Ensemble Methods for Dynamic Data Assimilation of Chemical Observations in Atmospheric Models

Adrian Sandu, Emil Constantinescu, Gregory R. Carmichael, Tianfeng Chai, Dacian Daescu and John H. Seinfeld

reprinted from

Journal of
Algorithms & Computational Technology

Volume 5 · Number 4

December 2011

Multi-Science Publishing Co. Ltd

Ensemble Methods for Dynamic Data Assimilation of Chemical Observations in Atmospheric Models*

Adrian Sandu[1], Emil Constantinescu[1], Gregory R. Carmichael[2], Tianfeng Chai[2], Dacian Daescu[3], and John H. Seinfeld[4]

[1]Virginia Polytechnic Institute and State University, Department of Computer Science, Blacksburg, VA 24061.
E-mail: *sandu@cs.vt.edu*, emconsta@vt.edu
[2]The University of Iowa, Center for Global and Regional Environmental Research, 424 IATL, Iowa City, IA 52242.
E-mail: *gcarmich@engineering.uiowa.edu,* tchai@cgrer.uiowa.edu
[3]Portland State University, Department of Mathematics and Statistics, Portland, OR 97207. E-mail: *daescu@pdx.edu*
[4]California Institute of Technology, Department of Chemical Engineering, Pasadena, CA 91125. E-mail: Seinfeld@caltech.edu

Submitted: January 2010; Accepted: January 2011

ABSTRACT

The task of providing an optimal analysis of the state of the atmosphere requires the development of dynamic data-driven systems (DDDAS) that efficiently integrate the observational data and the models. Data assimilation, the dynamic incorporation of additional data into an executing application, is an essential DDDAS concept with wide applicability. In this paper we discuss practical aspects of nonlinear ensemble Kalman data assimilation applied to atmospheric chemical transport models. We highlight the challenges encountered in this approach such as filter divergence and spurious corrections, and propose solutions to overcome them, such as background covariance inflation and filter localization. The predictability is further improved by including model parameters in the

*This work was supported by the National Science Foundation through the award NSF ITR AP&IM 0205198 managed by Dr. Frederica Darema.

[1]Corresponding author. Adrian Sandu, Computer Science Department, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, Phone: 540-231-2193. Fax: 540-231-9218. Email address: *sandu@cs.vt.edu*

assimilation process. Results for a large scale simulation of air pollution in North-East United States illustrate the potential of nonlinear ensemble techniques to assimilate chemical observations.

Keywords: Dynamic data-driven systems, data assimilation, ensemble Kalman filter, chemical transport models

1. INTRODUCTION

The chemical composition of the atmosphere has been (and is being) significantly perturbed by emissions of trace gases and aerosols associated with a variety of anthropogenic activities. This changing of the chemical composition of the atmosphere has important implications for urban, regional and global air quality, and for climate change. In the US alone 474 counties with nearly 160 million inhabitants, are currently in some degree of non-attainment with respect to the 8-hour National Ambient Air Quality Standards (NAAQS) standard for ground-level ozone (80 ppbv). Because air quality problems relate to immediate human welfare, their study has traditionally been driven by the need for information to guide policy.

Chemical transport models (CTMs) have become an essential tool for providing science-based input into best alternatives for reducing urban pollution levels, for designing cost-effective emission control strategies, for the interpretation of observational data, and for assessments into how we have altered the chemistry of the global environment. The use of CTMs to produce air quality forecasts has become a new application area, providing important information to the public, decision makers and researchers. Currently hundreds of cities world-wide are providing real time air quality forecasts. In addition, the U.S. National Weather Service (NWS) has recently started to provide mesoscale numerical model forecast guidance for short-term air quality predictions, beginning with next-day ozone (O_3) forecasts for the northeastern, and plans to expand this air quality capability over the next ten years to include the entire U.S., to lengthen the forecast period to 3-days, and to add fine particulate matter ($PM_{2.5}$) to the forecasts. The use of CTMs in support of large field experiments is another important application [Lee et al., 1997; von Kuhlman et al. 2003].

Over the last decade our ability to measure atmospheric chemistry, transport and removal processes has advanced substantially. We are now able to measure at surface sites and on mobile platforms (such as vans, ships and aircraft), with

fast response times and wide dynamic range, many of the important primary and secondary atmospheric trace gases and aerosols (e.g., carbon monoxide, ozone, sulfur dioxide, black carbon, etc.), and many of the critical photochemical oxidizing agents (such as the OH and HO2 radicals). Not only is our ability to characterize a fixed atmospheric point in space and time expanding, but the spatial coverage is also expanding through growing capabilities to measure atmospheric constituents remotely using sensors mounted at the surface and on satellites.

While significant advances in CTMs have taken place, predicting air quality remains a challenging problem due to the complex processes occurring at widely different scales and by their strong coupling across scales. Figure 1 illustrates some of the complexities in air quality predictions. Models have been developed for the simulation of these processes at each scale (right). These models have to balance fidelity (i.e., the accuracy of the description of the physical and chemical processes) and computational cost. Very detailed zero-dimensional ("box") models incorporate high fidelity descriptions of the chemistry, aerosol and atmospheric dynamics, and thermodynamics. For larger areas, models incorporate more processes and employ more grid points; but for computational feasibility the spatial and temporal resolution is decreased, and the fidelity of each component is reduced.

Air quality predictions have large uncertainties associated with: incomplete and/or inaccurate emissions information; lack of key measurements to impose initial and boundary conditions; missing science elements; and poorly parameterized processes. Improvements in the analysis capabilities of CTMs require them to be better constrained through the use of observational data. The ability to dynamically incorporate additional data into an executing application is a fundamental DDDAS concept (*http://www.cise.nsf.gov/dddas*). We refer to this process as data assimilation. Borrowing lessons learned from the evolution of numerical weather prediction (NWP) models, improving air quality predictions through the assimilation of chemical data holds significant promise. The dynamic data feedback loops that relate chemical transport models and observations are presented in Figure 1.

In this paper we focus on the particular challenges that arise in the application of nonlinear ensemble filter data assimilation to atmospheric CTMs. This paper addresses the following issues: (1) Background covariance inflation is investigated in order to avoid filter divergence, (2) localization is used to prevent spurious filter corrections caused by small ensembles, and (3)

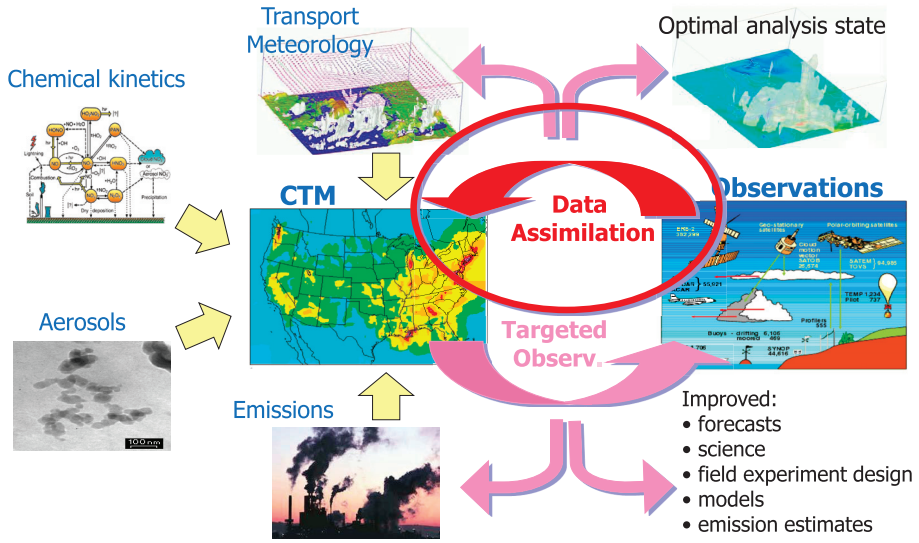

Figure 1. Dynamic data feedback loops between models and observations as they relate to predicting air quality. Complex CTMs incorporate chemical, aerosol, radiation modules, and use information from meteorological simulations (e.g., wind and temperature fields, turbulent diffusion parameterizations) and from emission inventories to produce chemical weather forecast. Yellow arrows represent the data flow for predictions using the first principles. Another source of information for concentrations of pollutants in the atmosphere is the observations. Data assimilation combines these two sources of information to produce an optimal analysis state of the atmosphere, consistent with both the physical/chemical laws of evolution through the model (first principles) and with reality through measurement information. Pink arrows illustrate the data flow for dynamic feedback and control loop from measurements/data assimilation to simulation. Targeted observations locate the observations in space and time such that the uncertainty in predictions is minimized.

parameters are assimilated together with the model states in order to reduce the model errors and improve the forecast. The paper is organized as follows. Section 2 presents the ensemble Kalman data assimilation technique, Section 3 illustrates the use of the tools in a data assimilation test, and Section 4 summarizes our results.

