Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 3082–3094
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Ensemble-based data assimilation and targeted observation of a chemical tracer in a sea breeze model

Amy L. Stuart[a,*], Altug Aksoy[b], Fuqing Zhang[c], John W. Nielsen-Gammon[c]

[a]*Departments of Environmental and Occupational Health and Civil and Environmental Engineering, University of South Florida, 13201 Bruce B. Downs Boulevard, MDC-56, Tampa, FL 33612-3805, USA*
[b]*National Center for Atmospheric Research, Boulder, CO, USA*
[c]*Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA*

Received 5 June 2006; received in revised form 25 September 2006; accepted 28 November 2006

Abstract

We study the use of ensemble-based Kalman filtering of chemical observations for constraining forecast uncertainties and for selecting targeted observations. Using a coupled model of two-dimensional sea breeze dynamics and chemical tracer transport, we perform three numerical experiments. First, we investigate the chemical tracer forecast uncertainties associated with meteorological initial condition and forcing error. We find that the ensemble variance and error builds during the transition between land and sea breeze phases of the circulation. Second, we investigate the effects on the forecast variance and error of assimilating tracer concentration observations extracted from a truth simulation for a network of surface locations. We find that assimilation reduces the variance and error in both the observed variable (chemical tracer concentrations) and unobserved meteorological variables (vorticity and buoyancy). Finally, we investigate the potential value to the forecast of targeted observations. We calculate an observation impact factor that maximizes the total decrease in model uncertainty summed over all state variables. We find that locations of optimal targeted observations remain similar before and after assimilation of regular network observations.

Keywords: Air quality modeling; Data assimilation; Ensemble modeling; Adaptive observations

1. Introduction

Significant strides in understanding and mitigating air pollution have been achieved over the past several decades (Seinfeld, 2004). Despite this progress, modeling and prediction of air pollution and its effects remains difficult and is prone to significant uncertainties (Seigneur, 2005). The complexity comes in large part from three areas. First, air pollution involves numerous nonlinear physical and chemical processes, many of which are not well understood. Second, the characteristic spatial and time scales of these processes span many orders of magnitude. Due to poor understanding of some processes and the computational cost of resolving all processes at the appropriate scale, many processes are parameterized in models. These parameterizations are, by nature, only approximations,

*Corresponding author. Tel.: +1 813 974 6632; fax: +1 813 974 4986.
E-mail address: astuart@hsc.usf.edu (A.L. Stuart).

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2006.11.046

and hence have errors and uncertainties associated with them. Finally, uncertainties in the initial and boundary conditions lead to uncertain predictions of air quality. The chaotic nature of the meteorological system governing underlying atmospheric transport (Lorenz, 1963) can often cause even small uncertainties to grow substantially in time.

Due to the uncertain nature of the prediction of air pollution hazards (whether due to routine emissions or emergency releases), simulation techniques combining deterministic modeling with statistical methods (i.e. Monte Carlo methods) have sometimes been used to determine air pollution hazard probabilities and to explore the sensitivity of predictions to underlying input and model uncertainties (e.g. Boybeyi et al., 1995; Stuart et al., 1996; Bergin et al., 1999; Dabberdt and Miller, 2000; Sax and Isakov, 2003; Zhang et al., 2006). One such statistical technique that has significant promise for improving air quality modeling and prediction is four-dimensional data assimilation (FDDA). FDDA combines observational data, knowledge of the physical and chemical system behavior (as represented in predictive models), and information on the uncertainty in both the observed data and the model representation over space and time (Kalnay, 2003). By integrating this information, data assimilation can provide a more accurate description of the system state (i.e., multidimensional information on wind speed, wind direction, temperature, and chemical concentration) and its expected evolution in time. Data assimilation can also be used to constrain errors associated with the uncertain model parameterizations and to determine parameter values, through on-line optimization of the parameters used (''parameter estimation'') (e.g., Navon, 1998).

In addition to improving model predictions, data assimilation research has led to the development of tools such as adjoint models for the selection and design of observational networks and targeted observations, as it provides a mechanism for on-line optimization of this selection (Morss et al., 2001). Existing long-term air quality observational networks are generally fixed in space and have very sparse spatial and temporal resolution. The sparse resolution is due to the number of variables that need to be observed (several meteorological variables and up to hundreds of toxic chemical pollutants) and large equipment and operational expenses. In addition to long-term fixed locations, targeted observational locations are often used for short-term monitoring, for specific management and research purposes. Recently, interest has escalated in observation targeting for response to air pollution hazards that result from emergency releases. In the weather forecasting field, there has been significant research on adaptive location of targeted observations using techniques originating with data assimilation (Emanuel et al., 1995; Emanuel and Langland, 1998; Lorenz and Emanuel, 1998; Berliner et al., 1999) that could be applied to air quality network design.

Although data assimilation is used operationally for meteorological modeling and prediction, its use for air quality modeling is less developed. Nonetheless, it has been found to be useful as an inverse modeling technique for diagnosing pollutant emission source locations and strengths (i.e. parameter estimation) (Chang et al., 1997; Elbern et al., 2000; Mendoza-Dominguez and Russell, 2001) and for identifying locations (in time and space) for field observation networks and adaptive observations (Daescu and Carmichael, 2003). Much of this previous work has focused on variational data assimilation techniques (three-dimensional and four-dimensional variational assimilation, 3DVAR and 4DVAR).

Ensemble-based Kalman filtering is an alternative data assimilation approach that is undergoing significant investigation for many environmental modeling applications (Evensen, 2003). Advantages include lower computational costs than extended Kalman filtering and explicit calculation of the nonlinear evolution of background covariances through the ensemble forecast. Explicit calculation of covariance eliminates the assumption of stationarity of covariances used in 3DVAR, and alleviates the need for the development of tangent linear and adjoint models of the dynamics used in extended Kalman filtering and 4DVAR (Tippett et al., 2003; Kalnay, 2003). Heemink and collaborators have investigated ensemble-based Kalman filtering and other Kalman filtering techniques for prediction of ozone concentrations over Europe with two- and three-dimensional chemical transport models that are decoupled (calculated off-line) from the underlying meteorological dynamics simulation (Hanea et al., 2004; Heemink and Segers, 2002). Their work indicates improvement in model predictions with filtering, and demonstrates the feasibility of Kalman filtering techniques for modeling air quality.

Here, we study the utility of ensemble-based Kalman filtering in the context of a two-dimensional

sea breeze model in which chemical tracer transport is directly coupled to the underlying nonlinear meteorological dynamics. The sea breeze circulation is an important weather pattern affecting coastal areas. Since many cities are located near coasts, these circulations have important impacts on the formation and transport of urban air pollution. In the United States, two of the most polluted cities, Houston and Los Angeles, are located on coastlines for which sea breeze circulations affect air quality. For example, observational and modeling studies (Banta et al., 2005; Bao et al., 2005; Zhang et al., 2006) have found that the sea breeze circulation was a significant contributing factor to high ozone events in Houston during the Texas 2000 Air Quality Study, as polluted air was recirculated over emissions sources.

Section 2 of this manuscript describes the sea breeze model, the chemical tracer algorithm, and the ensemble-based Kalman filter technique used. Section 3 describes and discusses our numerical experiments. Section 4 provides conclusions and implications.

