating scheme for high-dimensional data with correlated errors and we derive a new ensemble Kalman smoother for retrospective state estimation.

The rest of the paper is outlined as follows. Section 2 presents the general framework for combining advection-diffusion models with satellite data. Section 3 presents the state-space models and ensemble Kalman filter and smoothing algorithms for online and retrospective state estimation. In Section 4 we demonstrate the methodology through a case study of sediment transport in Lake Michigan. Discussion and extensions of the work are given in Section 5.

2. PHYSICAL–STATISTICAL MODEL

2.1 Measurement Model

We consider the following setup. Let $\{y(\mathbf{s},t), \mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d, t \in \mathcal{T} \subset \mathbb{R}\}$ denote the observed satellite data at location $\mathbf{s}$ and time t, and let $\{c(\mathbf{s},t), \mathbf{s} \in \mathcal{S}, t \in \mathcal{T}\}$ denote the geophysical variable of interest. For example, in our application, $y(\mathbf{s},t)$ represents the water-leaving reflectance from the SeaWiFS satellite, and $c(\mathbf{s},t)$ represents the suspended sediment concentration in Lake Michigan. We assume the following measurement model:

$$y(\mathbf{s},t) = h(c(\mathbf{s},t)) + b(\mathbf{s},t) + \nu(\mathbf{s},t), \tag{1}$$

where $h(\cdot)$ is a possibly nonlinear measurement function mapping the geophysical variable onto the observation scale; $b(\mathbf{s},t)$ is the observation bias; and $\nu(\mathbf{s},t)$ is the observation error. In our application, we specify a parametric nonlinear measurement function, $h(c;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a set of unknown measurement parameters.

The observational bias, $b(\mathbf{s},t)$, is modeled as a linear function of covariates. Let $\mathbf{z}(\mathbf{s},t) = (z_1(\mathbf{s},t),\ldots,z_p(\mathbf{s},t))'$ denote the vector of covariates at location $\mathbf{s}$ and time t and $\boldsymbol{\beta}_t = (\beta_{t1},\ldots,\beta_{tp})'$ the vector of unknown bias coefficients at time t. We assume the following model:

$$b(\mathbf{s},t) = \mathbf{z}(\mathbf{s},t)'\boldsymbol{\beta}_t. \tag{2}$$

The covariates might include variables such as satellite viewing angle, total brightness, or simple functions of the spatial coordinates. Space–time correlation in $b(\mathbf{s},t)$ can be incorporated through the choice of covariates or through a smoothness prior on the coefficients (see Section 2.6). In our application, we model the bias for each image time as a spatial constant, and assume the coefficients are independent across time.

The observation errors, $\nu(\mathbf{s},t)$, represent the difference between the satellite data and the predicted values when the concentrations and the observation bias are known. We assume the errors are independent in time and correlated in space, and model $\nu(\cdot,t)$ as a stationary Gaussian random field with mean zero and covariance function

$$\operatorname{cov}(\nu(\mathbf{s},t),\nu(\mathbf{s}',t)) = K(\|\mathbf{s}-\mathbf{s}'\|;\boldsymbol{\theta}_\nu). \tag{3}$$

Here $K(\cdot)$ is an isotropic covariance function, $\|\cdot\|$ is Euclidean distance, and $\boldsymbol{\theta}_\nu$ is a set of parameters. We consider the general class of Matérn covariance models (Stein 1999), which includes the exponential model $K(d) = \sigma_\nu^2\exp(-d/\rho_\nu)$ as a special case. We also consider a *tapered* exponential model, obtained by multiplying the exponential covariance by a compactly supported correlation function. This choice leads to sparse covariance matrices and allows for fast simulation of random fields in the filtering algorithms described in Section 3.

2.2 Advection-Diffusion Model

In what follows, we let $c(\mathbf{s},t)$ represent the concentration of the geophysical variable of interest. We assume the spatio-temporal dynamics of $c(\mathbf{s},t)$ are dominated by transport processes such as winds or water currents, and model its evolution through a linear advection-diffusion model:

$$\frac{\partial c}{\partial t} = -\nabla\cdot(\mathbf{u}c) + D\nabla^2 c + S. \tag{4}$$

Here $\mathbf{u} = \mathbf{u}(\mathbf{s},t)$ is the velocity vector, which varies over space and time, ∇ is the vector gradient operator, D is the diffusion coefficient, and $S = S(\mathbf{s},t)$ is a source-sink term, which may depend on a set of forcing variables $\boldsymbol{\tau}(\mathbf{s},t)$ and a vector of physical parameters $\boldsymbol{\psi}$. The model is completed with a set of initial and boundary conditions.

In general, the partial differential equations (4) cannot be solved analytically, so we rely on a numerical finite-difference scheme to solve the system, as described below. Throughout the paper, we will assume that the velocities, the diffusion coefficient, and the forcing variables are all known. The objective here is to estimate the unknown concentration field, conditional on these variables and the observed satellite images.

2.3 Discretization Scheme

To describe our numerical approach, we consider a two-dimensional spatial domain and denote the concentration field by $c = c(x,y,t)$, where $(x,y) \in \mathcal{S} \subset \mathbb{R}^2$ denotes the spatial location. The advection-diffusion model can then be written as

$$\frac{\partial c}{\partial t} = -\frac{\partial(uc)}{\partial x} - \frac{\partial(vc)}{\partial y} + D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) + S, \tag{5}$$

where $u = u(x,y,t)$ and $v = v(x,y,t)$ are the velocities in the x and y directions, D is the diffusion coefficient, and $S = S(x,y,t)$ is the source-sink term, which may depend on a set of forcing variables $\boldsymbol{\tau}(x,y,t)$ and a vector of physical parameters $\boldsymbol{\psi}$.

Many finite-difference schemes can be used to solve the system (5). Here, we consider the forward-time, central-space (FTCS) method, which uses the following approximations to the partial derivatives: $\partial c/\partial t \approx (c(x,y,t) - c(x,y,t-\Delta t))/\Delta t$, $\partial c/\partial x \approx (c(x+\Delta x,y,t) - c(x-\Delta x,y,t))/2\Delta x$, and $\partial^2 c/\partial x^2 \approx (c(x+\Delta x,y,t) - 2c(x,y,t) + c(x-\Delta x,y,t))/(\Delta x)^2$, where Δx, Δy, and Δt denote the spatial and temporal grid spacings. A similar method was used by Xu and Wikle (2007). Using this discretization scheme, the equation for the concentrations at the interior grid points can be written as

$$\begin{aligned} c(x,y,t+\Delta t) = {} & \phi_1(x,y,t)c(x,y,t) \\ & + \phi_2(x,y,t)c(x+\Delta x,y,t) \\ & + \phi_3(x,y,t)c(x-\Delta x,y,t) \\ & + \phi_4(x,y,t)c(x,y+\Delta y,t) \\ & + \phi_5(x,y,t)c(x,y-\Delta y,t) \\ & + \alpha(x,y,t), \end{aligned} \tag{6}$$

where the coefficients are given by

$$\begin{aligned} \phi_1(x,y,t) &= 1 - \left(\frac{2D}{(\Delta x)^2} + \frac{2D}{(\Delta y)^2}\right)\Delta t, \\ \phi_2(x,y,t) &= \left(\frac{D}{(\Delta x)^2} - \frac{u(x,y,t)}{2\Delta x}\right)\Delta t, \\ \phi_3(x,y,t) &= \left(\frac{D}{(\Delta x)^2} + \frac{u(x,y,t)}{2\Delta x}\right)\Delta t, \\ \phi_4(x,y,t) &= \left(\frac{D}{(\Delta y)^2} - \frac{v(x,y,t)}{2\Delta y}\right)\Delta t, \\ \phi_5(x,y,t) &= \left(\frac{D}{(\Delta y)^2} + \frac{v(x,y,t)}{2\Delta y}\right)\Delta t, \\ \alpha(x,y,t) &= S(x,y,t)\Delta t. \end{aligned}$$

The equations for the exterior grid points are determined by the specified boundary conditions. In our application, the lake has a closed coastline, which results in a slight modification of (6) along the boundary, with some of the coefficients being set to zero. Thus, the discretized advection-diffusion model can be viewed as a deterministic linear autoregression with a nearest-neighbor structure, where the autoregressive coefficients $\phi_1, \ldots, \phi_5$ vary across space and time. Note that a model with no advection, diffusion, sources, or sinks implies the static evolution equation: $c(x,y,t+\Delta t) = c(x,y,t)$.

2.4 Evolution Equation

To implement our approach, we solve the advection-diffusion model over a regular grid with n spatial locations, denoted by $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. We assume a time step of $\Delta t = 1$. At each time $t \in \mathbb{N}$, define $\mathbf{c}_t = (c(\mathbf{s}_1,t), \ldots, c(\mathbf{s}_n,t))$ as the $n \times 1$ concentration vector. The discrete-time evolution equation can then be written as a vector autoregression

$$\mathbf{c}_{t+1} = \mathbf{\Phi}_t \mathbf{c}_t + \boldsymbol{\alpha}_t + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{7}$$

where $\mathbf{\Phi}_t$ is the $n \times n$ sparse transition matrix implied by the discretized advection-diffusion model (see Appendix), $\boldsymbol{\alpha}_t = (S(\mathbf{s}_1,t), \ldots, S(\mathbf{s}_n,t))'$ is the vector of source-sink terms, and $\boldsymbol{\omega}_t = (\omega(\mathbf{s}_1,t), \ldots, \omega(\mathbf{s}_n,t))'$ is the vector of model errors included to account for the various sources of uncertainty in the advection-diffusion model.