2. CHEMICAL TRANSPORT MODELING

An atmospheric CTM solves for the mass balance equations for concentrations y_i of tracer species $1 \le i \le n$.

$$\begin{aligned}
&\frac{\partial y_i}{\partial t} = -\vec{u}\cdot\nabla y_i + \frac{1}{\rho}\nabla(\rho K\cdot\nabla y_i) + \frac{1}{\rho} f_i(\rho y) + E_i\,, \quad 1 \le i \le n \\
&y_i(t^0, x) = y_i^0(x), \qquad\qquad t^0 \le t \le t^F \\
&y_i(t, x) = y_i^{IN}(t, x) \qquad \text{on} \quad \Gamma^{IN} \\
&K\frac{\partial y_i}{\partial n} = 0 \qquad\qquad \text{on} \quad \Gamma^{OUT} \\
&K\frac{\partial y_i}{\partial n} = V_i^{DEP} y_i - Q_i \quad \text{on} \quad \Gamma^{GROUND}
\end{aligned} \tag{1}$$

Here u represents the wind velocity vector, K is the turbulent diffusion tensor, and ρ is the air density. These variables are typically prescribed from simulations with a numerical weather prediction model. The concentrations y_i are expressed as a mole fraction (e.g., the number of molecules of tracer per 1 billion molecules of air); the absolute concentration of tracer i is ρy_i (molecules per cm^3). The rate of chemical transformations of species i is f_i, and depends on all other concentrations at the same spatial location. The elevated emissions of species i are E_i and the ground level emissions are Q_i. The deposition velocity is V_i^{DEP}. The model has prescribed initial conditions C^0 and is subject to Dirchlet boundary conditions at the inflow (lateral and top) boundary Γ^{IN}, to no diffusive flow condition at the outflow (lateral and top) boundary Γ^{OUT}, and to Neumann boundary conditions at the ground level boundary Γ^{GROUND}.

$$y^k = M(t^{k-1}, y^{k-1}, p)\,, \quad y^0 = y(t^0)\,, \quad k = 1, 2, \ldots \tag{2}$$

where y^k is the discrete state vector containing the dependent variables at time t^k (e.g., concentrations of chemical species), p is the vector of model parameters (e.g., the emission rates, deposition velocities, boundary fluxes), M and is the discrete model solution operator.

Air quality forecasts built upon CTM predictions (in contrast to other techniques such as statistical methods) contain components related to emissions, transport, transformation and removal processes. Since the 4-dimensional distribution of pollutants in the atmosphere is heavily influenced

by the prevailing meteorological conditions, air quality models are closely aligned with weather prediction. Air quality forecasting differs in important ways from the problem of weather forecasting. One important difference is that weather prediction is typically focused on severe, adverse weather conditions (e.g., storms), while the meteorology of adverse air quality conditions frequently is associated with benign weather. Air quality predictions also differ from weather forecasting due to the additional processes associated with emissions, chemical transformations, and removal. Because many important pollutants (e.g., ozone and fine particulate sulfate) are secondary in nature (i.e., formed via chemical reactions in the atmosphere), air quality models must include a rich description of the photochemical oxidant cycle. As a result of these processes air quality models typically include hundreds of chemical variables (including gas phase constituents and aerosol species distributed by composition and size). The resulting system of equations is stiff and highly coupled, which greatly adds to the computational burden of air quality forecasting. It is also important to note that the chemical and removal processes are highly coupled to meteorology variables (e.g., temperature and water vapor), as are many of the emission terms (directly in the case of wind blown soils whose emission rates correlate with surface winds and evaporative emissions that correlate with temperature, and indirectly in the case of those associated with heating and cooling demand that respond to ambient temperatures).

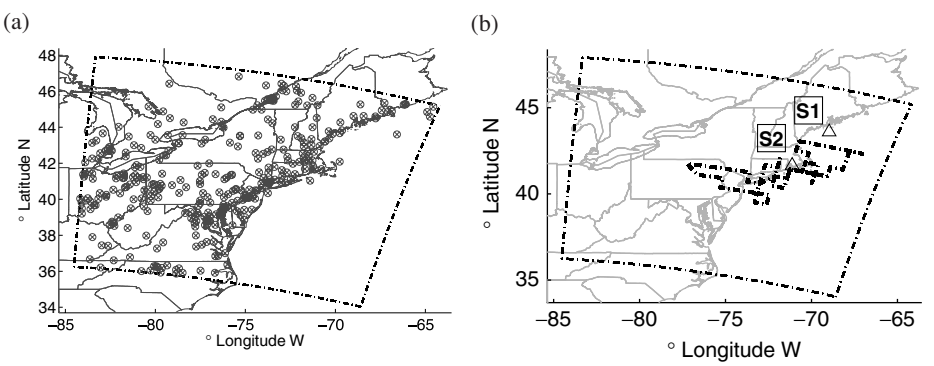

Figure 2. (a) Ground measuring stations in support of the ICARTT campaign (340 in total). (b) Location of two ozonesondes (S1, S2) and the flight path of a P3 plane.

3. DATA ASSIMILATION

Data assimilation is the process by which model predictions utilize measurements to obtain an optimal representation of the state of the atmosphere. The ability to dynamically incorporate additional data into an executing application is a fundamental DDDAS concept (*http://www.cise.nsf.gov/dddas*).

For the predictive capabilities of CTMs to improve, they must be better constrained through the use of observational data. The close integration of observational data is recognized as essential in weather/climate analysis, and it is accomplished by a mature experience/infrastructure in data assimilation—the process by which models use measurements to produce an optimal representation of the state of the atmosphere. This is equally desirable in CTMs.

Data assimilation combines information from three different sources: the physical and chemical laws of evolution (encapsulated in the model), the reality (as captured by the observations), and the current best estimate of the distribution of tracers in the atmosphere (all with associated errors). As more chemical observations in the troposphere are becoming available, chemical data assimilation is expected to play an essential role in air quality forecasting, similar to the role it has in numerical weather prediction.

Assimilation techniques fall within the general categories of variational (3D-Var, 4D-Var) and Kalman filter–based methods, which have been developed in the framework of optimal estimation theory. The variational data assimilation approach seeks to minimize a cost functional that measures the distance from measurements and the "background" estimate of the true state. In the 3D-VAR [Lorenc, 1986; Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987] method the observations are processed sequentially in time. The 4D-VAR [Courtier et al. 1994, Elbern et. al. 1999, 2000, Fisher and Lary 1995, Rabier et al. 2000] generalizes this method by considering observations that are distributed in time. These methods have been successfully applied in meteorology and oceanography [Navon 1998], but they are only just beginning to be used in nonlinear atmospheric chemical models [Menut et al., 2000, Elbern and Schmidt, 2001, Sandu et al. 2005]. When chemical transformations and interactions are considered, the complexity of the implementation and the computational cost of the data assimilation are highly increased. Some of the important challenges in chemical data assimilation include:

- Memory shortage (~100 concentrations of various species at each grid points, check-pointing required);
- Stiff differential equations (>200 various chemical reactions coupled

together, lifetimes of different species vary from seconds to months) ;
- Chemical observations are limited, compared to meteorological data;
- Emission inventories are often out-dated, and uncertainties are not well-quantified.