2. Model description

2.1. The sea breeze model

A detailed description of the sea breeze model can be found in Aksoy et al. (2005). For the purposes of this paper, we note that the meteorological model equations are two-dimensional, nonlinear, hydrostatic, non-rotating, and incompressible. Perturbation buoyancy (b) and vorticity (η) are the prognostic variables, as described by the following equations:

$$\frac{\partial \eta}{\partial t} + u\frac{\partial \eta}{\partial x} + w\frac{\partial \eta}{\partial z} + \frac{\partial b}{\partial x} = k_\eta \frac{\partial^2 \eta}{\partial z^2} \tag{1}$$

and

$$\frac{\partial b}{\partial t} + u\frac{\partial b}{\partial x} + w\frac{\partial b}{\partial z} + N^2 w = k_b \frac{\partial^2 b}{\partial z^2} + Q, \tag{2}$$

where k_η and k_b are the vertical dispersion coefficients, Q is a buoyancy source term, N is the Brunt–Vaisala frequency, x and z are the horizontal (coastline normal) and vertical domain variables, and t is time. The horizontal and vertical wind speeds, u and w, are diagnosed from the vorticity through the stream function, ψ, with the following definitional equations:

$$\eta = \frac{\partial^2 \psi}{\partial z^2}, \quad u = \frac{\partial \psi}{\partial z}, \text{ and } w = -\frac{\partial \psi}{\partial x}. \tag{3}$$

The model spatial domain represents the cross-shore horizontal distance and the vertical above-ground altitude. The coast is located at horizontal domain center. The force that drives the sea breeze circulation is modeled as an explicit volume buoyancy source that represents the differential heating over land and sea. The source function varies horizontally as an arc tangent with the inflection point at the coast-line, decays exponentially with vertical distance from the ground, and varies sinusoidally in time but with added stochastic noise. Free slip and thermal insulation are assumed at the vertical domain boundaries, and zero flux is assumed at the horizontal boundaries. The forecast domain size is 500 km horizontally and 3 km vertically. Rayleigh-damping sponge layers are used beyond the sides and top of the forecast domain, to lessen the effects of the boundaries on the forecast domain. The horizontal and vertical sponge layers are 300 and 2 km thick, respectively. Grid spacing is 4 km horizontally and 50 m vertically. Hence, the model resolves the mesoscale features of the sea breeze circulation and marginally resolves the nonlinear frontal structure associated with the sea breeze.

2.2. The chemical tracer model

To simulate tracer concentrations, the following tracer transport equation has been coupled with the sea breeze meteorological dynamics model:

$$\frac{\partial A}{\partial t} + u\frac{\partial A}{\partial x} + w\frac{\partial A}{\partial z} = k_A \frac{\partial^2 A}{\partial z^2} + S, \tag{4}$$

where A is the tracer concentration, k_A is the vertical dispersion coefficient, and S is a source term. The numerical solution is accomplished with the scheme consistent with that used for meteorological dynamics (as described in Aksoy et al., 2005, 2006). All spatial derivatives are discretized using second order central differences. Dispersion terms are discretized in time with the trapezoidal method (resulting in an overall Crank–Nicolson scheme) and advection terms are discretized in time with the leap frog method. Lagged-in-time numerical horizontal dispersion and an Asselin time filter (Asselin, 1972) are used to increase the stability of the simulations. Boundary conditions for the tracer concentrations are implemented to ensure no diffusive or advective flux at all boundaries. Source terms can include externally set grid-cell-average tracer sources and lower boundary flux sources.

2.3. The ensemble-based Kalman filtering method

The filtering method applied here is described in detail in Aksoy et al. (2005) and Snyder and Zhang (2003). The ensemble Kalman filter assimilation equations are the following:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}(\mathbf{y}^0 - \mathbf{H}\bar{\mathbf{x}}^b), \tag{5a}$$

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}. \tag{5b}$$

Here, $\bar{\mathbf{x}}^b$ is the mean forecast (background) state vector, $\mathbf{P}^b$ is the forecast covariance matrix. $\bar{\mathbf{x}}^a$ is the analysis mean state vector (after the assimilation update), $\mathbf{y}^0$ is the observation vector, $\mathbf{R}$ is the observational error matrix, $\mathbf{H}$ is an operator matrix mapping model state to observational space, $\mathbf{H}^T$ is its transpose, and $\mathbf{K}$ is the Kalman gain matrix. Essentially, the assimilation updates the state vector variables using linear combinations of the observations and the model forecast. The weighting factors in the linear combination (given by $\mathbf{K}$) are determined by the degree to which observational variables covary with state variables (represented by $\mathbf{P}^b$) and the error in the observations (represented by $\mathbf{R}$) versus the uncertainty in the forecast (also in $\mathbf{P}^b$). Features of the ensemble-based method applied here include sequential processing of observational data (Whitaker and Hamill, 2002), square root filtering (Whitaker and Hamill, 2002), and the use of covariance localization (Gaspari and Cohn, 1999; Houtekamer and Mitchell, 2001) with a radius of influence of 100 grid points. No covariance inflation is used in the technique applied here.

In ensemble-based Kalman filtering, $\mathbf{P}^b$ is forecast directly through the evolution of ensemble member perturbations. Here, initialization of the ensemble is achieved through a "climatological" scheme, in which initial values of vorticity and buoyancy for each ensemble member are statistically sampled from a time series of data generated from a prior model run. A normal probability distribution that is centered at the maximum diurnal heating phase (local noon) and has a standard deviation of 8 h is used for the random sampling of initial states (more details are provided in Aksoy et al., 2005). This assimilation method has been used to study the nonlinearity of the sea breeze circulation and the effects of assimilating buoyancy observations on meteorological state uncertainties (Aksoy et al., 2005, 2006). However, neither chemical tracer dynamics nor chemical observations were represented or considered.

In this paper, we extend the work on ensemble-based Kalman filtering to consider a coupled meteorological and chemical tracer model application. For this work, tracer concentrations were built into the observational and state vectors for assimilation calculations in order to investigate the use and optimization of concentration observations for improved meteorological and air quality predictions.

3. Numerical experiments in chemical tracer forecasting

Using the coupled model, we performed three numerical experiments. Through these, we investigate (1) uncertainty in tracer concentration forecast predictions, (2) effects of the assimilation of tracer concentration data on the prediction, and (3) use of the ensemble-based Kalman filtering assimilation system for observation targeting.

3.1. Ensemble forecasts of tracer concentrations

In order to understand the predictability of tracer concentrations and for benchmarking improvements in the prediction due to data assimilation, it is important to understand the underlying uncertainty in predicted tracer concentrations. To investigate this uncertainty in the coupled system, we performed pure forecast simulations with a 50 member ensemble and no data assimilation. Simulations were initialized at the maximum diurnal heating phase (local noon). Initial values of vorticity and buoyancy were varied among ensemble members using the climatological scheme discussed above. Initial concentration profiles were not varied. Model parameters were also not varied among ensemble members. The variability in model results is therefore representative of uncertainty due to meteorological initial conditions and buoyancy forcing (the stochastic uncertainty of heating). The initial tracer concentration profile was assumed to be a constant value throughout the domain ($3 \times 10^{-5}\,\mathrm{kg\,m^{-3}}$). To simulate tracer emissions, an idealized lower boundary tracer source flux was also applied. As shown in Fig. 1, the source flux varies horizontally from 10^{-10} to $10^{-6}\,\mathrm{kg\,m^{-2}\,s^{-1}}$ with a maximum at 28 km to the land side of the coastline. The source flux remained constant in time. (Tracer concentration and source values and shapes used here were chosen to be simple representations of those observed for pollutants in the atmosphere.)