We assume that the model errors are independent in time but correlated in space, and specify their covariance at each time through a dimension-reduction approach. Specifically, we define $\mathbf{Q}_t = \mathbf{F}_t \mathbf{\Lambda} \mathbf{F}_t'$, where $\mathbf{F}_t$ is an $n \times q$ matrix of known coefficients and $\mathbf{\Lambda}$ is a $q \times q$ diagonal matrix with unknown parameters. In our application, we define the matrix $\mathbf{F}_t$ based on forcing variables $\boldsymbol{\tau}_t$, which allows the magnitude of the errors to depend on the forcing of the system. This also implies spatial dependence in the errors since the forcing variables are typically spatially correlated. Finally, we note that the concentrations are constrained to be nonnegative. We impose this restriction in the ensemble algorithms in Section 3 by setting negative concentrations to zero.

2.5 Measurement Equation

We assume the satellite images are available on the modeling grid at integer times. Let $\{\mathbf{r}_{t1}, \ldots, \mathbf{r}_{tm_t}\}$ denote the m_t observation locations at time t, which are a subset of $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. Let $\mathbf{y}_t = (y(\mathbf{r}_{t1},t), \ldots, y(\mathbf{r}_{tm_t},t))$ denote the $m_t \times 1$ vector of satellite measurements at time t. The measurement equation (1) can then be written in vector form as

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{c}_t) + \mathbf{Z}_t \boldsymbol{\beta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t), \tag{8}$$

where $\mathbf{h}_t(\mathbf{c}_t) = (h(c(\mathbf{r}_{t1},t)), \ldots, h(c(\mathbf{r}_{tm_t},t))) : \mathbb{R}^n \to \mathbb{R}^{m_t}$ is the vector measurement function; $\mathbf{Z}_t$ is the $m_t \times p$ matrix of covariates with ith row $\mathbf{z}(\mathbf{r}_{ti},t)'$; $\boldsymbol{\beta}_t$ is the $p \times 1$ vector of unknown bias coefficients; $\boldsymbol{\nu}_t = (\nu(\mathbf{r}_{t1},t), \ldots, \nu(\mathbf{r}_{tm_t},t))$ is the vector of measurement errors; and $\mathbf{R}_t$ is the observation covariance matrix with elements $R_{ij} = K(\|\mathbf{r}_{ti} - \mathbf{r}_{tj}\|; \boldsymbol{\theta}_\nu)$.

2.6 State Augmentation

In many cases, the source-sink term and observation bias are unknown and need to be estimated along with the concentrations. To do this, we use the idea of state augmentation (Evensen 2007; Stroud and Bengtsson 2007). Here we define an augmented state vector, $\mathbf{x}_t$, which includes the concentrations, the source term and the bias coefficients. We then specify evolution equations for $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$. One possible choice is to assume assume the coefficients are temporally independent. Another possibility is to use a smoothness prior, where $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ follow independent random walks. Under this assumption, the augmented state follows a autoregressive evolution equation $\mathbf{x}_{t+1} = \mathbf{\Phi}_t^* \mathbf{x}_t + \boldsymbol{\omega}_t^*$, where

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{c}_t \\ \boldsymbol{\alpha}_t \\ \boldsymbol{\beta}_t \end{pmatrix}, \qquad \mathbf{\Phi}_t^* = \begin{pmatrix} \mathbf{\Phi}_t & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}, \qquad \boldsymbol{\omega}_t^* = \begin{pmatrix} \boldsymbol{\omega}_{t1} \\ \boldsymbol{\omega}_{t2} \\ \boldsymbol{\omega}_{t3} \end{pmatrix}, \tag{9}$$

and $\boldsymbol{\omega}_t^* \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t^*)$, where $\mathbf{Q}_t^*$ is a block-diagonal covariance matrix. The evolution equation can then be combined with the observation equation $\mathbf{y}_t = \mathcal{H}_t(\mathbf{x}_t) + \boldsymbol{\nu}_t$, where $\mathcal{H}_t(\mathbf{x}_t) = \mathbf{h}_t(\mathbf{c}_t) + \mathbf{Z}_t \boldsymbol{\beta}_t$, and written as a nonlinear state-space model. We can then estimate the augmented state vector, $\mathbf{x}_t$, using the methods described in the next section.

3. STATE–SPACE FRAMEWORK

Let $\mathbf{y}_t$ denote the $m_t \times 1$ vector of satellite observations and let $\mathbf{x}_t$ denote the unobserved state vector at time $t \in \mathbb{N}$. The state may include the concentrations, the bias coefficients, or the source-sink term, depending on the setting. The observation and evolution equations can be written as a Gaussian state-space model of the form

$$\mathbf{y}_t = \mathcal{H}_t(\mathbf{x}_t) + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t), \tag{10}$$

$$\mathbf{x}_{t+1} = \mathcal{M}_t(\mathbf{x}_t) + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{11}$$

where $\mathcal{H}_t(\cdot)$ is the observation operator, $\mathcal{M}_t(\cdot)$ is the model operator, $\mathbf{R}_t$ is the observation error covariance matrix and $\mathbf{Q}_t$ is the model error covariance matrix. The model is completed with an initial distribution $\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{P}_0)$.

The main goal of our analysis is to estimate the space–time concentration fields given the sequence of images. The second goal is to predict concentrations at future time periods. These estimates, along with the associated uncertainties, can be obtained from the state filtering distribution $p(\mathbf{x}_t|\mathbf{Y}_t)$, the forecast distribution $p(\mathbf{x}_{t+k}|\mathbf{Y}_t)$ and the smoothing distribution $p(\mathbf{x}_{t-k}|\mathbf{Y}_t)$, where k is a positive integer and $\mathbf{Y}_t = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ denotes the observations up to time t.

For linear models, the state distributions are Gaussian and the corresponding moments are obtained using the Kalman filter and smoothing algorithms (Shumway and Stoffer 2006). However, when the state dimension is large (say greater than a few thousand), then the recursions become impracticable, requiring storage and multiplication of matrices of the dimension of the state. Furthermore, when the model is nonlinear, the state distributions are unavailable in closed form. Thus, we rely on sequential Monte Carlo algorithms, the ensemble Kalman filter and smoother, for state estimation in this high-dimensional and nonlinear setting. The algorithms are described below.

3.1 Ensemble Kalman Filter

The ensemble Kalman filter (EnKF; Evensen 1994) is a sequential Monte Carlo algorithm used to approximate the forecast and filtering distributions in nonlinear high-dimensional state-space models. In the EnKF, the state distribution is represented at each time period by an equally weighted sample or "ensemble" of states. The ensemble is propagated forward through time using the evolution equation and is updated using linear regression when new data arrive. In contrast to sequential importance sampling methods such as the particle filter (Gordon, Salmond, and Smith 1993), it does not reweigh or resample states. This allows the EnKF to remain stable in high dimensions and avoid sample degeneracy problems that hinder particle filters (Snyder et al. 2008).

We use a version of the EnKF known as *perturbed observations* (Burgers, van Leeuwen, and Evensen 1998). The algorithm is based on the following idea, which is known as *conditional simulation* in the geostatistics literature. Assume $(\mathbf{x}, \mathbf{y})$ are jointly normal, and we observe data $\mathbf{y}$. The goal is to simulate from the posterior distribution $\mathbf{x}^* \sim p(\mathbf{x}|\mathbf{y})$. To do this, we first generate the pair $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ from $p(\mathbf{x}, \mathbf{y})$. We then set $\mathbf{x}^* = \tilde{\mathbf{x}} + \mathrm{cov}(\mathbf{x}, \mathbf{y})\,\mathrm{var}(\mathbf{y})^{-1}(\mathbf{y} - \tilde{\mathbf{y}})$ to obtain the desired posterior draw. *Perturbed observations* uses the same idea but instead generates an ensemble of $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ pairs from $p(\mathbf{x}, \mathbf{y})$. The update is then completed for each ensemble member by replacing the variance and covariance with their sample estimates in the updating formula for $\mathbf{x}^*$. By Slutsky's theorem, it can be shown that the resulting draws of $\mathbf{x}^*$ converge to samples from $p(\mathbf{x}|\mathbf{y})$ as the ensemble size goes to infinity.

The EnKF algorithm proceeds as follows. Let $\{\mathbf{x}_t^{f(i)} : i = 1, \ldots, N\}$ and $\{\mathbf{x}_t^{u(i)} : i = 1, \ldots, N\}$ denote the forecast and filtered ensemble at time t, respectively. The algorithm is initialized at time $t = 0$ by drawing $\mathbf{x}_0^{u(i)} \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{P}_0)$, for $i = 1, \ldots, N$. The ensemble is then propagated forward through time, alternating between the forecast and update steps. Starting with the filtered ensemble at time $t-1$, the one-step-ahead forecasts at time t are obtained by drawing from the evolution equation

$$\mathbf{x}_t^{f(i)} = \mathcal{M}_{t-1}\big(\mathbf{x}_{t-1}^{u(i)}\big) + \boldsymbol{\omega}_{t-1}^{(i)}, \qquad \boldsymbol{\omega}_{t-1}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{t-1}). \tag{12}$$

This provides draws from the state forecast distribution $p(\mathbf{x}_t|\mathbf{Y}_{t-1})$. Due to their reduced-rank specification, the model errors, $\boldsymbol{\omega}_{t-1}$, can be generated efficiently by drawing q-dimensional normals.