A discussion of current approaches follows.

3.1. Problem Formulation

Consider the chemical transport model (1) discretized in time and space (2). Observations of quantities that depend on system state are available at discrete times t^k

$$y_{obs}^k = h(y^k) + \varepsilon_{obs}^k \approx H_k y^k + \varepsilon_{obs}^k, \quad \langle \varepsilon_{obs}^k \rangle = 0, \quad \left\langle (\varepsilon_{obs}^k)(\varepsilon_{obs}^k)^T \right\rangle = R_k, \qquad (3)$$

where $y_{obs}^k \in \Re^m$ is the observation vector at t^k, h is the (model equivalent) observation operator and H_k is the linearization of h about the solution y^k. Each observation is corrupted by observational (measurement and representativeness) errors $\varepsilon_{obs}^k \in \Re^m$ [Cohn, 1997]. We denote by $\langle \cdot \rangle$ the ensemble average over the uncertainty space. The observational error is the experimental uncertainty associated with the measurements and is usually considered to have a Gaussian distribution with zero mean and a known covariance matrix R_k.

The aim of data assimilation is to find $P[y(t^k) | y_{obs}^k \ldots y_{obs}^0]$, the PDF of the true state at time t^k conditioned by all previous observations (including the most recent one). From Bayes' rule

$$P\left[y(t^k) \middle| y_{obs}^k \ldots y_{obs}^0 \right] = \frac{P[y_{obs}^k \middle| y(t^k)] \cdot P[y(t^k) \middle| y_{obs}^{k-1} \ldots y_{obs}^0]}{\int P[y_{obs}^k | y] \cdot P[y \middle| y_{obs}^{k-1} \ldots y_{obs}^0] \, dy} \qquad (4)$$

$P[y_{obs}^k | y(t^k)] = P(\varepsilon_{obs}^k)$ is the PDF of the latest observational error $P[y(t^k) | y_{obs}^{k-1} \ldots y_{obs}^0]$ is the "model forecast PDF" (conditioned by all previous observations minus the most recent one) and $P[y(t^k) | y_{obs}^k \ldots y_{obs}^0]$ is the "assimilated PDF".

In the 4D-Var approach an optimal solution is sought by adjusting chosen parameters according to available measurements in the analysis time interval. Such parameters are often called control variables and they may include initial concentrations, emission rates, concentration and flux at domain boundaries, and other physical or chemical parameters. The gradients of the cost functional

with respect to all control parameters are calculated simultaneously through the adjoint model. With the gradients, the optimal solution can be found efficiently by applying various minimization routines. Quasi-Newton limited memory L-BFGS [Byrd et al., 1995] is used by most 4D-Var applications. Chai, et al [2006] found that adding constraints to the admissible solution space through L-BFGS-B [Zhu et al, 1997] improved the optimization efficiency. Variational techniques for data assimilation are well-established in numerical weather prediction (NWP). Building on the early variational approach [Lorenc, 1986; Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987], the 4D-Var framework is the current state-of-the-art in meteorological [Courtier et al., 1994; Rabier et al., 2000] and chemical [Elbern et al., 2000a, 2001a; Liao et al., 2005; Sandu et al., 2003, 2005; Sandu, 2006; Segers, 2002] data assimilation. Lorenc [2003] performs a comparison of 4D-Var versus EnKF.

3.2. The Ensemble Kalman Filter

Kalman filters [Kalman, 1960] provide a stochastic approach to the data assimilation problem. The filtering theory is described in Jazwinski [1970] and the applications to atmospheric modeling in [Menard et al., 2000]. The computational burden associated with the filtering process has prevented the implementation of the full Kalman filter for large-scale models. Ensemble Kalman filters (EnKF) [Burgers et al,, 1998; Evensen, 2003] may be used to facilitate the practical implementation as shown by van Loon et al. [2000]. There are two major difficulties that arise in EnKF data assimilation applied to CTMs: (1) CTMs have stiff components [Sandu et al., 1997] that cause the filter to diverge [Houtekamer et al., 1998] due to the lack of ensemble spread and (2) the ensemble size is typically small in order to be computationally tractable and this leads to filter spurious corrections due to sampling errors. Kalman filter data assimilation has been discussed for DDDAS in another context by Jun and Bernstein [2006].

The ensemble Kalman filter (EnKF) approach to data assimilation has recently received considerable attention in meteorology. The Kalman filter [Kalman, 1960; Evensen, 1992; Evensen, 1993; Fisher, 2002] solves eqn (4) under the assumptions that the model is linear, and the model state at previous time t^{k-1} is normally distributed with mean y_a^{k-1} and covariance matrix P_a^{k-1}. The Extended Kalman Filter (EKF) allows for nonlinear models and observations by assuming the error propagation is linear (through the tangent linear model) and by linearizing the observation operators, $y_{obs}^k = H_k y^k + \varepsilon_{obs}^k$. However, the

(extended) Kalman Filter is impractical for large systems due to the high cost of propagating covariance matrices. A practical approach is provided by the ensemble Kalman Filter (EnKF) [Fisher, 2002; Evensen, 1994; Burgers, 1998] which estimates covariances through sampling the state space. Consider an ensemble of N states $\{y_a^{k-1}[i]\}_{1\le i\le N}$ at t^{k-1}. Each of the ensemble states is evolved in time using model equation to obtain a forecast ensemble at t^k,

$$y_f^k[i] = M(t^{k-1}, y_a^{k-1}[i]), \quad 1 \le i \le N \tag{5}$$

The mean and the covariance of the forecast PDF are approximated by the ensemble statistics:

$$\langle y_f^k \rangle = \frac{1}{N}\sum_{i=1}^{N} y_f^k[i], \quad (P_f^k)_{i,j} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_f^k[i] - \langle y_f^k \rangle\right)\left(y_f^k[i] - \langle y_f^k \rangle\right)^T \tag{6}$$

An ensemble of observation vectors $\{y_{obs}^k[i]\}_{1\le i\le N}$ is constructed by adding to the most recent observation vector y_{obs}^k perturbations drawn from a normal distribution with zero mean and covariance R_k. Each member of the ensemble is assimilated using the EKF to obtain the ensemble of analyzed states $\{y_a^k[i]\}_{1\le i\le N}$:

$$y_a^k[i] = y_f^k[i] + P_f^k H_k^T \left(R_k + H_k P_f^k H_k^T\right)^{-1}\left(y_{obs}^k[i] - H_k y_f^k[i]\right) \tag{7}$$

The ensemble mean and covariance describe the PDF of the assimilated field. The cost of updating the covariance matrix is that of N model evaluations. The ensemble implicitly describes a density function that can be non-Gaussian. Experience gained in numerical weather prediction indicates that relatively small ensembles (50–100 members) are sufficient to accurately capture this density function [Houtekamer, 1998]. Extensions of this approach proposed in the literature include the Ensemble Kalman Smoother [Evensen, 2000], the 4D-EnKF method [Hunt et al., 2003], the Ensemble Transform Kalman Filter [Bishop et al., 2000], the hybrid approach [Hansen et al., 2001] and ensemble nonlinear filters [Anderson et al., 1999; Anderson, 2001; Pham, 2001].