Physical model parameter values used were a mean wind of $0.5\,m\,s^{-1}$, Brunt–Vaisala frequency of $10^{-2}\,s^{-1}$, and vertical eddy diffusivities of $0.75\,m^{-2}\,s^{-1}$ for vorticity, buoyancy, and tracer concentration.

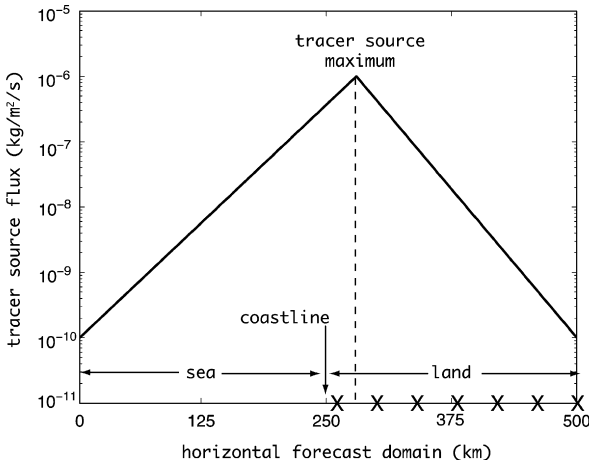

Fig. 1. Horizontal domain and tracer source configuration. The network of seven land surface observation locations are marked by X's. The network starts at 8 km to the land side of the coastline an has 40 km spacing. The solid line (and *y*-axis) provide the tracer source flux profile.

Diurnal heating profile parameters include an amplitude of $7\times10^{-6}\,m\,s^{-3}$ and horizontal and vertical scales of 10 km and 500 m, respectively.

Fig. 2 shows the evolution of ensemble mean winds and concentration distributions for a 36 h pure forecast. 0 and 24 h correspond to local noon and the maximum in the differential day-time heating source, 6 and 36 h (6 pm) correspond to peak temperatures over land, 9 and 33 h (9 pm) correspond to peak sea breeze winds, and 21 h (9 am) corresponds to the peak land breeze. Released tracer concentrations are largely confined to the area near the source maximum, tailing to the sea side after about 24 h. The sea-land breeze recirculation also allows concentrations to build somewhat near the surface source maximum, with advection away from the local area confined both horizontally and vertically.

Fig. 3 shows the corresponding evolution of predicted uncertainty (the ensemble standard deviation) in tracer concentrations. We see that uncertainty spreads relatively quickly in the horizontal, with maxima migrating diurnally between the land and sea sides of the coast. Uncertainty is largely confined to the near surface, with greater vertical

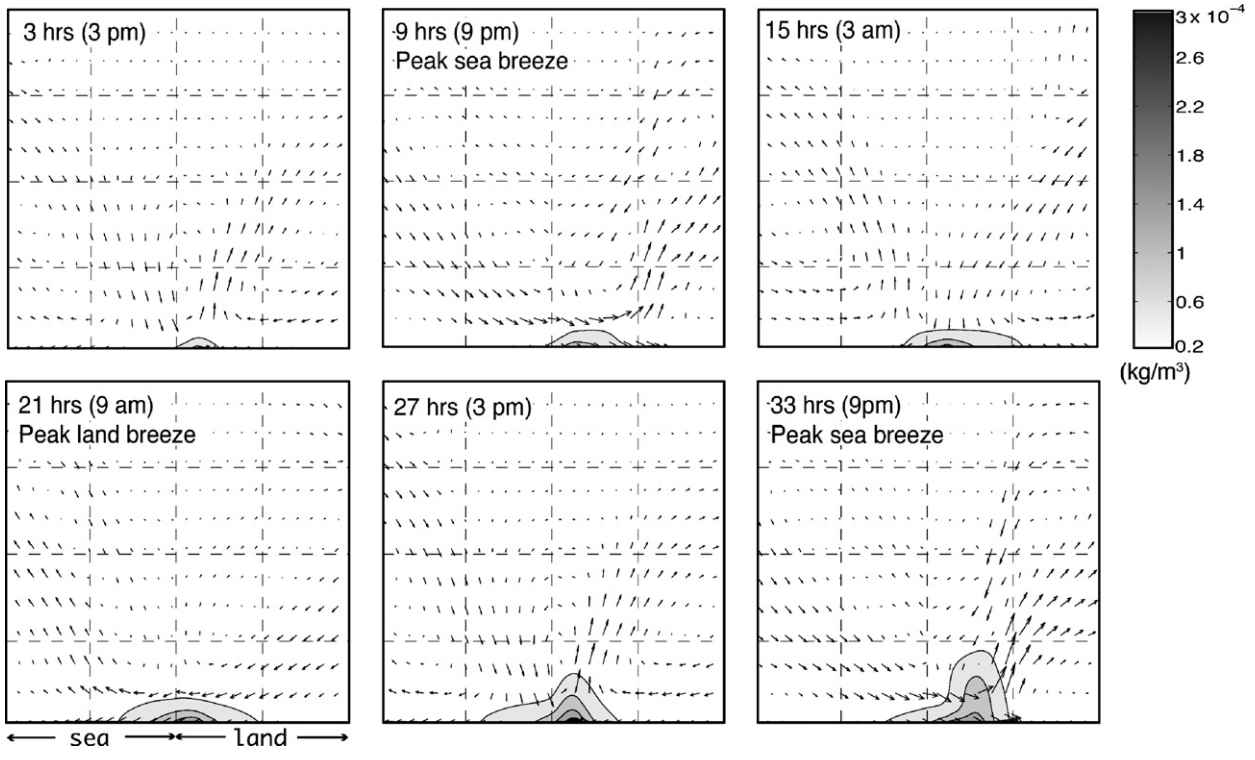

Fig. 2. Evolution of predicted mean winds (arrows) and tracer concentrations (contours) for a pure forecast simulation. The domain shown is the forecast model domain (without numerical sponge layers) of 500 km (horizontal) by 3 km (vertical).