If no data are available at time t, the update step is trivial and we set $\mathbf{x}_t^{u(i)} = \mathbf{x}_t^{f(i)}$ for $i = 1, \ldots, N$. If observations are available at time t, then we update the ensemble using the perturbed observations algorithm as described above. We first generate synthetic observations from the measurement equation

$$\mathbf{y}_t^{f(i)} = \mathcal{H}_t\big(\mathbf{x}_t^{f(i)}\big) + \boldsymbol{\nu}_t^{(i)}, \qquad \boldsymbol{\nu}_t^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t). \tag{13}$$

This provides samples from the joint state and observation forecast distribution $p(\mathbf{x}_t, \mathbf{y}_t|\mathbf{Y}_{t-1})$. Due to their stationarity assumption and the gridded domain, $\boldsymbol{\nu}_t$ can be generated efficiently in $\mathcal{O}(n \log n)$ operations using the circulant embedding approach of (Wood and Chan 1994). The update is completed using a linear regression step

$$\mathbf{x}_t^{u(i)} = \mathbf{x}_t^{f(i)} + \mathbf{K}_t\big(\mathbf{y}_t - \mathbf{y}_t^{f(i)}\big), \tag{14}$$

where $\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t'(\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t' + \mathbf{R}_t)^{-1}$ is the Kalman gain matrix, $\mathbf{H}_t$ is the linearized observation matrix with elements $H_{ij} = (\partial \mathcal{H}_i / \partial \mathbf{x}_j)(\boldsymbol{\mu}_t^f)$, and $\boldsymbol{\mu}_t^f$ and $\mathbf{P}_t^f$ are the sample forecast mean and covariance matrix computed from the ensemble, as described below.

Since the state and observation dimensions are large, computing and storing the full ensemble covariance and Kalman gain matrices is infeasible. To reduce storage costs and stabilize covariance estimation, we use a technique known as *covariance tapering* or localization (Houtekamer and Mitchell 1998). Here we define the forecast covariance matrix as

$$\mathbf{P}_t^f = \mathbf{C} \circ \left(\frac{1}{N-1} \sum_{i=1}^{N} \big(\mathbf{x}_t^{f(i)} - \boldsymbol{\mu}_t^f\big)\big(\mathbf{x}_t^{f(i)} - \boldsymbol{\mu}_t^f\big)' \right),$$

where $\boldsymbol{\mu}_t^f = N^{-1} \sum_{i=1}^{N} \mathbf{x}_t^{f(i)}$ is the ensemble forecast mean, $\circ$ denotes the Schur product, and $\mathbf{C}$ is a sparse correlation matrix defined over the model gridpoints. The tapering matrix $\mathbf{C}$ is defined through an isotropic correlation function with compact support (identically zero beyond some distance). The correlation function is typically chosen to be smooth at the origin with a tapering radius which is relatively small (Furrer and Bengtsson 2007). This preserves the ensemble-based correlation structure at short distances while removing spurious long-range correlations.

Covariance tapering provides a number of benefits. First, it regularizes the covariance matrix $\mathbf{P}_t^f$, increases its rank and guarantees that is positive definite (Furrer and Bengtsson 2007). This is important because the sample covariance matrix is severely rank deficient, as the ensemble size N is typically 10–100 while the matrix dimension is often on the order of 10,000 or more. Second, it preserves the flow-dependencies (i.e., spatial nonstationarities and anisotropies) due to model dynamics, which are represented in the forecast ensemble. Finally, it induces sparsity in the forecast covariance matrix, which reduces storage costs and speeds up matrix multiplications.

The main computational cost in our application arises in the update step (14). This involves computing $\mathbf{x}^{u(i)} = \mathbf{x}^{f(i)} + \mathbf{P}^f\mathbf{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}^{(i)}$, for each ensemble member $i = 1, \ldots, N$, where $\boldsymbol{\Sigma} = \mathbf{H}\mathbf{P}^f\mathbf{H}' + \mathbf{R}$ is the $m \times m$ innovation covariance matrix, $\boldsymbol{\epsilon}^{(i)} = \mathbf{y} - \mathbf{y}^{f(i)}$ is the $m \times 1$ innovation vector, and we omit time indices for simplicity. Since the observation dimension is large ($m > 3500$ in our application), direct matrix inversion of $\boldsymbol{\Sigma}$ is computationally expensive. Instead, we propose an efficient variational approach that exploits the sparsity of $\boldsymbol{\Sigma}$. We first solve the system $\boldsymbol{\Sigma}\mathbf{w}^{(i)} = \boldsymbol{\epsilon}^{(i)}$ iteratively using a conjugate gradient algorithm (Golub and Van Loan 1996). This algorithm requires only matrix–vector multiplications of the form $\boldsymbol{\Sigma}\mathbf{x}$, which can be performed efficiently using sparse matrix routines. The update is completed by setting $\mathbf{x}^{u(i)} = \mathbf{x}^{f(i)} + \mathbf{P}^f\mathbf{H}'\mathbf{w}^{(i)}$, which requires two sparse matrix–vector multiplications.

3.2 Likelihood Function

The ensemble Kalman filter also provides an approximate likelihood function for parameter estimation. Let Θ denote the vector of unknown parameters, which may include the physical parameters, the measurement parameters, and the error covariance parameters. For linear Gaussian state-space models (Shumway and Stoffer 2006), the likelihood function $L(\Theta)$ is given by

$$-2\log L(\Theta) = \sum_{t=1}^{T} \log|\boldsymbol{\Sigma}_t(\Theta)| + \boldsymbol{\epsilon}_t(\Theta)'\boldsymbol{\Sigma}_t(\Theta)^{-1}\boldsymbol{\epsilon}_t(\Theta) \quad (15)$$

(plus a constant), where $\boldsymbol{\epsilon}_t(\Theta)$ is the innovation vector and $\boldsymbol{\Sigma}_t(\Theta)$ is the innovation covariance matrix at time t obtained using the parameter value Θ. To approximate the likelihood function $L(\Theta)$, we run the EnKF for a fixed value of Θ and replace the innovation and its covariance matrix in (15) by $\boldsymbol{\epsilon}_t = N^{-1}\sum_{i=1}^{N}\boldsymbol{\epsilon}_t^{(i)}$ and $\boldsymbol{\Sigma}_t = \mathbf{C} \circ ((N-1)^{-1}\sum_{i=1}^{N}(\boldsymbol{\epsilon}_t^{(i)} - \boldsymbol{\epsilon}_t)(\boldsymbol{\epsilon}_t^{(i)} - \boldsymbol{\epsilon}_t)')$. The likelihood can then be maximized numerically using Quasi-Newton or simplex methods. To obtain a smooth likelihood surface for maximization, we adopt the approach of Pitt (2002), and use common random numbers for each likelihood evaluation.

3.3 Ensemble Kalman Smoother

The ensemble Kalman smoother (EnKS; Evensen and van Leeuwen 2000) provides approximate samples from the smoothing distribution $p(\mathbf{x}_t|\mathbf{Y}_T)$, for each time t. Let $\{\mathbf{x}_t^{s(i)} : i = 1, \ldots, N\}$ denote the smoothed ensemble at time t. The smoothing algorithm requires two passes through time. We first run the EnKF forward for $t = 0, \ldots, T$, storing the forecast and filtered ensembles at each time t. We then run a backward pass to obtain the smoothed ensemble. The smoother is initialized at time T by setting $\mathbf{x}_T^{s(i)} = \mathbf{x}_T^{u(i)}$ for $i = 1, \ldots, N$. We then proceed backward for times $t = T-1, \ldots, 0$, using the recursive updating rule

$$\mathbf{x}_t^{s(i)} = \mathbf{x}_t^{u(i)} + \mathbf{B}_t\big(\mathbf{x}_{t+1}^{s(i)} - \mathbf{x}_{t+1}^{f(i)}\big), \qquad i = 1, \ldots, N. \quad (16)$$

Here $\mathbf{B}_t = \mathbf{P}_t^u\mathbf{M}_t'(\mathbf{P}_{t+1}^f)^{-1}$ is the backward gain matrix, $\mathbf{M}_t$ is the linearized evolution matrix with elements $M_{ij} = (\partial\mathcal{M}_i/\partial x_j)(\boldsymbol{\mu}_t^u)$, $\boldsymbol{\mu}_t^u$ is the filtered mean, and $\mathbf{P}_t^u$ and $\mathbf{P}_{t+1}^f$ are the filtered and forecast covariance matrices computed from the respective ensembles.

As in the EnKF, covariance tapering is used to improve covariance estimation in the smoothing algorithm. This provides substantial computational savings because the recursion (16) is applied at each time period and involves matrices of the dimension of the state rather than the observation. The smoothing recursion is implemented in two steps. We first solve the system $\mathbf{P}^f\mathbf{w}^{(i)} = \boldsymbol{\epsilon}^{(i)}$, where $\boldsymbol{\epsilon}^{(i)} = \mathbf{x}^{s(i)} - \mathbf{x}^{f(i)}$. We then set $\mathbf{x}_t^{s(i)} = \mathbf{x}_t^{u(i)} + \mathbf{P}^u\mathbf{M}'\mathbf{w}^{(i)}$. The first step is implemented by variational methods using the conjugate gradient algorithm. The second step is computed efficiently with two sparse matrix–vector operations. This provides an ensemble-based approximation to the smoothing distribution $p(\mathbf{x}_t|\mathbf{Y}_T)$ at each time period. In the next section, we apply the EnKF and EnKS algorithms described above to a study of sediment transport in Lake Michigan.