The application of EnKF presents several challenges: (1) the rank of estimated covariance matrix is (much) smaller than its dimension; a solution is presented [Houtekamer et al., 2001]; (2) the random errors in the statistically

estimated covariance decrease only by the square-root of the ensemble size; (3) the subspace spanned by random vectors for explaining forecast error is not optimal [Hershel et al., 2002]; and (4) the estimation and correct treatment of model errors is possible but difficult [Daley, 1992; Dee, 1995; Shubert et al., 1996; Houtekamer et al., 1997; Hansen, 2001; Babovic et al., 2002]. In addition, a careful implementation is required for efficiency [Houtekamer et al., 2001].

In spite of these challenges, EnKF has many attractive features including: (1) it is able to propagate the PDFs through highly nonlinear systems; (2) it does not require additional modeling efforts such as the construction of tangent linear model and its adjoint; and (3) the method is highly parallelizable.

3.3. The Role of Chemical Observations

As we have discussed throughout this paper, improved predictions require a closer integration of measurements with models. The weather forecast system is supported by a comprehensive observing system designed to improve forecasting skill. No such system exists to support air quality forecasts. The chemical observations presently available were designed largely for environmental compliance and not to enhance predictive skill. However that opens the question as to what chemical data is needed to improve the predictions? The chemical data assimilation techniques can be used to help address this issue.

4. ENSEMBLE-BASED CHEMICAL DATA ASSIMILATION

Our data assimilation numerical experiments use the state-of-the-art regional atmospheric photochemistry and transport model STEM (Sulfur Transport Eulerian Model) (Carmichael et al., 2003) to solve the mass-balance equations for concentrations of trace species in order to determine the fate of pollutants in the atmosphere [Sandu et al., 2005].

The test case is a real-life simulation of air pollution in North–Eastern U.S. in July 2004 as shown in Figure 3.a (the dash-dotted line delimits the domain). The observations used for data assimilation are the ground-level ozone (O3) measurements taken during the ICARTT [ICARTT; Tang et al., 2006] campaign in 2004 (which also includes the initial concentrations, meteorological fields, boundary values, and emission rates). Figure 3.a shows the location of the ground stations (340 in total) that measured ozone concentrations and an ozonesonde (not used in the assimilation process). The computational domain

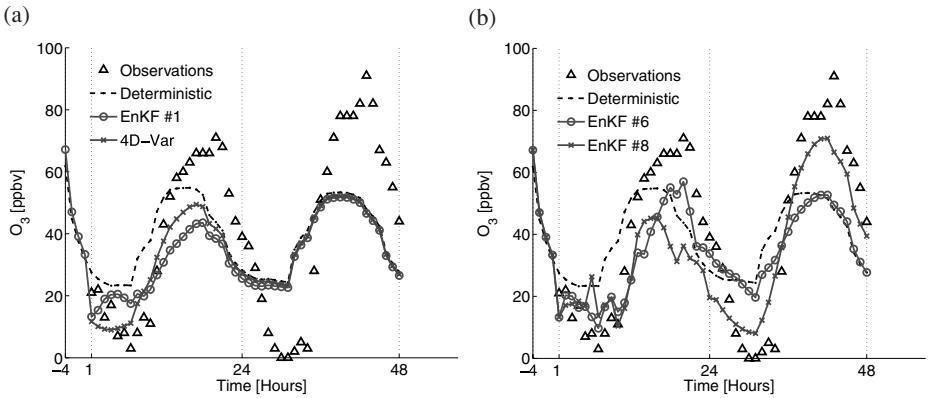

Figure 3. (a) Ozone concentrations measured at one selected AirNow station and predicted by EnKF#1 (50 members, "noiseless application") and 4D-Var (50 iterations). (b): Ozone concentrations measured at the selected station and predicted by EnKF #6 (multiplicative inflation) and EnKF#8 (parameter inflation with standard deviations of 10% for emissions, 10% for lateral boundary conditions, and 3% for the wind field).

covers $1500 \times 1320 \times 20$ Km with a horizontal resolution of 60×60 Km and a variable vertical resolution. The simulations are started at 0 GMT July 20th with a four hour initialization step ([−4,0] hours). The "best guess" of the state of the atmosphere at 0 GMT July 20th is used to initialize the deterministic solution. The ensemble members are formed by adding a set of unbiased perturbations to the best guess, and then evolving each member to 4 GMT July 20th. The perturbation is formed according to an AR model [3] making it flow de- pendent. The 24 hours assimilation window starts at 4 GMT July 20th (denoted by [1, 24] hours). Observations are available at each integer hour in this window, i.e., at 1, 2, . . ., 24 hours (Figure 1.a). EnKF adjusts the concentration fields of 66 "control" chemical species in each grid point of the domain every hour using (2). The ensemble size was chosen to be 50 members (a typical size in NWP). A 24 hour forecast window is also considered to start at 4 GMT July 21st (denoted by [24, 48] hours).

The performance of each data assimilation experiment is measured by the R^2 correlation and RMS factors between the observations and the model solution (separate R^2 and RMS factors are computed in the assimilation and in the forecast windows). The R^2 correlation and RMS factor of two series x and y of length n are

$$R^2(x,y)=\frac{\left(n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i\right)^2}{\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)\left(n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)},$$
$$RMS(x,y)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-y_i)^2}\ . \tag{8}$$

4.1. Models of the Background Errors

Our current knowledge of the state of the atmosphere (at the beginning of the simulation) is represented by the "background" field and its error. In practice, little is known about the background error [Fisher, 1995, 2003]; it is typically assumed to be Gaussian and with zero mean (the model is unbiased) and covariance B. In [Constantinescu et al., 2007; Sandu et al., 2005] we consider background errors modeled by autoregressive (AR) processes of the form

$$A\delta c^B=S\xi,\quad S=diag(\sigma_{i,j,k}),\quad \xi\in(N(0,1))^n \tag{9}$$

A is a correlation coefficient matrix, and σ represents the state covariances. The AR background accounts for spatial correlations, distance decay, and chemical lifetime. For more details on the construction and application of the AR background model the reader is referred to [Constantinescu et al., 2007].

4.2. Preventing Filter Divergence

The textbook application of EnKF [Evensen, 2003] (perfect model assumption) to our particular scenario leads to filter divergence: EnKF shows a decreasing ability to correct the ensemble state toward the observations at the end of the assimilation window. Filter divergence [Houtekamer and Mitchell, 1998; Hamill, 2004] is caused by progressive underestimation of the model error covariance magnitude during the integration; the filter becomes "too confident" in the model and "ignores" the observations in the analysis process. The cure is to artificially increase the covariance of the ensemble (effectively accounting for model errors) and therefore decrease the filter's confidence in the model results. In this section we investigate several ways to "inflate" the ensemble covariance in order to prevent filter divergence.

The additive inflation process [Corazza et al., 2002] consists of adding random noise to the model solution; the noise can be thought of as a representation of the unknown model error. The most intuitive way is to add noise to the forecast solution

$$y_i^f(e) \leftarrow M\left(y_{i-1}^a(e)\right) + \eta(e), \quad e = 1, \quad ,E, \quad \eta \in N(0,Q) \Rightarrow P^f \leftarrow P^f + Q. \quad (10)$$

but in principle the noise can also be added to the analysis.

The multiplicative approach to covariance inflation [Anderson, 2001] is to enlarge the spread of the ensemble about its mean by a scalar factor $\gamma > 1$. This can be applied to either the forecast or the analyzed ensembles:

$$\begin{aligned} y_i^f(e) &\leftarrow \langle y_i^f \rangle + \gamma_-\left(y_i^f(e) - \langle y_i^f \rangle\right), \quad e = 1, \quad ,E, \quad \Rightarrow \quad P^f \leftarrow \gamma_-^2 P^f \\ y_i^a(e) &\leftarrow \langle y_i^a \rangle + \gamma_+\left(y_i^a(e) - \langle y_i^a \rangle\right), \quad e = 1, \quad ,E, \quad \Rightarrow \quad P^a \leftarrow \gamma_+^2 P^a \end{aligned} \quad (11)$$

The choice for the inflation factors is based on Kalman filtering theory which requires that the ensemble and innovation spreads be of similar magnitude [Evensen, 2003]. At each assimilation cycle the inflation factor is chosen as

$$\gamma_- = \max\left(\frac{trace\left(\langle dd^T \rangle - R\right)}{trace\left(HP^f H^T\right)}, 1\right) \quad (12)$$

where $d = y_{obs} - H \cdot y$ is the vector of innovations for all observations, R is the observational covariance, and H is the observation operator.