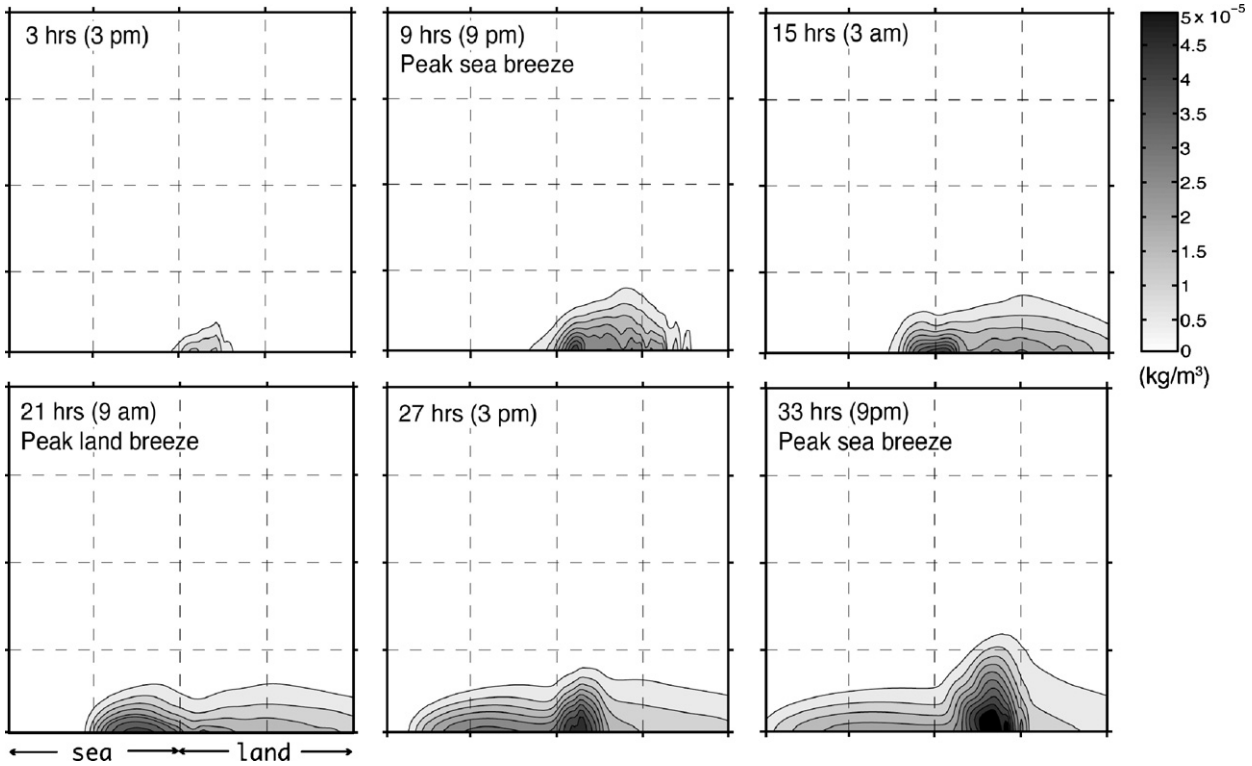

Fig. 3. Evolution of the predicted tracer concentration uncertainties (ensemble standard deviation) for the pure forecast simulation.

error growth occurring at times and locations near the peak sea breeze front. This correlation in timing and location with the sea breeze front is similar to that observed by Aksoy et al. (2005) for uncertainty in vorticity. The similar behavior of concentration uncertainty is consistent with our experimental set-up in which meteorological initial conditions and forcing are the sources of forecast uncertainty.

Fig. 4 provides the evolution of the domain averaged ensemble uncertainty for all prognostic state variables (vorticity, buoyancy, and concentration) for this 36 h simulation. (Here, we will discuss the pure forecast predictions.) We see that although there are short periods of error growth, the meteorological variable errors are damped with overall time during this period. Strong error growth is apparent for the chemical prediction. However, concentration errors also eventually decay overall with time, with a maximum at about 30 h. The chemical prediction errors appear to grow during the transition from the peak land to peak sea breeze, from about 0 to 9 h and from 21 to 30 h. Results are also shown for two longer (12 days) ensemble forecast simulations: one case with the same set-up as the 36 h run, and one case for which the initial values of vorticity and buoyancy were not varied among ensemble errors. (Hence, in this simulation the only source of error is the stochastic buoyancy forcing.) Comparing the long simulations with and without uncertainty in the initial meteorological conditions, we see that at earlier times, the total error is dominated by initial condition error, which largely propagates out of the domain after a few days. Nonetheless, the error due to stochastic buoyancy forcing grows in time. Aksoy et al. (2005) discuss the meteorological error growth characteristics of this representation of a chaotic forced dissipative system.

3.2. Assimilation of concentration observations

To investigate the effect of assimilating concentration observations on meteorological and tracer predictions, we performed a second experiment using ensemble-based Kalman filtering data assimilation. A 50 member ensemble was also used here. An arbitrarily chosen 51st member was used to represent the true evolution of the state and for extraction of observations. Concentration values were extracted for a network of seven land surface