4. CASE STUDY OF LAKE MICHIGAN

The Episodic Events Great Lakes Experiment (EEGLE) was an intensive data collection effort from 1998–2000 sponsored by NOAA and the National Science Foundation, aimed to study the impact of episodic events on Great Lakes ecosystems. We consider a one-month period, March 1998, when a large storm event occurred and the development of a large sediment plume (50 km wide) was observed in satellite images of Lake Michigan (see Figures 1 and 2). The goal of the analysis is to provide a complete picture of the space–time development of the suspended sediment field, which can aid understanding of the physical process and help to calibrate the parameters and input variables of a numerical sediment transport model.

Figure 3 shows a time series of observed wind vectors during the March 1998 modeling period at buoy 45002, located in the northern basin of Lake Michigan. The plot shows the development of the storm event, which began around March 8 and produced winds in excess of 20 m/s for the first 24 hours. The initial winds were from the east, but quickly shifted to the north where the peak winds occurred on March 9–10. After the initial event, three additional wind bursts occurred during the month: a strong wind from the south beginning on March 13, a northern wind starting on March 19, and a southern wind on March 25.

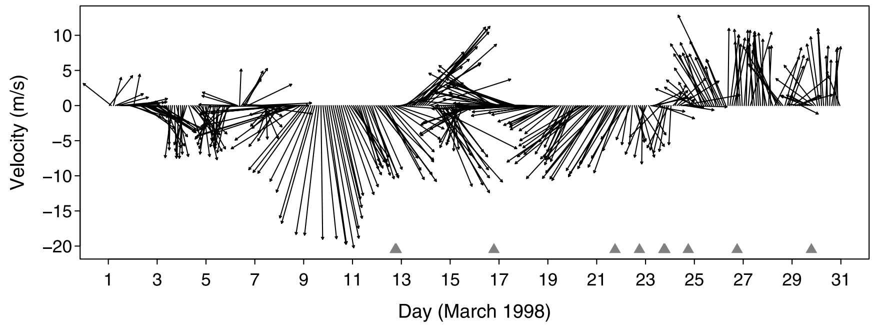

Figure 3. Observed wind vectors during the March 1998 modeling period at a buoy located in the center of northern Lake Michigan. The triangles along the horizontal axis indicate the 10 satellite image times. Note that there are two images on March 12 and 23.

4.1 Satellite Data

Satellite radiance data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) were downloaded from the National Aeronautic and Space Administration's Goddard Space Flight Center web page (*http://oceancolor.gsfc.nasa.gov/SeaWiFS*). The raw radiances were processed using the SeaDAS software (Baith et al. 2001), which includes an atmospheric correction to convert satellite radiances to water-leaving radiances, and a mapping onto the 2-km modeling grid. The derived water-leaving radiances were then converted to remote-sensing reflectances (RSR), which are recorded on a percentage scale. The available data consisted of RSR in eight spectral bands, six in the visible and two in the near infrared. After exploring different relationships between in situ sediment concentrations and RSR in different bands, we found that Band 5 (555 nm), located in the green part of the spectrum, provided the strongest relationship to TSM. Thus we use RSR(555) as our satellite data throughout the analysis.

Before running the analysis, a cloud-screening algorithm was applied to the raw radiances. We defined clouds as pixels having an albedo (satellite reflectance at 865 nm) value of greater than 1.25%. This resulted in a large number of pixels being removed during our March 1998 modeling period, with many of the images being nearly completely cloud covered. Thus, to simplify matters, we limit our analysis to the southern basin of Lake Michigan and use only images with at least half of the southern basin pixels ($m_t > 3500$) cloud free. This provides 10 valid images during the March 1998 modeling period, whose times are indicated in Figure 3.

4.2 Measurement Function

As part of the EEGLE study, in situ measurements were collected in Lake Michigan, primarily along five transects in the southern basin of the lake: Chicago, Gary, St. Joseph, Muskegon, and Racine. Additional samples were taken at a deep water station and at auxiliary stations near Chicago, Michigan City, Saugatuck, and Waukegan. Figure 4 shows a map of southern Lake Michigan along with the measurement locations. A total of 52 surface measurements of sediment concentration were collected during the study.

To derive a measurement function relating sediment concentration to satellite reflectance, we matched the 52 available in situ sediment concentration measurements with the nearest cloud-free satellite reflectance value in space and time. Figure 4 shows a scatterplot of the matched reflectance–sediment concentration pairs. The plot indicates that the relationship between reflectance and sediment concentration is roughly linear for low concentrations and logarithmic for high concentrations. To capture these features, we choose a nonlinear measurement function of the form

$$h(C;\boldsymbol{\theta}) = \theta_0 + \theta_1 \log(1 + \theta_2(C + \theta_3)), \tag{17}$$

where C is the sediment concentration and $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ is a vector of unknown parameters. We fit the model to the 52 observations using nonlinear least squares, and obtained the parameter estimates $\hat{\boldsymbol{\theta}} = (0.003, 0.054, 0.474, 0.55)$. The fitted measurement function and the calibration data are shown in Figure 4. The parameters were fixed at these estimates in the analysis below.

4.3 Sediment Transport Model

To describe the space–time evolution of the suspended sediment field, we assume a two-dimensional sediment transport model which includes advection, sources and sinks (no diffusion). The model for the depth-averaged suspended sediment concentration, $C = C(x, y, t)$, is given by

$$\frac{\partial(HC)}{\partial t} = -\frac{\partial(HUC)}{\partial x} - \frac{\partial(HVC)}{\partial y} + S, \tag{18}$$

where $H = H(x, y)$ are the water depths, $U = U(x, y, t)$ and $V = V(x, y, t)$ are the depth-averaged water currents, and $S = S(x, y, t)$ is the source-sink term, which incorporates settling and resuspension processes:

$$S = \begin{cases} -\psi_1(HC) + \psi_2\left(\dfrac{\tau}{\psi_3} - 1\right), & \text{if } \tau \ge \psi_3 \\ -\psi_1(HC), & \text{if } \tau < \psi_3, \end{cases} \tag{19}$$

where $\tau = \tau(x, y, t)$ is the bottom shear stress. The model includes three parameters: the settling rate, ψ_1; the resuspension rate, ψ_2; and the critical shear stress required to cause a resuspension event, ψ_3. We denote these physical parameters collectively by $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)$. The inputs for the model are the water depths, $H(x, y)$, the water velocities, $u(x, y, t)$ and $v(x, y, t)$, and the bottom shear stress, $\tau(x, y, t)$. The input variables are assumed to be known for the analysis, while the model parameters $\boldsymbol{\psi}$ are estimated using maximum likelihood, as described below.

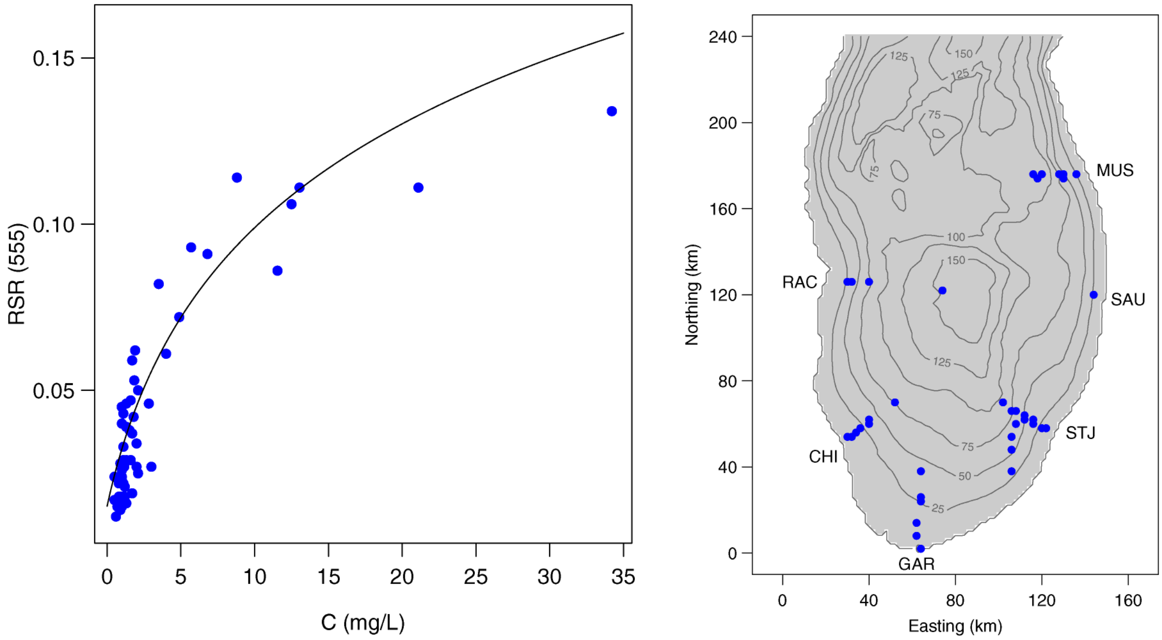

Figure 4. Nonlinear relationship between sediment concentration and remote sensing reflectance. Left: matched in situ sediment measurements and satellite reflectance observations and the fitted measurement function Equation (17). Right: map of southern Lake Michigan with the locations of the in situ sediment measurements. The contours in the map represent water depths (in meters). The three-letter abbreviations indicate the transect names.