A third approach for covariance inflation is through perturbations applied to key model parameters, and we refer to it as model-specific inflation. This approach focuses on sources of uncertainty that are specific to each model (for instance in CTMs: boundary conditions, emissions, and meteorological fields). Each ensemble member is run with different values of model parameters, drawn from a specific probability distribution.

The filter behavior for different settings is shown in Figure 4 below. The noiseless application filter diverges, while the parameter and the multiplicative inflation strategies alleviate the problem.

4.3. Covariance Localization

The practical Kalman filter implementation employs a small ensemble of Monte Carlo simulations in order to approximate the background covariance (P^f). In its initial formulation, EnKF may suffer from spurious correlations

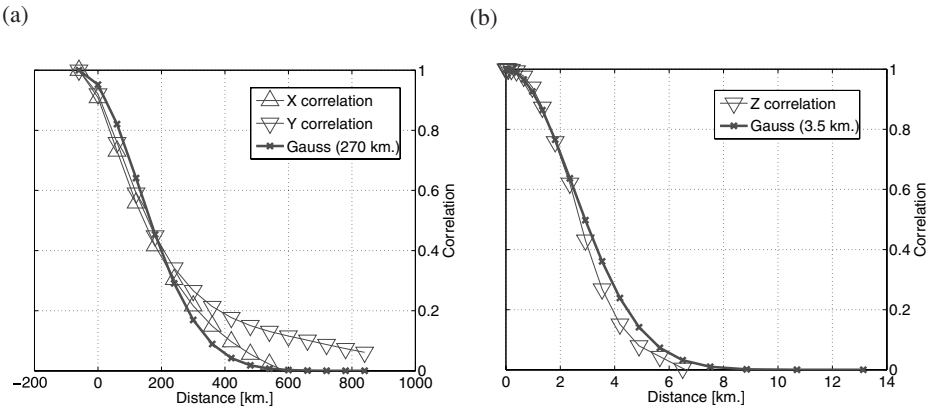

Figure 4. Horizontal (a) and vertical (b) correlation-distance relationships obtained from an ensemble of runs with the same verification time. The horizontal and vertical experimental correlation-distance curves are fitted with Gaussian functions with specific decorrelation distances (in parenthesis).

caused by sub-sampling errors in the background covariance estimates. This allows for observations to incorrectly impact remote model states. The filter localization introduces a restriction on the correction magnitude based on its remoteness. One way to impose localization in EnKF is to apply a decorrelation function $\rho(D)$ that decreases with the distance D, to the background covariance. Following [Houtekamer and Mitchell, 2001] the EnKF relation becomes

$$y_a^k[i] = y_f^k[i] + \rho(D^c) \circ P_f^k H_k^T \left(R_k + \rho(D^y) \circ \left(H_k P_f^k H_k^T \right) \right)^{-1} \left(y_{obs}^k[i] - H_k y_f^k[i] \right) \tag{13}$$

where the distance matrix D^y is calculated as the distance among the observation sites, and D^c contains the distance from each state variable to each observation site. The decorrelation function is applied to the distance matrix and produces a decorrelation matrix (decreasing with the distance). The operation '$\circ$' denotes the Schur product that applies. We consider a Gaussian decorrelation function that accounts for the anisotropy in the horizontal–vertical flow

$$\rho\left(D^h, D^v\right) = \exp\left(-\left(\frac{D^h}{L^h} \right)^2 - \left(\frac{D^v}{L^v} \right)^2 \right) \tag{14}$$

where D^h, D^v, are the horizontal and vertical distances and L^h, L^v, are the horizontal and vertical decorrelation lengths. The horizontal correlation-distance relationship is determined by fitting Gaussian distributions to the experimental ozone decorrelation distances obtained from multiple forecast verifying at the same time [Parrish and Derber, 1992] This is illustrated in the Figure below. The data fit gives $L^h = 270$ km and $L^v = 5$ grid points.

To exemplify the importance of covariance localization consider the vertical profiles of the assimilated ozone fields using the localized and non-localized versions of EnKF. Figure 5 represents the vertical profile of the ozone concentrations measured by the two ozonesondes (S1 and S2) together with the concentrations predicted by the model after the EnKF and LEnKF data assimilation with model-specific inflation. The ozonesondes were launched at 14 GMT (S1) and at 22 GMT (S2) July 20th. The EnKF solutions near the observation sites (on or close to the ground level) where the solution is constrained show a good fit, and the vertically developed correlations improve the solution in that vicinity. At higher altitudes, however, the ozonesondes show an oscillatory behavior of the ozone profile. LEnKF solution gives a fit as good as EnKF does close to the observation sites, and comes closer to the non-assimilated solution at higher altitudes, where there is no information about the true profile, and thus the model prediction prevails. The LEnKF approach forces the correction that each observation exerts on the concentration field to decrease with the distance from the observation site, and thus limits the spatial influence.

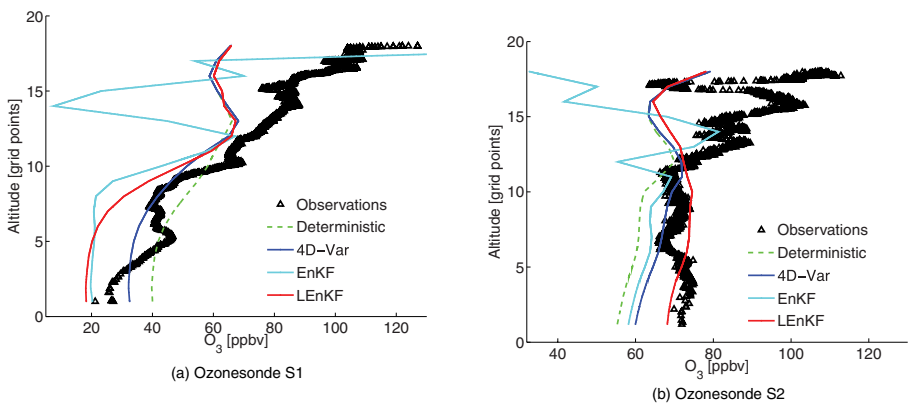

Figure 5. Ozone concentrations measured by the two ozonesondes and predicted by the model after data assimilation with 4D-Var, EnKF, and LEnKF. Model-specific inflation is used. The assimilated results without localization show considerable errors at high altitudes.

4.4. Inflation Localization

The traditional approach to covariance inflation increases the spread of the ensemble equally throughout the computational domain. In the LEnKF framework, the corrections are restricted to a region that is rich in observations. These states are corrected and their variance is reduced, while the remote states (i.e., the states that are relatively far from the observations' locations) maintain their initial variation which is potentially reduced only by the model evolution. The spread of the ensemble at the remote states may be increased to unreasonably large values through successive inflation steps. Therefore, the covariance inflation needs to be restricted in order to avoid the over-inflation of the remote states. A sensible inflation restriction can be based on the localization operator, $\rho(D)$, which is applied in the same way as for the covariance localization. The localized multiplicative inflation factor, γ_{loc}, is given by

$$\gamma_{loc}(i,j,k) = \rho(D^c(i,j,k)) \cdot (\gamma - 1) + 1 \tag{15}$$

where γ is the (non-localized) multiplicative inflation factor and i, j, k refer to the spatial coordinates. In this way, the localized inflation increases the ensemble spread only in the information-rich regions where filter divergence can occur.