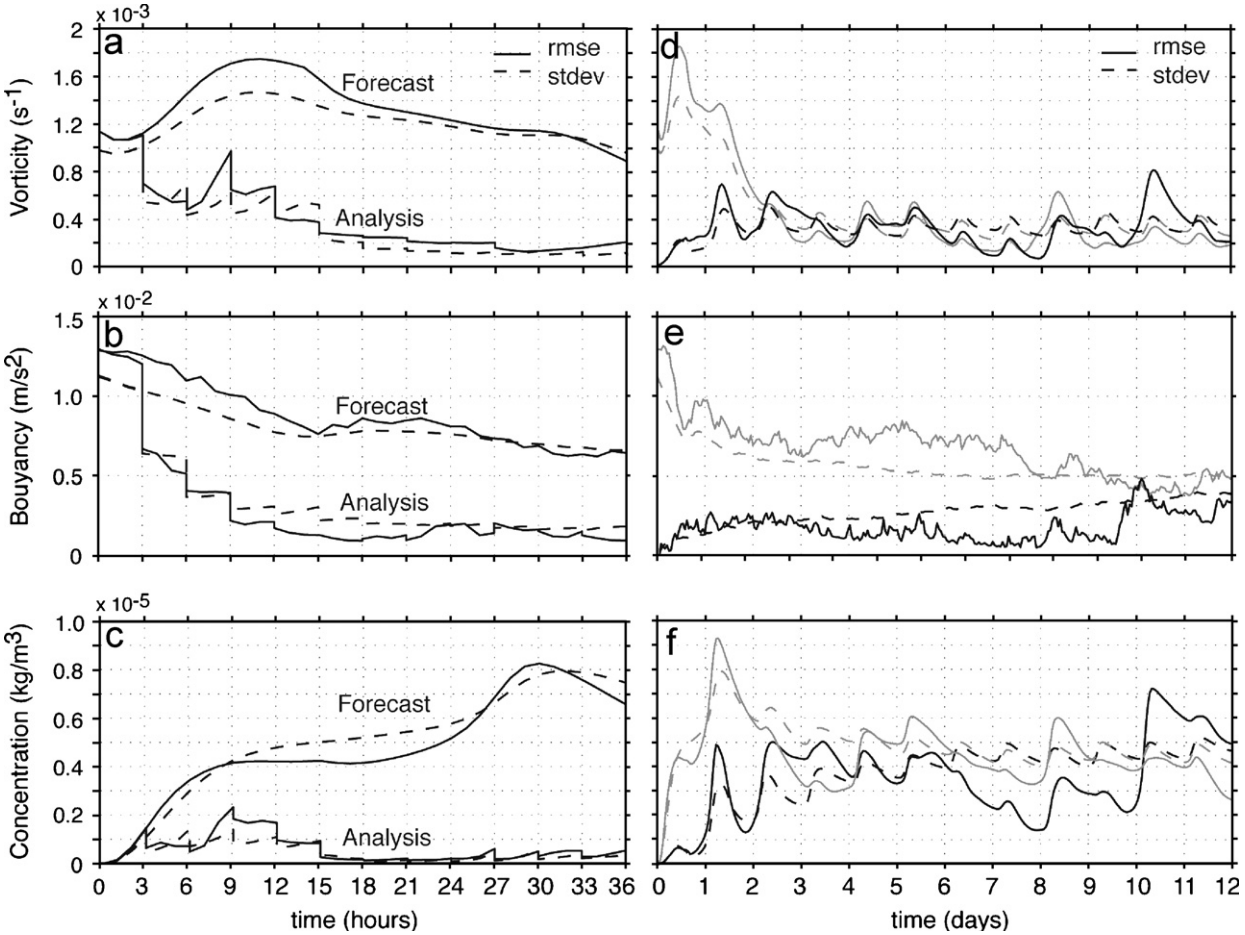

Fig. 4. Domain averaged ensemble root mean squared errors (rmse) with respect to the truth simulation and standard deviations (stdev) in model prognostic variables. Plots (a–c) provide trends for the 36 h forecast and analysis simulations. Plots (d–f) provide trends for two long-term (12 days) forecast simulations. The gray lines are for a case with the same set-up as the 36 h case (but with slight differences due to the forcing stochasticity). The black lines are for a case with no ensemble variability in the initial conditions.

locations (see Fig. 1). These truth values were stochastically perturbed to create observations, assuming a Gaussian distribution of errors with standard deviation of $1 \times 10^{-7}\,\mathrm{kg\,m^{-3}}$. The observations were assimilated into the predicted state every 3 h. Note that only concentration observations and no meteorological observations were used in the analysis. Ensemble member initialization, model set-up, and physical parameters used in this simulation were as described above for the forecast.

Figs. 5 and 6 show the evolution of winds and concentration distributions for the truth and mean analysis forecast, respectively. In the truth simulation, a strong vertical circulation during the sea breeze portion of the diurnal cycle is present. This results in advection of peak concentrations to higher heights and slightly less horizontal spread. The analysis mean is a significantly better representation of the truth than the pure forecast for both winds and concentrations. It captures the stronger vertical circulation and predicts concentrations fields that are very similar to the truth. In the pure forecast ensemble mean, the circulation wind speeds are significantly damped, with more horizontal and less vertical concentration spread.

Domain averaged ensemble uncertainty and mean error for analysis predictions of all prognostic state variables are compared with the pure forecast prediction in Fig. 4(a–c). Assimilation of concentration observations is clearly effective at constraining the meteorological errors, as well as the concentration errors. Hence, the filter is effective for both the

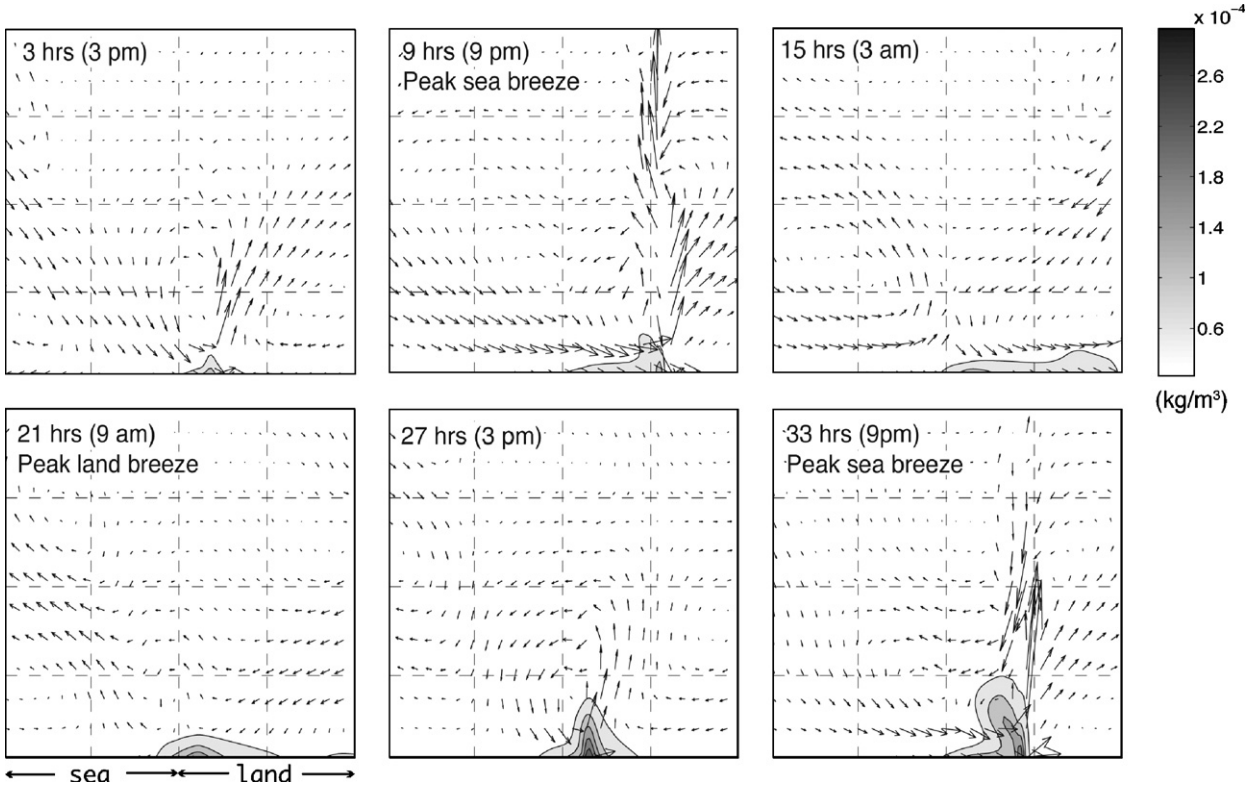

Fig. 5. Evolution of winds and tracer concentrations for the truth simulation.