The system is solved over a 2-km modeling grid of Lake Michigan at an hourly time step using a first-order upwind finite-difference scheme (Schwab and Beletsky 2002; Lee et al. 2007). This scheme is a slight modification of the method described in Section 2.2, but also results in a linear discrete-time system. The grid dimensions are 131×251 which includes $n = 14{,}558$ water pixels, and the total number of hourly time steps is $T = 744$ during the modeling period. The lateral boundary conditions assume no sources, and the bottom boundary condition assumes that the sediment bed is an unlimited source. The initial sediment concentrations were unknown and assumed to follow a Gaussian distribution.

The input variables to the sediment model include the gridded water velocities and bottom shear stress at each hour during the modeling period. The gridded velocity fields $u(x, y, t)$ and $v(x, y, t)$ are obtained by taking depth averages of hourly output from a three-dimensional hydrodynamic model (Beletsky et al. 2003), which is based on the Princeton Ocean Model (Blumberg and Mellor 1987). The gridded shear stress fields $\tau(x, y, t)$ are defined as $\tau = \sqrt{\tau_1^2 + \tau_2^2}$, where τ_1 and τ_2 are the sheer stresses due to advection and waves, respectively. The former is obtained as a deterministic function of the water currents, while the latter is obtained from a numerical wave model (Schwab et al. 1984). The hydrodynamic and wave models are forced by gridded wind fields derived from observations at 18 National Weather Service stations and National Data Buoy Center buoy 45002. The modeled velocities and bottom shear stress fields at three image times during March 1998 are shown in Figure 5.

4.4 Results

The filtering and smoothing algorithms described in Section 3 were run using the following specifications. The observation bias was modeled as a spatial constant [$z(\mathbf{s}, t) = 1$ and $\boldsymbol{\beta}_t = \beta_t$], and the bias coefficients were assumed to be independent over time, $\beta_t \sim \mathcal{N}(0, 0.01)$ for each t. The observational errors were assumed to follow a tapered exponential covariance model with unknown sill and range parameters σ_v^2 and ρ_v. The model errors were specified using the dimension-reduction approach with $q = 6$ basis functions defined by the shear stress fields $\boldsymbol{\tau}_t$, and $\boldsymbol{\Lambda}$ depends on an unknown variance parameter σ_ω^2. Finally, we assumed the initial distribution $\mathbf{c}_0 \sim \mathcal{N}(\bar{\mathbf{c}} + \mu_0 \mathbf{1}', \sigma_0^2 \mathbf{I})$ where $\bar{\mathbf{c}}$ is the climatological mean field in Stroud et al. (2009) and μ_0 and σ_0 are unknown parameters.

The tapering matrix $\mathbf{C}$ was defined using the 5th-order polynomial correlation function of Gaspari and Cohn (1999), with a cutoff radius of $r = 3$ pixels (6 km). The radius was chosen based on a grid search where we considered values for r from 1 to 10 pixels. We found that increasing r beyond 3 pixels increased the value of the likelihood function, but had little effect on the parameter estimates and the predictions. Given these results and the added computational burden of a larger radius, we used $r = 3$ pixels throughout the analysis.

The algorithms were run with and without observation bias. When bias was included in the model, the filter and smoother implementations were slightly modified from Section 3. Since the bias coefficient is spatially constant, the tapering function was applied only to the first $n \times n$ block of the $(n+1) \times (n+1)$ state covariance matrix. The filtering algorithm proceeds as described in Section 3.1, with the bias coefficients being generated from the prior distribution $\beta_t^{(i)} \sim \mathcal{N}(0, 0.01)$ for each t. The update step is performed as in Equation (14) to obtain posterior samples of the augmented state $\mathbf{x}_t = (\mathbf{c}_t, \beta_t)$. Retrospective state estimation is performed using the ensemble Kalman smoother recursion in Equation (16), and the taper is again applied only to the first $n \times n$ block of the state covariance matrices.

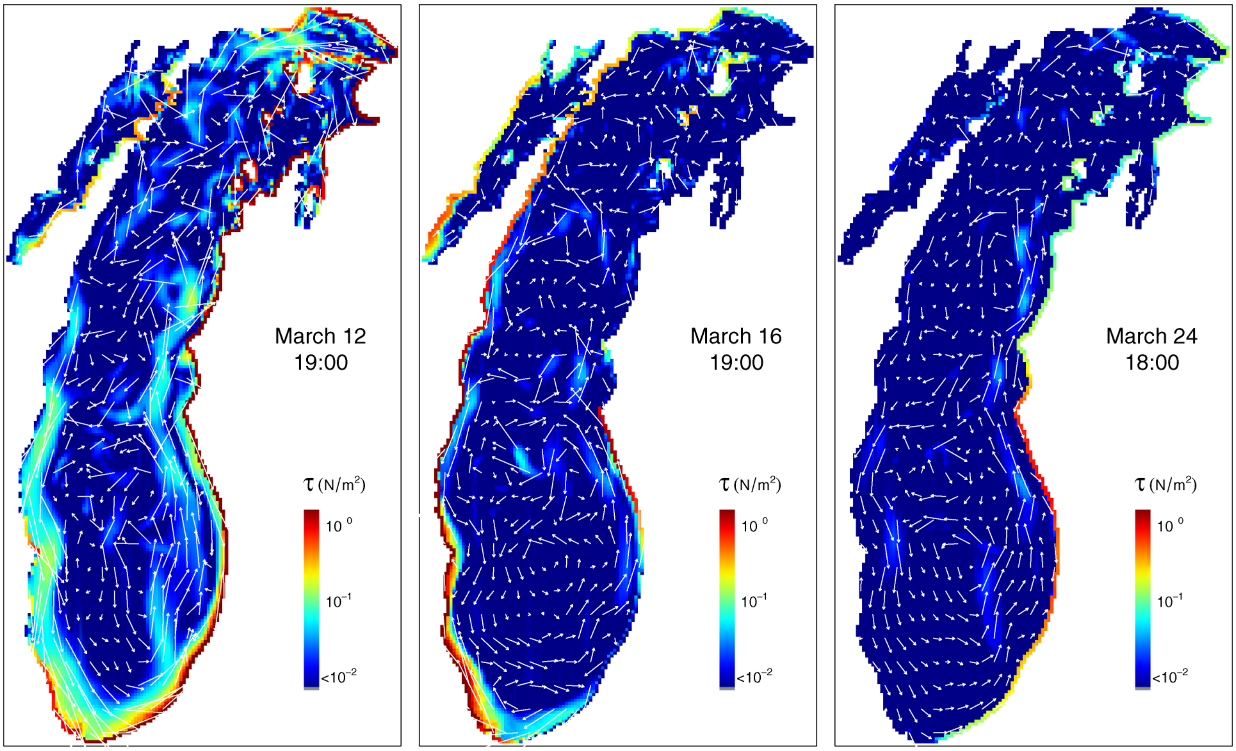

Figure 5. Vertically averaged velocity fields and bottom shear stress in Lake Michigan on March 12, 16, and 24, 1998. The shear stresses are shown as image intensities, while the velocities are shown as arrows (at every seventh pixel). The velocities are obtained as output from the Princeton Ocean Model (POM); the shear stresses are obtained by combining output from the POM and a wave model. All times are GMT.

We first ran a Newton–Raphson algorithm to obtain maximum likelihood estimates for the parameters $\Theta = (\psi_1, \psi_2, \psi_3, \sigma_\nu, \rho_\nu, \sigma_\omega, \mu_0, \sigma_0)$. The estimates are presented in Table 1. After fixing the parameters at their estimates, we ran the ensemble filtering and smoothing algorithms to obtain sequential and retrospective state estimates. Throughout the analysis, we used an ensemble size of $N = 25$ (larger ensemble sizes were considered but did not substantially change the results). The computational run time for the filtering and smoothing algorithms was about seven minutes for the March 1998 modeling period (744 hourly time steps), using C code on a 8-core 2.8 GHz Intel Xeon processor with 12 GB of memory. (The data and code are available in the online supplements.)

Figure 6 shows the satellite images and the one-image-ahead forecast mean and standard deviation at eight times in March 1998, using bias correction. Also shown are the water velocities at the same time periods. The satellite image on March 12 shows a spiral-shaped sediment plume extending roughly 50 km off the eastern shore. The March 16 image shows an enlarged plume which has shifted to the northwest. In subsequent images, the plume is advected westward, reaching about 100 km off the east coast on March 29. The EnKF forecasts do an excellent job predicting the movement of the sediment plume, closely tracking its location and shape. The forecast standard deviations are also quite reasonable, with uncertainties that are roughly proportional to the estimated concentrations. We note that the algorithm also provides realistic forecasts at locations with missing data.