4.5. Joint State-Parameter Data Assimilation

In regional CTMs the influence of the initial conditions is rapidly diminishing with time, and the concentration fields are "driven" by emissions and by lateral boundary conditions. Since both of them are generally poorly known, it is of considerable interest to improve their values using information from observations. In this setting we have to solve a joint state-parameter data assimilation problem. The emission rates and lateral boundary conditions are multiplied by specific correction coefficients, with a different coefficient for each species and each gridpoint. These correction coefficients are appended to the model state. The LEnKF data assimilation is then carried out with the augmented model state to recover corrected emissions and boundary conditions as well.

5. DATA ASSIMILATION RESULTS

We now illustrate the discussion with representative results. The data assimilation setting for Northeastern U.S. in July 2004 was discussed in Section

4. The behavior of the ensemble filter is shown in Figure 6, where the distribution of ground level ozone during the afternoon peak (2 pm) as predicted by the model before assimilation (Figure 7.a) and after assimilation (Figure 7.b) are plotted. The assimilated field more closely matches the observations (especially near the West inflow boundary) and displays finer scale structures.

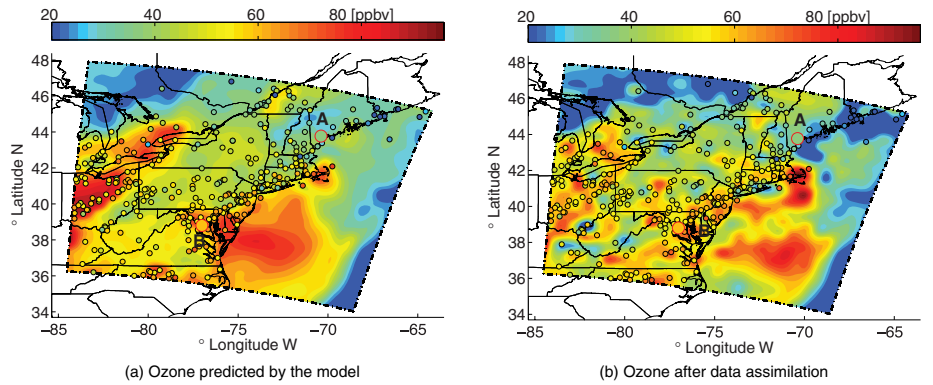

Figure 6. Ground-level ozone at 2 pm EDT on July 20, 2004 (in the assimilation window). Shown are the model predictions (a) without data assimilation, and (b) with data assimilation. The data is provided by the AirNow network (shown as circles colored by the measured ozone level).

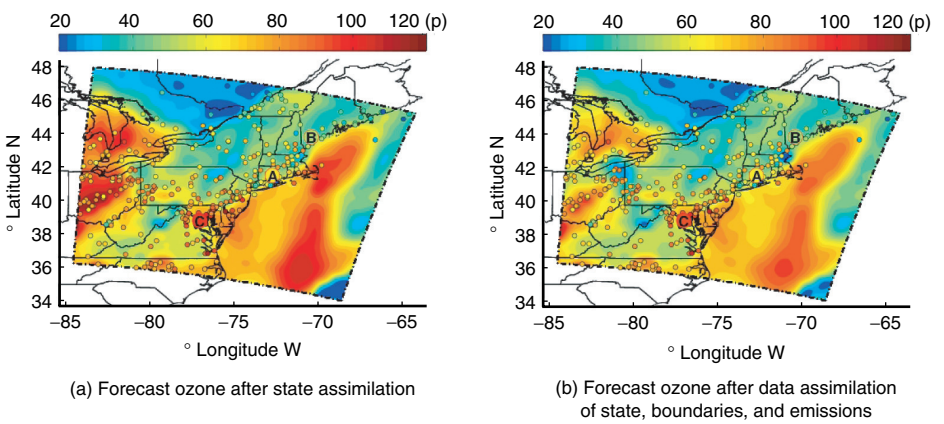

Figure 7. Ground-level ozone at 2 pm EDT on July 21, 2004 (in the forecast window). The forecast is based on two strategies. (a) Data assimilation corrects only the model state the state, and (b) data assimilation corrects for state, lateral boundary conditions, and emission rates. The forecast is in better agreement with the new observations when boundaries and emissions are also corrected.

Figure 8 illustrates the forecasted ground level ozone concentrations at 2 pm on the next day (July 21, 2004). In Figure 8 (a) he forecast uses the corrected initial conditions. In Figure 8(b) the forecast uses the corrected initial conditions, emissions, and boundary conditions. A comparison of the two plots with the AirNow observations (colored circles) reveals that the joint assimilation of state and parameters leads to an improved forecast.

The time evolution of ozone concentrations at selected ground stations (Figure 8) show how the assimilated ozone series follow the observations much closer than the non-assimilated ones in the analysis window.

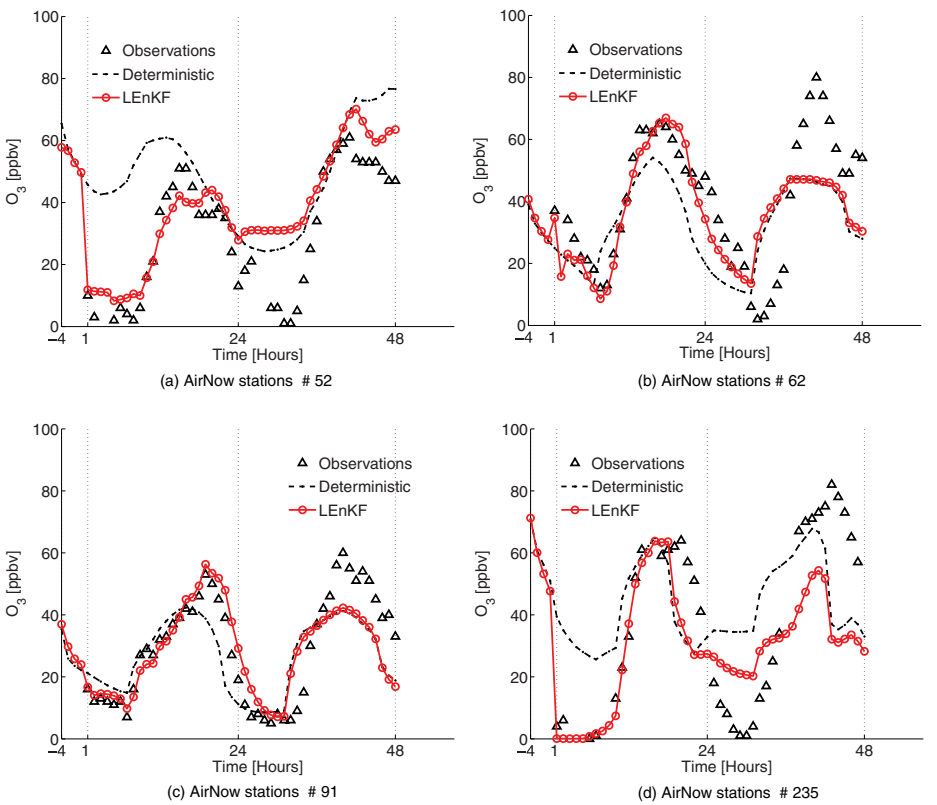

Figure 8. The time evolution of ozone at four selected stations. Data assimilation of the state is performed with the localized, inflated EnKF.

Table 1 contains performance results for data assimilation with different filter choices. The performance of each data assimilation experiment is measured by the R^2 correlation factor as well as the RMS distance between the model prediction and observations. The correlation factor between the observations and the model solution in the assimilation window is $R^2 = 0.24$ for the non-assimilated solution, $R^2 = 0.52$ for 4D-Var (results not shown), and $R^2 \approx 0.8 - 0.9$ for EnKF (with various forms of covariance inflation and localization). We note that the performance of the 200 member ensemble is better than the performance of the 50 members ensemble. However, with localization the 50 member ensemble results are very good. This number of members is to be preferred to the high computational overhead associated with large ensembles.