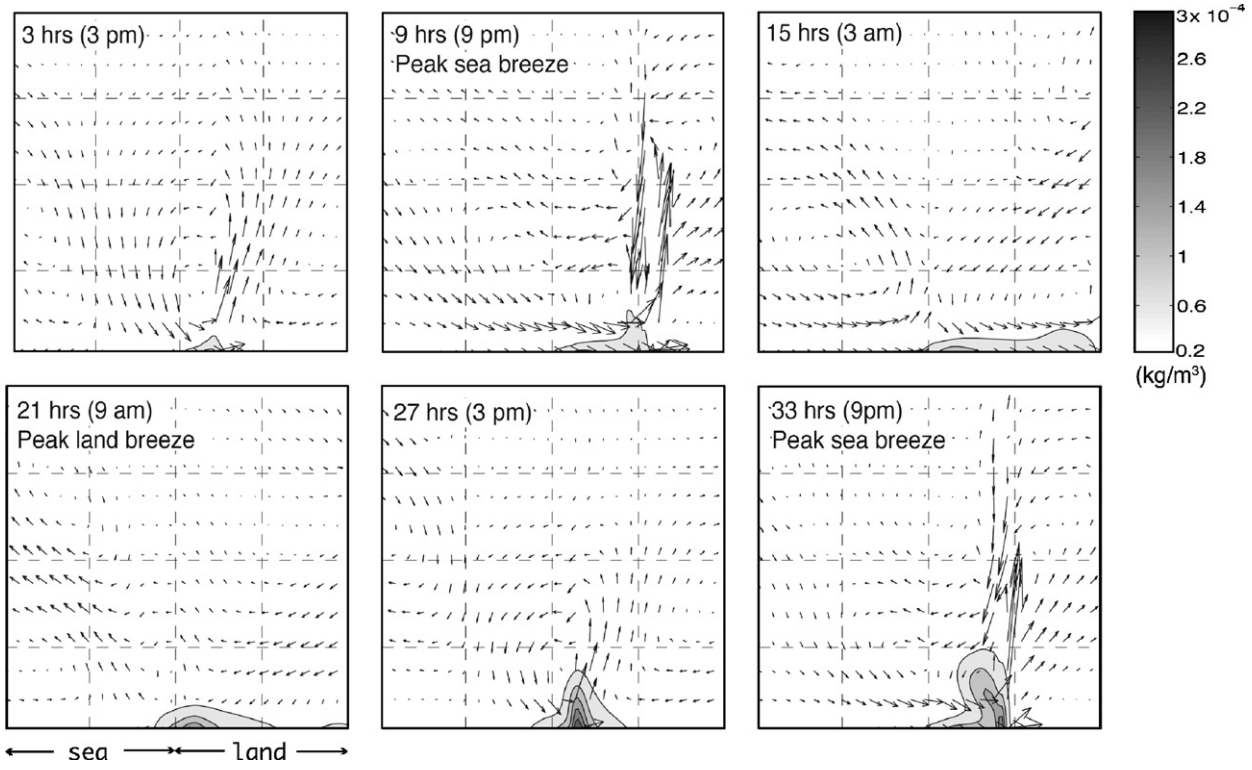

Fig. 6. Evolution of analysis mean winds and tracer concentrations.

observed variable (concentration) and the unobserved variables (vorticity and buoyancy).

To investigate the robustness of our assimilation results, several additional ensemble simulation cases were also performed. We investigated the sensitivity of the assimilation results to the interval between analyses, the radius of influence, the observational spacing, the observational error, and the observational variables assimilated. Simulations were also performed with a distinct initial vertically varying concentration profile. Fig. 7 shows results of several sensitivity cases. The error in predicted variables is somewhat sensitive to all the filter parameters, particularly at early times. However, in all cases the assimilation significantly reduced the uncertainty in the forecast. Sensitivity simulations with single observed variables other than concentration (i.e. buoyancy alone or vorticity alone) indicate that concentration is a somewhat less effective observed variable than the meteorological variables for improving the forecast of meteorological variables. Simulations with both concentration and a meteorological variable observed, performed better overall than any single observed variable.

3.3. Observation targeting

The ensemble-based Kalman filter provides a mechanism for optimizing the spatial and temporal location of fixed observational networks and targeted observations. Hamill and Snyder (2002) developed a method to select observation locations based on maximizing the improvement (decrease in uncertainty) in model fields that would result due to an assimilation cycle and applied it to a weather forecasting application. In this optimization method, the norm used for total decrease in model uncertainty is the sum over all state variables of the individual differences in variances, or the trace of $\mathbf{P}^{b}-\mathbf{P}^{a}$, where $\mathbf{P}^{a}$ is the post-analysis background error covariance matrix. $\mathbf{P}^{b}-\mathbf{P}^{a}$ can be written as (Hamill and Snyder, 2002)

$$\mathbf{P}^{b}-\mathbf{P}^{a}=\mathbf{P}^{b}\mathbf{H}^{T}(\mathbf{H}\mathbf{P}^{b}\mathbf{H}^{T}+\mathbf{R})^{-1}\mathbf{H}\mathbf{P}^{b}. \quad (6)$$

To investigate the potential use of Hamill and Snyder's method in the context of chemical observational network design and air quality forecast improvement, we have applied their development to our system. Using post-processing, we calculate the effect of a single observation (y_o) located at a model grid point. For this case, the trace of $\mathbf{P}^{b}-\mathbf{P}^{a}$ reduces to

$$\Gamma=\sum_{i=1}^{M}(\sigma_b^2-\sigma_a^2)_i=\frac{1}{\sigma_{x_{y_0}}^2+\sigma_{y_0}^2}\sum_{i=1}^{M}\mathrm{cov}^2\left(x_i,x_{y_0}\right), \quad (7)$$

where M is the size of the state vector (the number of state variables times the number of domain grid points), σ^2 is the variance, x_i is a state vector variable, x_{y_0} is the state variable corresponding to the observation (or Hx), and cov is the covariance. We refer to Γ as the observational impact factor. Here, we have applied it to calculate values of the impact factor for assimilation of individual tracer concentration observations, in order to investigate implications for the location of optimal chemical observations.

Fig. 8 shows the spatial and temporal evolution of the observation impact factor fields over the simulation period. Values are shown every 6 h after the assimilation of concentration observations from the fixed network (and all prior network analyses). Maximum values indicate locations where additional observations of tracer concentration are predicted to lead to greatest improvement (over all prognostic model variables) in the model prediction. The figure indicates optimal locations for observations that change with time. The most valuable observations are generally located primarily on the land side of the domain and closer to the surface (i.e., close to the source maximum), though complicated structures are apparent. When a sea breeze front is present, observations near the front appear most critical, while at other times, preferred locations are more broadly distributed. The value of assimilating observations also changes significantly with time, decreasing overall (see the logarithmic scale). We note that this optimization method, represented by Eqs. (6) and (7), values the variance of all state variables equally, regardless of the differences in magnitude between variables. To investigate impacts of this valuation, we also calculated impact factors for each state variable separately (not shown). Although the impact factor magnitudes change for each variable, the shape of the impact factor fields (and hence predicted locations of optimal observations), remains very similar.

The above impact factor was calculated after the assimilation cycle using the regular network observations. Hence, it provides optimal locations for