Table 1. Maximum likelihood estimates for the parameters in the static model (M0) and dynamic model (M1). Note that the physical parameters (ψ_1, ψ_2, ψ_3) are undefined in the static model

Parameter	Interpretation	M0	M1	Units
ψ_1	Settling rate	–	6.02×10^{-6}	m/s
ψ_2	Resuspension rate	–	1.60×10^{-8}	$kg/m^2/s$
ψ_3	Critical shear stress	–	0.284	N/m^2
σ_ν	Observation SD	0.012	0.007	RSR
ρ_ν	Observation range	2.000	2.000	km
σ_ω	Evolution SD	0.109	0.063	mg/L
μ_0	Initial mean	0.766	0.379	mg/L
σ_0	Initial SD	0.014	0.368	mg/L

Figure 7 illustrates the use of the ensemble smoother for temporal interpolation. This plot shows the satellite images on March 12, 16, and 21 along with the filtered and smoothed means at eight times within the interval. To obtain the state estimates, the EnKF was run forward from the initial time, assimilating only the images on March 12 and 21; the March 16 image was withheld for validation purposes. The ensemble Kalman smoother was then run backward from March 21 to the initial time. Of interest here is the comparison of the forecast and smoothed means on March 16 to the withheld image. We see that, while the forecast estimates have a more physically coherent structure for the plume, the smoothed estimates provide better predictions along the eastern coastline near Saugatuck (SAU in Figure 4), closely matching the width and shape of the sediment band.

Figure 8 compares our approach to the reduced-rank square-root Kalman filter (RRSQRT-KF, Verlaan 1998), which is a widely used technique in oceanographic data assimilation

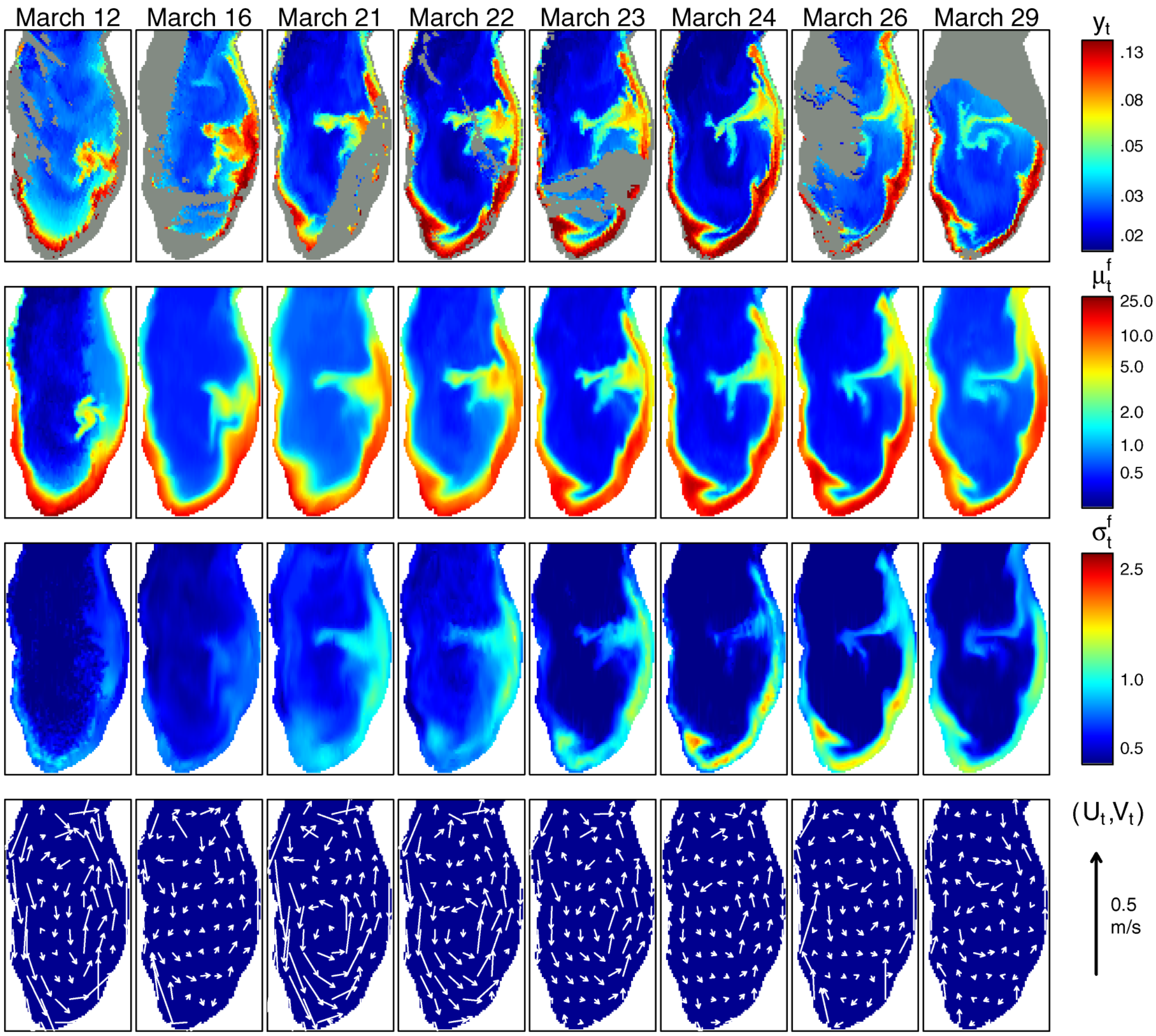

Figure 6. Satellite data and one-image-ahead forecasts of suspended sediment concentration at eight image times. Top row: satellite reflectance data. Second row: forecast mean. Third row: forecast standard deviation. Bottom row: depth-averaged water velocities. Sediment concentrations are in units of mg/L.

(Bertino, Evensen, and Wackernagel 2003). The RRSQRT is a dimension-reduction approach designed for high-dimensional linear systems. As shown in the figure, this approach performs quite poorly in our problem, removing small-scale features in the images. The problem with this method is that it reduces all of the spatial information to a small set of coefficients in the Kalman filter update step. In contrast, our approach produces excellent results, retaining the detailed spatial structures (the plume) from the satellite images. While the computational cost for the two approaches is roughly the same, our approach reduces forecast RMSE by more than 30% relative to the RRSQRT-KF.

Table 2 presents numerical results from a cross-validation study and compares our model to a simpler model with no dynamics. To obtain these results, we made 10 separate smoothing runs, one for each image. For the forecasting results, we ran the EnKF to one hour before the image time and generated a one-step-ahead prediction for the withheld image. For the smoothing results, we ran an EnKF to the last time period, ignoring the update for the withheld image, and then ran the EnKS backwards to the image time. For comparison, we also performed the same computations using a model with no dynamics (i.e., $\boldsymbol{\Phi}_t = \mathbf{I}$ and $\boldsymbol{\alpha}_t = \mathbf{0}$). This is referred to as the *persistence approach* in the forecasting literature. The persistence model was also run using optimized parameters, which are listed in Table 1. For both methods, we computed the forecast and smoothed root mean squared error (RMSE) by comparing the ensemble mean to the satellite data, both transformed to the log RSR scale.

The numerical results in Table 2 show that the dynamic modeling approach reduces the forecast and smoothing RMSE by 27% and 21%, respectively, relative to the persistence approach. However, these numbers understate the performance of the dynamic approach, as they combine the results for all ten images. We note that the largest forecast improvements correspond to the longer lead times (e.g., March 16 and 29), while the smallest improvements correspond to shorter lead times (March 12, 22, and 23). For example, the March 16 image, which is the only image within a nine-day interval, provides forecast and smoothing improvements of 38% and 24% relative to the persistence approach. This indicates that the dynamic model substantially improves predictions when the time interval between images is large.