The impact of data assimilation on the forecast skill is also shown. The period from 24–48 hours represents the forecast. Near surface ozone levels are strongly dependent on chemical production/destruction processes involving a variety of precursor species. The joint assimilation of state, lateral boundary conditions, and emissions leads to considerable improvements not only in the assimilation window, but also in the forecast window.

6. CONCLUSIONS

This paper discusses some of the challenges associated with the application of nonlinear ensemble filtering data assimilation to atmospheric CTMs. Several aspects are analyzed in this study: (1) ensemble initialization – using autoregressive models of the background errors; (2) filter divergence - CTMs tend to dampen perturbations; (3) spurious corrections - small ensemble size cause wrong increments; (4) over-estimation of the model errors in data sparse areas; and (5) model parameterization errors - without correcting model errors in the analysis, correcting the state only does not help in improving the forecast accuracy.

Experiments showed that the filter diverges quickly. The influence of the initial conditions fades in time as the fields are largely determined by emissions and by lateral boundary conditions. Consequently, the initial spread of the ensemble is diminished in time. Moreover, stiff systems (like chemistry) are stable - small perturbations are damped out quickly in time. In order to prevent filter divergence, the spread of the ensemble needs to be explicitly increased. We discuss three approaches to ensemble covariance inflation among which model- specific inflation is the most intuitive. The “localization” of EnKF is needed in order to avoid the spurious corrections noticed in the

Table 1. The R^2 and RMS [ppbv] measures of model-observations match in the assimilation and forecast windows for the EnKF (with different ensemble sizes) and 4D-Var data assimilation.

Simulation and data assimilation method	**R^2 (RMS) analysis**	**R^2 (RMS) forecast**
Best guess solution, no assimilation	0.24 (22.1)	0.28 (23.5)
4D-Var 50 iterations w/AR background	0.52 (16.0)	0.29 (22.4)
EnKF (50 members) "noiseless application"	0.38 (18.2)	0.30 (23.1)
EnKF (200 members) "noiseless application"	0.49 (16.3)	0.30 (23.7)
EnKF (50 members) adaptive multiplicative inflation	0.67 (12.7)	0.19 (62.0)
EnKF (200 members) adaptive multiplicative inflation	0.82 (9.36)	0.28 (37.6)
LEnKF (50 members), "noiseless application"	0.81 (9.79)	0.34 (22.0)
LEnKF (50 members) adaptive multiplicative inflation	0.82 (9.52)	0.34 (22.0)
LEnKF (50 members), "noiseless". Joint assimilation of state, emissions, and lateral boundary conditions	0.88 (7.75)	0.42 (20.3)
LEnKF (50 members) adaptive multiplicative inflation. Joint assimilation of state, emissions, and lateral boundary conditions	0.91 (6.52)	0.40 (20.5)

"textbook" application. The correlation distances are approximated using experimental correlations. Furthermore, covariance localization prevents over-inflation of the states that are remote from observation. LEnKF increased both the accuracy of the analysis and forecast at the observation

sites and at distant locations (from the observations). A localization was is also applied to the ensemble inflation to prevent overestimation of model errors in data-sparse areas. Since the solution of a regional CTM is largely influenced by uncertain lateral boundary conditions and by uncertain emissions it is of great importance to adjust these parameters through data assimilation. The assimilation of emissions and boundary conditions visibly improves the quality of the analysis.

More work is required to completely understand the use of ensemble data assimilation to reduce uncertainties in emission inventories and in boundary conditions. One challenge arises from the long integration times needed to develop meaningful correlations between the emission rates or boundary conditions and the concentration fields. Another challenge is posed by large spurious correlations which lead the filter to correct the emission rates and boundary conditions in order to compensate for other sources of error.

In this paper we considered the "perturbed observations" version of EnKF. The performance of the "square root" EnKF variants will need to be assessed in the context of chemical data assimilation. In the future we plan to develop hybrid methods that combine the advantages of the 4D-Var and EnKF data assimilation approaches.

REFERNCES

1. J.L. Anderson, and S.L. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly Weather Review* 127, 2741–2758, 1999.
2. J.L. Anderson, An Ensemble Adjustment Kalman Filter for Data Assimilation., *Monthly Weather Review* 129: 2884–2903, 2001.
3. T.V. Babovic, and D.R. Fuhrman, Data assimilation of local model error forecasts in a deterministic model, *International journal for numerical methods in fluids* 39: 887–918, 2002.
4. C.H. Bishop, and B.J. Etherton, Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, *Monthly Weather Review,* 2000.
5. G. Burgers, and P.J. van Leeuwen, Analysis scheme in the ensemble Kalman Filter, *Monthly Weather Review* 126: 1719–1724, 1998.
6. R. Byrd, P. Lu, and J. Nocedal, A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Stat. Comput.* 16(5), (1995), 1190–1208.
7. G.R. Carmichael, et. al. Regional-scale Chemical Transport Modeling in Support of the Analysis of Observations obtained During the Trace-P Experiment. *J. Geophys. Res.*, 108(D21 8823), 10649–10671. 2003.

8. T. Chai, G. R. Carmichael, A. Sandu, Y. Tang, and D. N. Daescu, Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements *J. Geophys. Res.* 111, (2006), Art. No. D02301, doi:10.1029/2005 JD005883.
9. S.E. Cohn, An introduction to estimation theory, *J. Meteor. Soc. Japan* 75(B): 257–288, 1997.
10. E.M. Constantinescu, A. Sandu, T. Chai, and G.R. Carmichael. Ensemble-based chemical data assimilation. I: General Approach. Quarterly Journal of the Royal Meteorological Society Volume 133, Issue 626, Pages 1229–1243, July 2007 Part A.
11. E.M. Constantinescu, A. Sandu, T. Chai, and G.R. Carmichael. Ensemble-based chemical data assimilation. II: Covariance Localization. Quarterly Journal of the Royal Meteorological Society Volume 133, Issue 626, Pages 1245–1256, July 2007 Part A.
12. E.M. Constantinescu, A. Sandu, T. Chai, and G.R. Carmichael. Autoregressive models of background errors for chemical data assimilation. Journal of Geophysical Research, Vol. 112, D12309, doi:10.1029/2006JD008103, 2007.
13. E.M. Constantinescu and A. Sandu: "On Adaptive Mesh Refinement for Atmospheric Pollution Models". V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3515, pp. 798–805, 2005.
14. Corazza, E. Kalnay, and D. Patil. Use of the breeding technique to estimate the shape of the analysis "errors of the day". *Nonlinear Processes in Geophysics*, 10: 233–243, 2002.
15. Courtier, J. Thepaut, and A. Hollingsworth, A strategy of operational implementation of 4D-Var using an incremental approach. Q.J.R. Meteorol. Soc., 120, p. 1367–1388, 1994.
16. D.N. Daescu, and G.R. Carmichael, An adjoint sensitivity method for the adaptive location of the observations in air quality modeling. J. Atmos. Sci., Vol. 60, No. 2, 434–450, 2003.
17. Daley, Estimating model-error covariances for application to atmospheric data assimilation, *Monthly Weather Review* 120: 1735–1746, 1992.
18. V. Damian, A. Sandu, M. Damian, F. Potra, and G.R. Carmichael: "The kinetic preprocessor KPP -a software environment for solving chemical kinetics". Computers and Chemical Engineering, 26, p. 1567–1579, 2002.
19. D.P. Dee, On-line estimating model-error covariances for application to atmospheric data assimilation, *Monthly Weather Review* 120: 164–177, 1995.
20. J.C. Derber, and F. Bouttier, A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus 51A, (1999) 195–221.