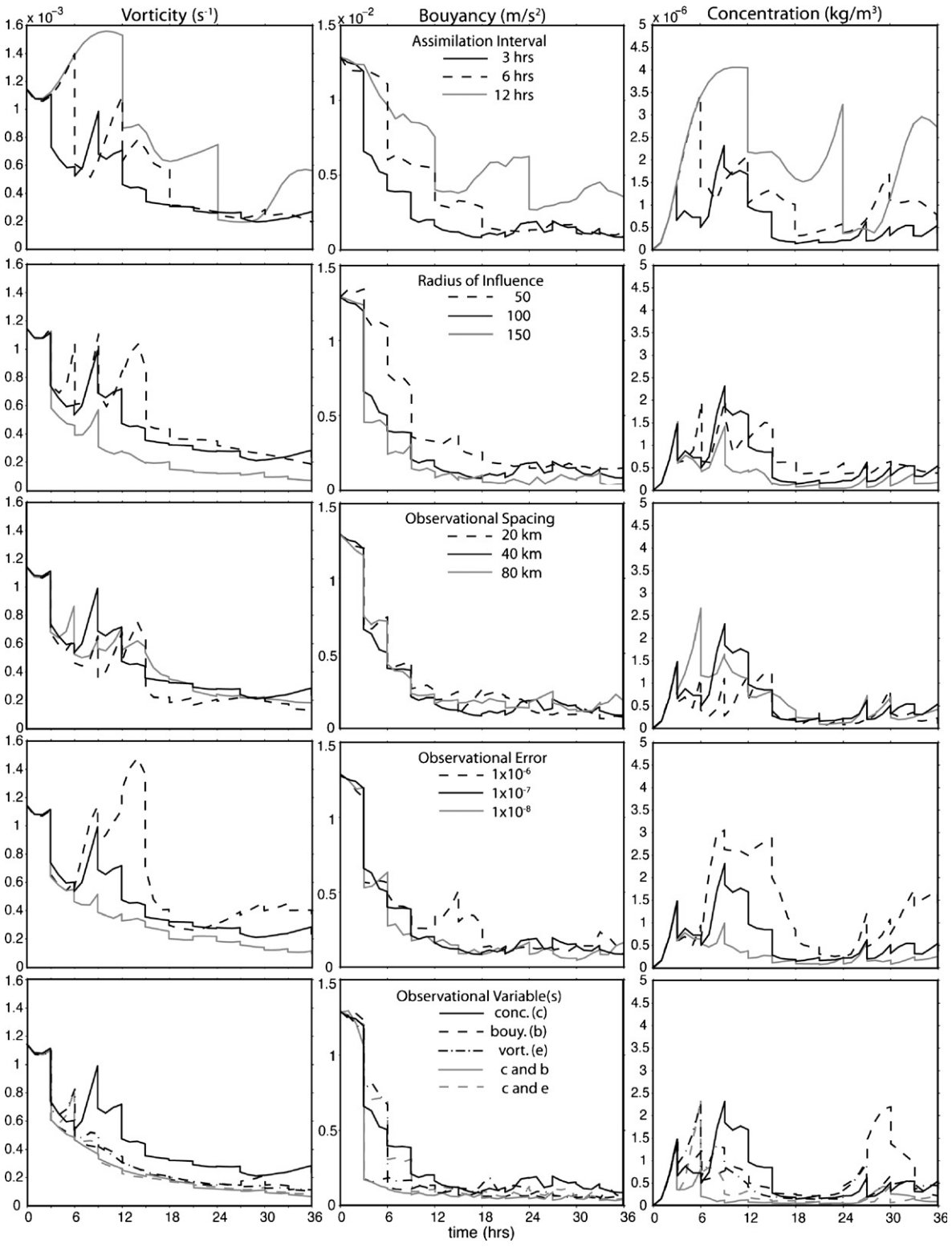

Fig. 7. Sensitivity of the rmse of the model prognostic variables. For observations of buoyancy and vorticity, observational errors of 1×10^{-3} and 1×10^{-4} were used, respectively.

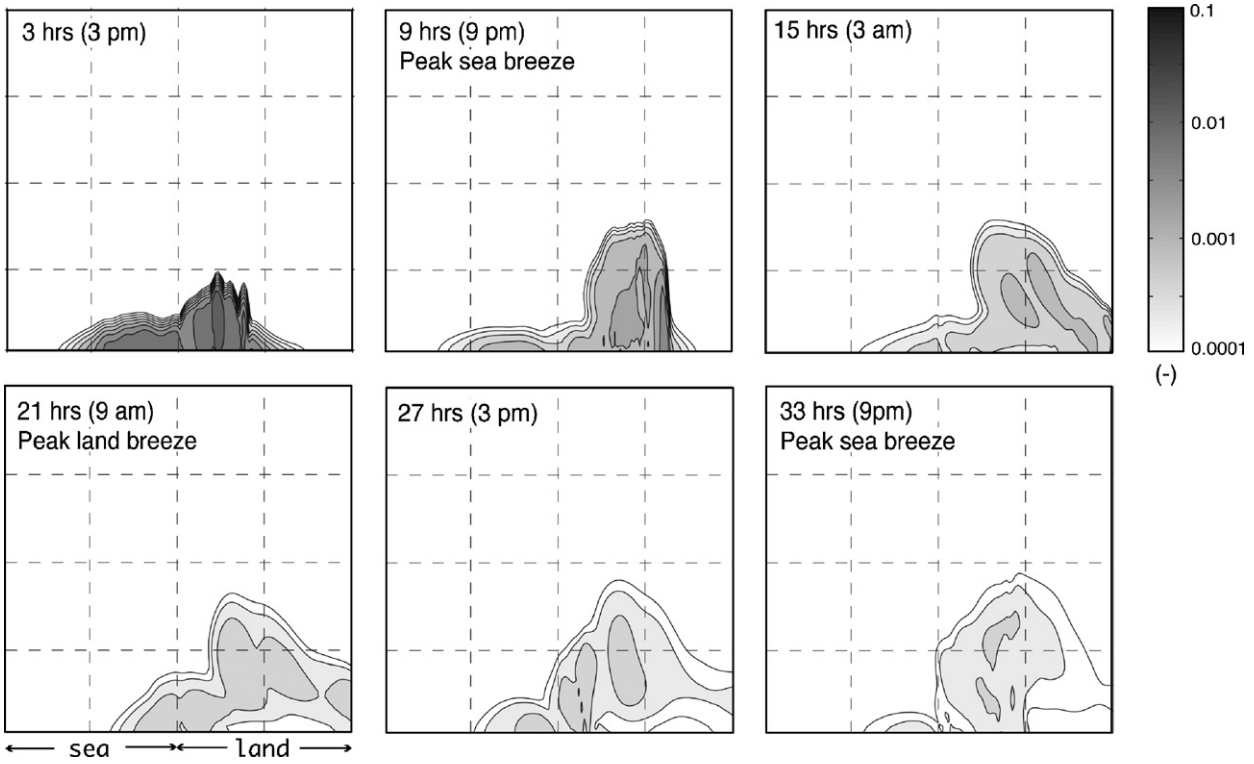

Fig. 8. Observational impact factor fields at analysis times after the assimilation of concentration observations from the fixed network (and all prior network analyses).

targeted observations additional to the regular network. However, the calculation also requires information on observation values at the analysis times. For real-time forecasting, network observations for a given analysis time are unavailable a priori. Additionally, impact factor values are only a function of state covariance and observational variance (error) information, which is often available before the analysis update. Fig. 9 shows impact factor fields at analysis times but prior to the assimilation of concentration observations from the fixed network compared with those after assimilation. Substantially, similar structures are apparent in the normalized impact factor fields both before and after analysis of the regular network observations. This demonstrates the potential usefulness of this norm for locating promising adaptive observations in a predictive mode, without a priori information on regular network observation values or even observation locations. However, further work is needed to test the success of this strategy through addition of targeted observations at the selected locations.

4. Conclusions

We study the impact of ensemble-based Kalman filter assimilation of chemical tracer concentration observations on improving the prediction of both chemical and meteorological model variables for an idealized sea breeze circulation. We also investigate the potential utility of an optimization technique based on the Kalman filtering equations for design of fixed air quality observational networks and targeted observations. By comparing the truth simulation with ensemble forecasts, both with and without assimilation of concentration data, we find that ensemble-based Kalman filtering assimilation of concentrations observations effectively improved the forecast of both the observed variable (concentration) and the unobserved meteorological variables. This suggests considerable potential value in using the plethora of air pollution data available from regulatory air quality monitoring networks for improving skill in meteorological forecasts (even for purposes unrelated to air pollution), when meteorological and chemical models are routinely coupled.