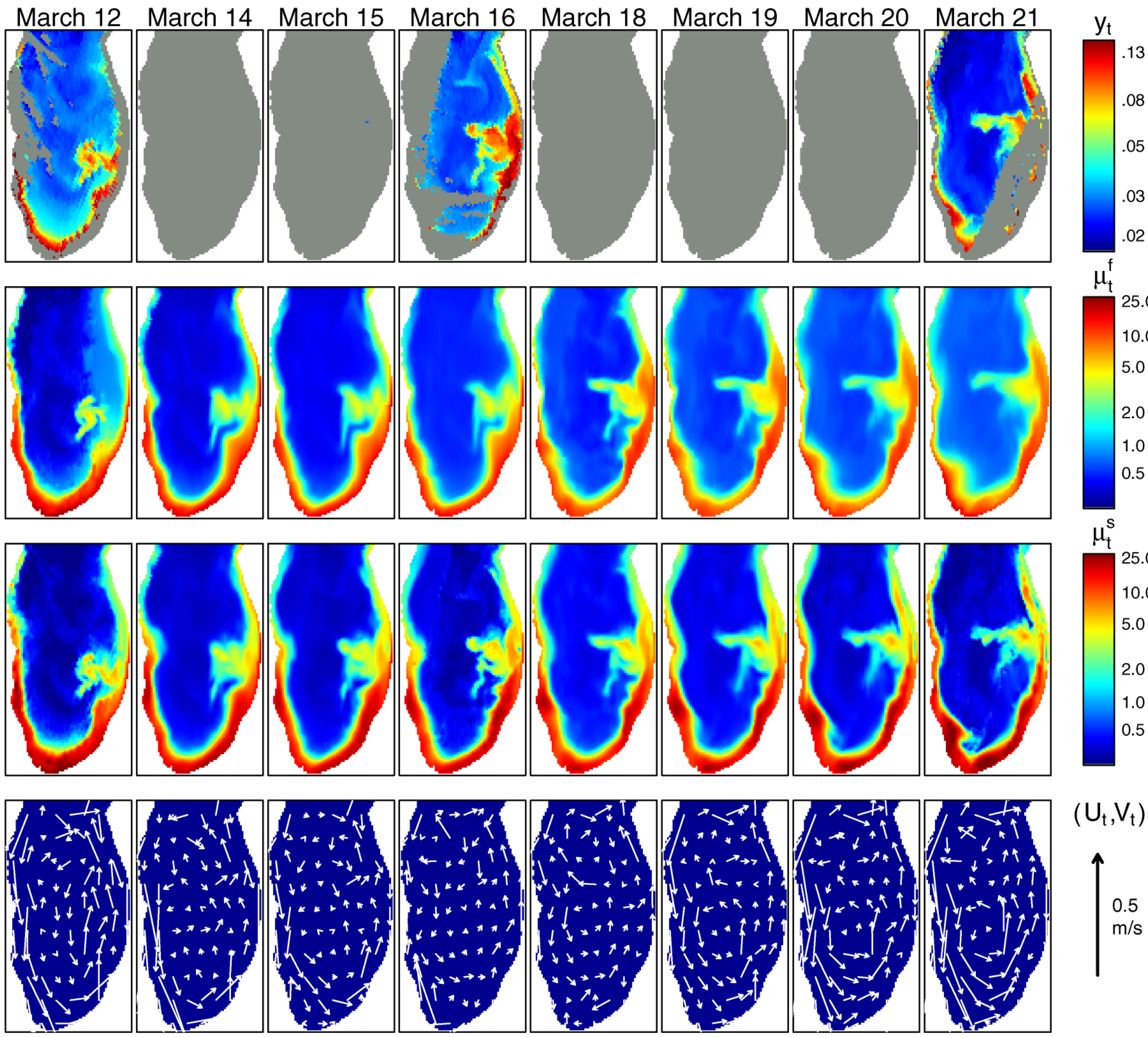

Figure 7. Satellite data and forecast and smoothed estimates of suspended sediment concentration at selected times. Top row: satellite reflectance data. Second row: forecast mean. Third row: smoothed mean. Bottom row: vertically averaged water velocities. Note that the March 16th image was not used in the assimilation. Sediment concentrations are in units of mg/L.

5. CONCLUSIONS

We have proposed a class of dynamic spatio-temporal models for satellite images based on advection-diffusion models. The model provides sequential and retrospective estimates of an unknown concentration field along with associated uncertainties over space and time. Our method handles the nonlinearies, high dimensionality, measurement bias, and missing data common in satellite images, and allows for fast computation through the use of covariance tapering. To obtain state and bias estimates, we rely on the ensemble Kalman filter and smoothing algorithms which have become extremely popular in atmospheric and oceanographic data assimilation over the last decade (see Geir Evensen's EnKF website at *http://enkf.nersc.no*). In this context, we provided two methodological innovations: a variational updating scheme for high-dimensional observations with correlated errors, and a variational ensemble Kalman smoother for retrospective state estimation.

Using a sequence of satellite images from Lake Michigan during a storm event, we applied our method to produce hourly forecast and smoothed maps of sediment concentration over a one-month period. We compared our approach to two other methods: a state-space model with a static evolution equation, and a reduced rank square-root Kalman filter (RRSQRT-KF), which is widely used for oceanographic data assimilation. We showed that our method improved forecast root mean squared error by 25% relative to the static model and 30% relative to the RRSQRT-KF. Larger improvements were obtained for longer forecast lead times. The proposed methods could be applied to a wide range of environmental variables, such as atmospheric aerosols, particulate matter, or total column ozone.

An interesting direction for future research is to use satellite images to jointly estimate the velocity fields and tracer concentrations. While conceptually straightforward, this presents challenges due to the nonlinear hydrodynamic model which governs the velocities. Using our ensemble approach, this could be carried out by augmenting the state vector to include the velocity fields. Although this would imply a nonlinear evolution for the state, it would require only minor changes in the algorithms presented here. Zhang et al. (2007) have proposed a method for assimilating current measurements into a hydrodynamic model of Lake Michigan, and we have recently begun work to combine the two ideas.

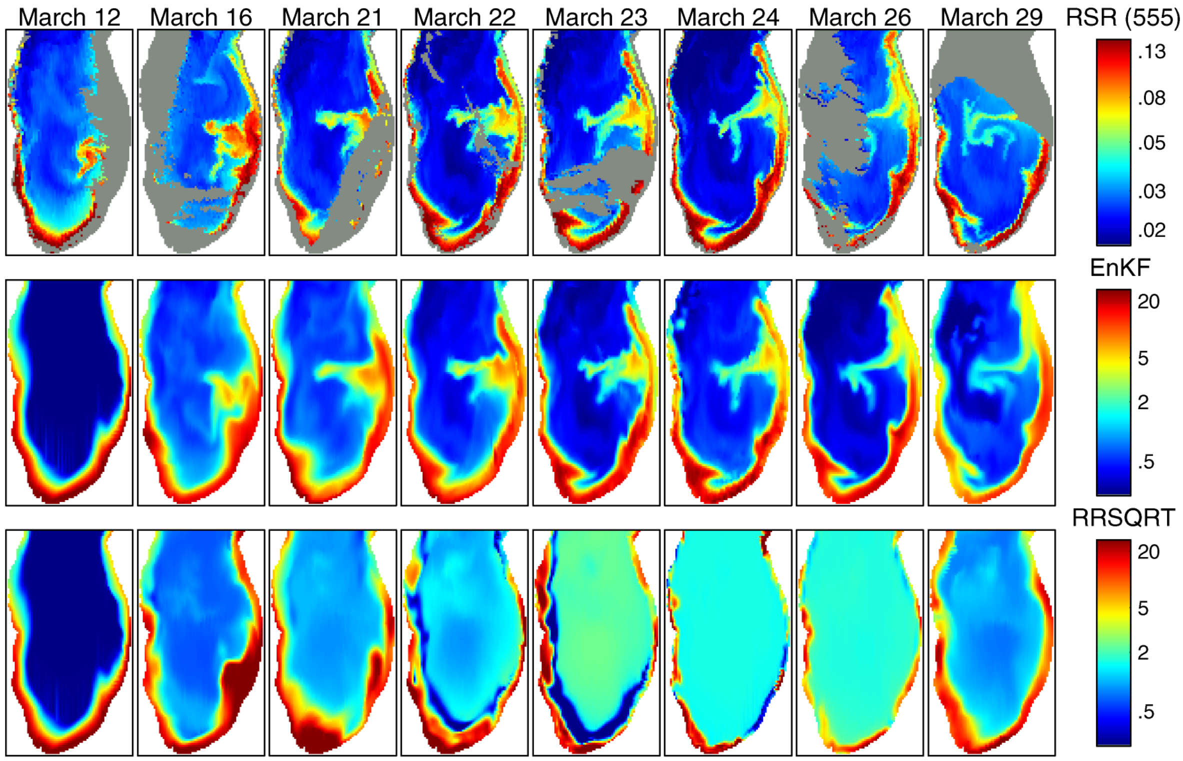

Figure 8. Comparison of forecast mean suspended sediment concentration from the ensemble Kalman filter (EnKF) and a reduced-rank square root Kalman filter (RRSQRT-KF). Top row: satellite data. Second row: EnKF forecast mean. Bottom row: RRSQRT-KF forecast mean. Sediment concentrations are in units of mg/L.

APPENDIX: DEFINITION OF $\mathbf{\Phi}_t$

Here, we define the transition matrix $\mathbf{\Phi}_t$ implied by the forward-time central-space discretization scheme in Section 2.3. Let Δt denote the model time step. We assume a two-dimensional $(I \times J)$ spatial modeling grid with spacing Δx and Δy, so that the model grid coordinates are $(x_i, y_j) = (i\Delta x, j\Delta y)$, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$. (Time subscripts are omitted throughout the rest of the appendix.) Let $c_{ij} = c(x_i, y_j, t)$ denote the concentration at location (x_i, y_j) and time t, and let $u_{ij} = u(x_i, y_j, t)$ and $v_{ij} = v(x_i, y_j, t)$ denote the corresponding x- and y-velocities. The coefficients of the transition matrix at time t are given by

$$\phi^1_{ij} = 1 - \left(\frac{2D}{(\Delta x)^2} + \frac{2D}{(\Delta y)^2}\right)\Delta t,$$

$$\phi^2_{ij} = \left(\frac{D}{(\Delta x)^2} - \frac{u_{ij}}{2\Delta x}\right)\Delta t,$$

$$\phi^3_{ij} = \left(\frac{D}{(\Delta x)^2} + \frac{u_{ij}}{2\Delta x}\right)\Delta t,$$

$$\phi^4_{ij} = \left(\frac{D}{(\Delta y)^2} - \frac{v_{ij}}{2\Delta y}\right)\Delta t,$$

$$\phi^5_{ij} = \left(\frac{D}{(\Delta y)^2} + \frac{v_{ij}}{2\Delta y}\right)\Delta t.$$

To define the transition matrix at time t, we order the gridpoints row by row and let $\mathbf{c} = (c_{11}, c_{21}, \ldots, c_{ij}, \ldots, c_{IJ})'$ denote the $IJ \times 1$ concentration vector at time t. We then define the transition matrix at time t as

$$\mathbf{\Phi} = \begin{pmatrix} \phi^1_{11} & \phi^2_{11} & \mathbf{0}' & \phi^4_{11} & & & & & \\ \phi^3_{21} & \phi^1_{21} & \phi^2_{21} & \mathbf{0}' & \phi^4_{21} & & & & \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & \phi^5_{ij} & \mathbf{0}' & \phi^3_{ij} & \phi^1_{ij} & \phi^2_{ij} & \mathbf{0}' & \phi^4_{ij} & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & & \phi^5_{IJ} & \mathbf{0}' & \phi^3_{IJ} & \phi^1_{IJ} \end{pmatrix},$$