21. H. Elbern and H. Schmidt. Ozone episode analysis by 4D-Var chemistry data assimilation, *J of Geophys Res*, 106(D4): 3569–3590, 2001.
22. H. Elbern, H. Schmidt, and A. Ebel. Implementation of a parallel 4D-Var chemistry data assimilation scheme, *Environmental Management and Health*, 10: 236–244, 1999.
23. H. Elbern, H. Schmidt, O. Talagrand, and A. Ebel. 4D-variational data assimilation with an adjoint air quality model for emission analysis. *Environmental Modeling and Software*, 15: 539–548, 2000.
24. G. Evensen, Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model, *J. Geophys. Res.* 97(C11): 17905–17924, 1992.
25. G. Evensen, Open boundary conditions for the extended Kalman filter with a quasi-geostrophic mode, *J. Geophys. Res.* 98(C19): 16529–16546, 1993.
26. G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.* 99(C5): 10143–10162, 1994.
27. G. Evensen, and P.J. van Leeuwen, An Ensemble Kalman Smoother for Nonlinear Dynamics, *Monthly Weather Review* 128: 1852–1867, 2000.
28. G. Evensen. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 2003.
29. M. Fisher and D.J. Lary. Lagrangian four-dimensional variational data assimilation of chemical species. Quarterly Journal of the Royal Meteorological Society, 121: 1681–1704, 1995.
30. M. Fisher, Assimilation Techniques (5): Approximate Kalman filters and singular vectors, 2002.
31. M. Fisher, Background error covariance modeling. Proceedings of the ECMWF Workshop on Recent Developments in Data Assimilation for Atmosphere and Ocean, September 8-12, 2003, Reading, UK, 45–64.
32. T.M. Hamill. Ensemble-based atmospheric data assimilation. *Technical report, University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, 2004.*
33. J.A. Hansen, Accounting for model error in ensemble-based state estimation and forecasting. *Monthly Weather Review*, 130: 2373–2391, 2001.
34. L.M. Herschel, and P.L. Houtekamer, Ensemble size, Balance, and Model-Error Representation in EnKF, *Monthly Weather Review* 125: 2416–2426, 2002.
35. P.L. Houtekamer, and L. Leafaivre, Using Ensemble Forecast for Model Validation, *Monthly Weather Review* 126: 2416–2426, 1997.
36. P.L. Houtekamer, and H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review* 126: 796–811, 1998.

37. P.L. Houtekamer, and H.L. Mitchell, A sequential Ensemble Kalman Filter for atmospheric data assimilation, *Monthly Weather Review* 129: 123–137, 2001.
38. B.R. Hunt, and E. Kalnay, Four-Dimensional Ensemble Kalman Filtering, *Tellus*. 2003.
39. ICARTT. ICARTT home page:http://www.al.noaa.gov/ICARTT.
40. A.H. Jazwinski. Stochastic Processes and Filtering Theory. Academic Press, 1970.
41. B.-E. Jun and D.S. Bernstein Least-correlation estimates for errors-in-variables models . Int'l J. Adaptive Control and Signal Processing, 20(7): 337–351, 2006.
42. R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Transaction of the ASME- Journal of Basic Engineering* 35–45, 1960.
43. R. von Kuhlmann, M.G. Lawrence, P.J. Crutzen, and P.J. Rasch, A model for studies of tropospheric ozone and nonmethane hydrocarbons: Model description and ozone results, J. Geophys. Res., 108(D9), doi:10.1029/2002JD002893, 2003.
44. F.X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations. *Tellus*, 38 A, 97–110, 1986.
45. A.C. Lorenc, Analysis methods for numerical weather prediction. *J of the Royal Meteorological Society, Soc.*, 112, 1177–1194, 1986.
46. A.C. Lorenc, The potential of the ensemble Kalman filter for NWP – a comparison with *4D-Var. Quarterly Journal of the Royal Meteorological Society,* 129(595): 3183–3203, 2003.
47. R. Menard, S.E. Cohn, L.-P. Chang, and P.M. Lyster. Stratospheric assimilation of chemical tracer observations using a Kalman filter. Part I: Formulation. Mon. Wea. Rev, 128: 2654–2671, 2000.
48. L. Menut, R. Vautard, M. Beekmann, and C Honor, Sensitivity of photochemical pollution using the adjoint of a simplifed chemistry-transport model, *J of Geophys Res - Atmospheres,* 105-D12(15): 15, 379 {15, 402, 2000.
49. M. Navon, Practical and Theoretical Aspects of Adjoint Parameter Estimation and Identifiability in Meteorology and Oceanography, *Dynamics of Atmospheres and Oceans. Special Issue in honor of Richard Pfeffer*, 27, Nos.1-4, 55–79 (1998).
50. D.F. Parrish, and J.C. Derber, National Meteorological Center's spectral statistical-interpolation analysis system. Mon. Wea. Rev. 120, (1992), 1747–1763.
51. D.T. Pham, Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Monthly Weather Review* 129: 1194–1207, 2001.
52. F. Rabier, H. Jarvinen, E. Klinker, J.F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four-dimensional variational assimilation. I:Experimental results with simplified physics. Quarterly Journal of the Royal Meteorological Society, 126: 1148–1170, 2000.

53. Sandu, E.M. Constantinescu, W. Liao, G.R. Carmichael, T. Chai, J.H. Seinfeld, and D. Daescu, Ensemble Filter Data Assimilation for Atmospheric Chemical and Transport Models. Pages 648–656. International Conference on Computational Science, 2005.
54. Sandu, D. Daescu, and G.R. Carmichael, Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: I – Theory and Software Tools, *Atmospheric Environment*, Vol. 37, p. 5083–5096, 2003.
55. Sandu, On the properties of Runge-Kutta discrete adjoints, International Conference on Computational Science, Reading, U.K., pages 550–557, 2006.
56. Sandu and R. Sander, Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmospheric Chemistry and Physics*, Vol. 6, pp 187–195, 26-1-2006.
57. Sandu, J.G. Verwer, M. Van Loon, G.R. Carmichael, F.A. Potra, D. Dabdub, and J.H. Seinfeld, Benchmarking stiff ODE solvers for atmospheric chemistry problems II – Rosenbrock solvers, *Atmospheric Environment* Vol. 31, No. 19, pp. 3151–3166, 1997.
58. Sandu, W. Liao, G.R. Carmichael, D.K. Henze, and J.H. Seinfeld, Inverse Modeling of Aerosol Dynamics using Adjoints - Theoretical and Numerical Considerations, *Aerosol Science and Technology*, Vol. 39, p. 1–18, DOI:10.1080/02786820500182289, 2005.
59. Sandu, D. Daescu, G.R. Carmichael, and T. Chai, Adjoint sensitivity analysis of regional air quality models, *J of Computational Physics*, 204, p. 222–252.
60. Sandu, Targeted Observations for Atmospheric Chemistry and Transport Models, V.N. Alexandrov et al. (Eds.): ICCS 2006, Part I, LNCS 3991, pp. 712–719, 2006.
61. A.J. Segers. Data assimilation in atmospheric chemistry models using Kalman filtering. *Ph.D. Thesis, TU Delft., 2002.*
62. S. Schubert, Y.H. Chang, An objective method for inferring sources of model error, *Monthly Weather Review* 124: 325–340, 1996.
63. O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory, *Quarterly Journal of the Royal Meteorological Society*, 113, 1311–1328, 1987.
64. Y. Tang, G.R. Carmichael, I. Uno, J.-H. Woo, G. Kurata, B. Lefer, R.E. Shetter, H. Huang, B.E. Anderson, M.A. Avery, A.D. Clarke, and D.R. Blake, Impacts of aerosols and clouds on photolysis frequencies and photochemistry during TRACE-P: 2. Three-dimensional study using a regional chemical transport model, J. Geophys. Res., 108(D21), doi:10.1029/2002JD003100, 2003.
65. Zhu, R. H. Byrd, and J. Nocedal, L-BFGS-B. Fortran routines for large scale bound constrained optimization. *ACM Trans. Math. Software* 23(4), (1997), 550–560.