Table 2. Forecast and smoothed root mean squared error for the 10 images during March 1998. M0 and M1 denote the static and dynamic model, respectively. Results are in units of log RSR

Satellite Images			Forecast		Smoothing	
Date	Time	Nobs	M0	M1	M0	M1
3/12	17:34	5398	0.380	0.290	0.241	0.195
3/12	19:14	5491	0.228	0.181	0.219	0.149
3/16	18:54	4580	0.380	0.234	0.313	0.237
3/21	17:42	5414	0.400	0.283	0.239	0.157
3/22	18:26	6600	0.382	0.259	0.272	0.203
3/23	17:32	5646	0.302	0.208	0.242	0.193
3/23	19:12	6291	0.234	0.199	0.203	0.192
3/24	18:17	7079	0.333	0.240	0.323	0.251
3/26	18:08	4176	0.229	0.242	0.215	0.230
3/29	18:44	4146	0.301	0.221	0.301	0.221
March	Total	54,821	0.325	0.238	0.261	0.205

where $\mathbf{0}'$ is the $(I-1)$-vector of zeros. Note that the transition matrix $\boldsymbol{\Phi}$ has at most five non-zero coefficient per row. Hence, the matrix can be stored in a sparse format and matrix multiplications of the form $\boldsymbol{\Phi}\mathbf{c}$ and $\boldsymbol{\Phi}'\mathbf{c}$ can be computed efficiently in $\mathcal{O}(IJ)$ operations by exploiting sparse matrix routines.

SUPPLEMENTAL MATERIALS

Data: The file contains data from the 10 satellite images used in the paper. The units of the data are log of remote sensing reflectance at 555 nm [$\log \mathrm{RSR}(555)$]. The dimensions of the file are 32881×10. Each column represents a different image and each row represents a different grid point. The grid dimensions are 131×251 and the points are ordered lexicographically. Below is a short R script to scan and plot the 10 images. (satellite-data.txt)

```
#######################################################
## R code to plot the 10 images described in Table 2
#######################################################
data = read.table("satellite-data.txt")
par(mfrow=c(3,4),mar=c(3,3,1,1),las=1)
for
(i in 1:10){
  zmat = matrix(data[,i],131,251)
  image(1:131,1:251,zmat,xlim=c(0,80),ylim=c(0,140))
}
```

[Received December 2007. Revised August 2009.]

REFERENCES

Baith, K., Lindsay, R., Fu, G., and McClain, C. (2001), "SeaDAS, a Data Analysis System for Ocean-Color Satellite Sensors," *EOS, Transactions of the American Geophysical Union*, 82, 202. [984]

Beletsky, D., Schwab, D. J., Roebber, P. J., McCormick, M. J., Miller, G. S., and Saylor, J. H. (2003), "Modeling Wind-Driven Circulation During the March 1998 Sediment Resuspension Event in Lake Michigan," *Journal of Geophysical Research*, 108, 3038. [985]

Bertino, L., Evensen, G., and Wackernagel, H. (2003), "Sequential Data Assimilation Techniques in Oceanography," *International Statistical Review*, 71, 223–242. [987]

Blumberg, A. F., and Mellor, G. L. (1987), "A Description of a Three-Dimensional Coastal Ocean Circulation Model," in *Three-Dimensional Coastal Ocean Models, Coastal Estuarine Studies*, ed. N. Heaps, Vol. 5, Washington, DC: American Geophysical Union, pp. 1–16. [985]

Brown, P. E., Karesen, K. F., Roberts, G. O., and Tonellato, S. (2000), "Blur-Generated Non-Separable Space–Time Models," *Journal of the Royal Statistical Society, Ser. B*, 62, 847–860. [978]

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998), "Analysis Scheme in the Ensemble Kalman Filter," *Monthly Weather Review*, 126, 1719–1724. [982]

Evensen, G. (1994), "Sequential Data Assimilation With a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics," *Journal of Geophysical Research*, 99, 10143–10162. [980,982]

——— (2007), *Data Assimilation: The Ensemble Kalman Filter*, New York: Springer. [981]

Evensen, G., and van Leeuwen, P. J. (2000), "An Ensemble Kalman Smoother for Nonlinear Dynamics," *Monthly Weather Review*, 128, 1852–1867. [983]

Furrer, R., and Bengtsson, T. (2007), "Estimation of High-Dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants," *Journal of Multivariate Analysis*, 98, 227–255. [983]

Gaspari, G., and Cohn, S. (1999), "Construction of Correlation Functions in Two and Three Dimensions," *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. [985]

Gelpke, V., and Künsch, H. (2001), "Estimation of Motion From Sequences of Images: Daily Variablility of Total Ozone Mapping Spectometer Ozone Data," *Journal of Geophysical Research D*, 106, 11,825–11,834. [978]

Golub, G., and Van Loan, C. (1996), *Matrix Computations*, Baltimore: Johns Hopkins University Press. [983]

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," in *IEE Proceedings*, Vol. F-140, IEE, pp. 107–113. [982]

Higdon, D. (2002), "Space and Space–Time Modeling Using Process Convolutions," in *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, London: Springer-Verlag, pp. 37–56. [978]

Houtekamer, P. L., and Mitchell, H. L. (1998), "Data Assimilation Using an Ensemble Kalman Filter Technique," *Monthly Weather Review*, 126, 796–811. [982]

Huang, H.-C., and Hsu, N.-J. (2004), "Modeling Transport Effects on Ground-Level Ozone Using a Non-Stationary Space–Time Model," *Environmetrics*, 15, 251–268. [978]

Johannesson, G., Cressie, N. A. C., and Huang, H.-C. (2007), "Dynamic Multi-Resolution Spatial Models," *Environmental and Ecological Statistics*, 14, 5–25. [978]

Lee, C., Schwab, D. J., Beletsky, D., Stroud, J., and Lesht, B. (2007), "Numerical Modeling of a Mixed Sediment Resuspension, Transport, and Deposition During the March 1998 Episodic Events in Southern Lake Michigan," *Journal of Geophysical Research—Oceans*, 112, C02018. [985]

Niu, X., and Tiao, G. C. (1995), "Modeling Satellite Ozone Data," *Journal of the American Statistical Association*, 90, 969–983. [978]

Pitt, M. K. (2002), "Smooth Particle Filters for Likelihood Evaluation and Maximization," technical report, University of Warwick, Dept. of Economics. [983]

Schwab, D., and Beletsky, D. (2002), "Hydrodynamic and Sediment Transport Modeling of Episodic Resuspension Events in Lake Michigan," in *Proceedings of the Seventh International Conference on Estuarine and Coastal Modeling*, ed. M. Spaulding, St. Petersburg, FL, pp. 266–279. [985]

Schwab, D. J., Bennett, J. R., Liu, P. C., and Donelan, M. A. (1984), "Application of a Simple Numerical Wave Prediction Model to Lake Erie," *Journal of Geophysical Research*, 89, 3586–3592. [985]

Shumway, R. H., and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications With R Examples* (2nd ed.), New York: Springer. [982,983]

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. (2008), "Obstacles to High-Dimensional Particle Filtering," *Monthly Weather Review*, 136, 4629–4640. [982]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer. [980]

——— (2007), "Seasonal Variation in the Spatial–Temporal Dependence of Total Column Ozone," *Environmetrics*, 18, 71–86. [978]

Stroud, J. R., and Bengtsson, T. (2007), "Sequential State and Variance Estimation Within the Ensemble Kalman Filter," *Monthly Weather Review*, 135, 3194–3208. [981]

Stroud, J. R., Lesht, B. M., Schwab, D. J., Beletsky, D., and Stein, M. L. (2009), "Assimilation of Satellite Images Into a Sediment Transport Model of Lake Michigan," *Water Resources Research*, 45, W02419. [985]

Verlaan, M. (1998), "Efficient Kalman Filtering Algorithms for Hydrodynamic Models," Ph.D. thesis, TU Delft, Delft, The Netherlands. [986]

Wikle, C. K. (2002), "A Kernel-Based Spectral Model for Non-Gaussian Spatial Processes," *Statistical Modelling: An International Journal*, 2, 299–314. [978]

——— (2003), "Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes," *Ecology*, 84, 1382–1394. [978,980]

Wikle, C. K., and Cressie, N. A. C. (1999), "A Dimension-Reduced Approach to Space–Time Kalman Filtering," *Biometrika*, 86, 815–829. [978]

Wikle, C. K., Berliner, L. M., and Cressie, N. A. C. (1998), "Hierarchical Bayesian Space–Time Models," *Environmental and Ecological Statistics*, 5, 117–154. [978]

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382–397. [978]

Wood, A. T. A., and Chan, G. (1994), "Simulation of Stationary Gaussian Processes in $[0, 1]^d$," *Journal of Computational and Graphical Statistics*, 3, 409–432. [982]

Xu, K., and Wikle, C. K. (2007), "Estimation of Parameterized Spatio-Temporal Dynamic Models," *Journal of Statistical Planning and Inference*, 137, 567–588. [978,980,981]

Zhang, Z., Beletsky, D., Schwab, D. J., and Stein, M. L. (2007), "Assimilation of Current Measurements Into a Circulation Model of Lake Michigan," *Water Resources Research*, 43, W11407. [988]