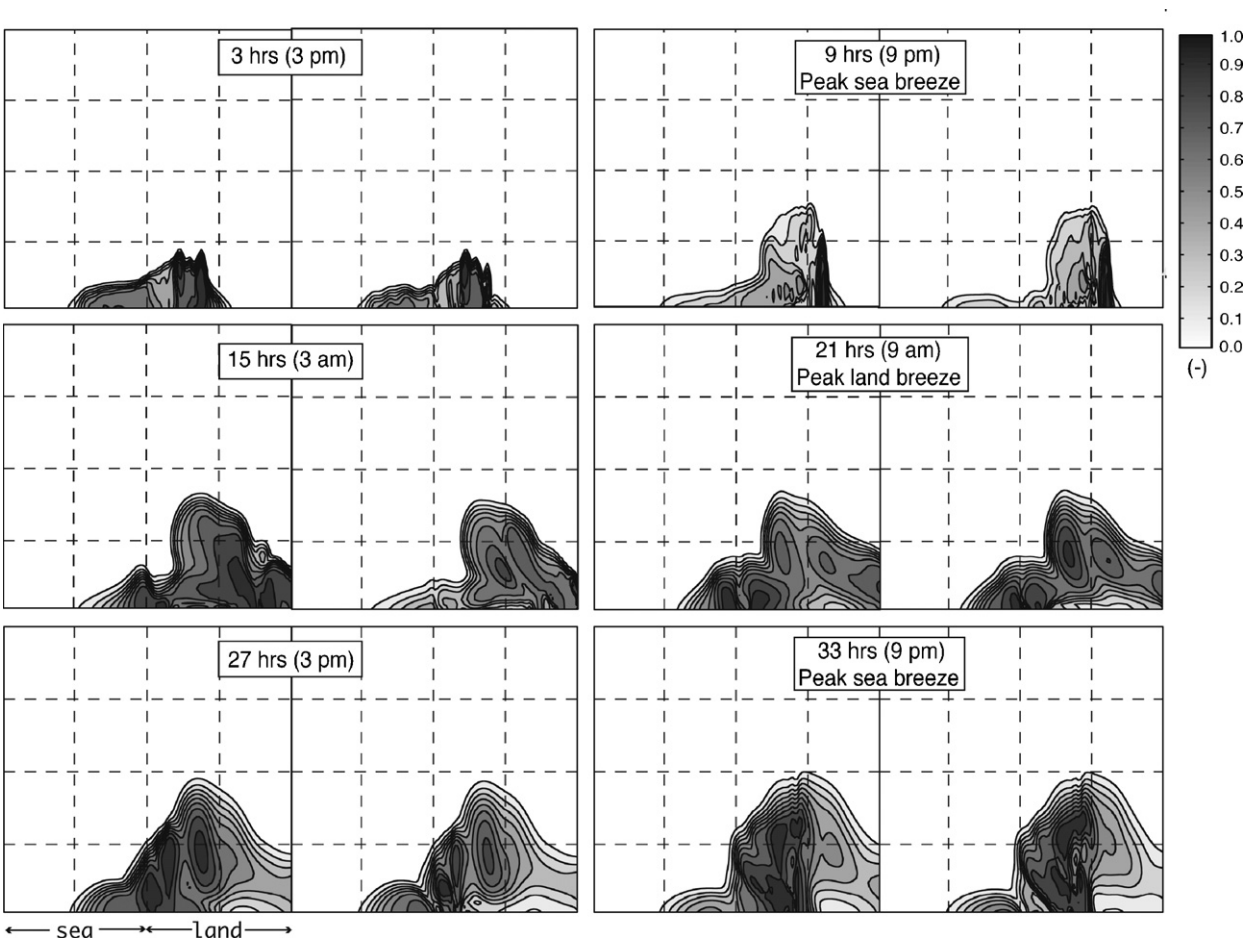

Fig. 9. Normalized observational impact factor fields before (left side of each pair) and after (right side) observation assimilation from the fixed network. Impact factor values have been normalized by dividing by the maximum value in the given field.

We also demonstrate the potential of the impact factor derived from ensemble-based Kalman filtering for selection of observation locations.

With this work, we advance understanding of the use of ensemble-based Kalman filter data assimilation of chemical observations with a nonlinear, two-dimensional model in which dynamics and chemical tracer transport are coupled. The context of our experiments is a chaotic force dissipative dynamic system, a sea breeze circulation. Such circulations are important for many municipalities with difficult air quality problems, but there has been little work in the past on using ensemble-based data assimilation for improving predictions for such systems. We focus here on studying the effects when the uncertainty in the forecast is driven by meteorological initial condition error. The results of our idealized modeling study suggest the potential value of chemical observations and suggest directions regarding targeted observation planning. However, we must note that realistic systems affecting air quality are very complex and have many sources of error (notably emissions data and model representation) that we have not investigated here. Further significant work is needed to understand the applicability of these results for a variety of realistic dynamic systems and complex coupled models.

Acknowledgments

This work was supported in part by the Texas Environmental Research Consortium Project No. H24-2003 and the Environmental Protection Agency Cooperative Agreement No. R-83037701. We would also like to acknowledge the services provided by Research Computing at USF.

References

Aksoy, A., Zhang, F., Nielsen-Gammon, J.W., Epifanio, C., 2005. Ensemble-based data assimilation for thermally forced circulations. Journal of Geophysical Research 110, D16105.

Aksoy, A., Zhang, F., Nielsen-Gammon, J.W., 2006. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea breeze model. Monthly Weather Review 134, 2951–2970.

Asselin, R., 1972. Frequency filter for time integrations. Monthly Weather Review 100, 487–490.

Banta, R.M., Senff, C.J., Nielsen-Gammon, J., Darby, L.S., Ryerson, T.B., Alvarez, R.J., Sandberg, S.P., Williams, E.J., Trainer, M., 2005. A bad air day in Houston. Bulletin of the American Meteorological Society 86 (5), 657–669.

Bao, J.-W., Michelson, S.A., Mckeen, S.A., Grell, G.A., 2005. Meteorological evaluation of weather-chemistry forecasting model using observations from the TEXAS AQS 2000 field experiment. Journal of Geophysical Research 110, 21105.

Bergin, M.S., Noblet, G.S., Petrini, K., Dhieux, J.R., Milford, J.B., Harley, R.A., 1999. Formal uncertainty analysis of a Lagranian photochemical air pollution model. Environmental Science and Technology 33 (7), 1116–1126.

Berliner, L.M., Lu, Z.-Q., Snyder, C., 1999. Statistical design for adaptive weather observations. Journal of Atmospheric Sciences 56, 2536–2552.

Boybeyi, Z., Raman, S., Zannetti, P., 1995. Numerical investigation of the possible role of local meteorology in Bhopal gas accident. Atmospheric Environment 24 (4), 479–496.

Chang, M.E., Hartley, D.E., Cardelino, C., Haas-Laursen, D., Change, W.L., 1997. On using inverse methods for resolving emissions with large spatial in homogeneities. Journal of Geophysical Research 102, 16023–16036.

Dabberdt, W.F., Miller, E., 2000. Uncertainty, ensembles and air quality dispersion modeling: applications and challenges. Atmospheric Environment 34 (27), 4667–4673.

Daescu, D.N., Carmichael, G.R., 2003. An adjoint sensitivity method for the adaptive location of the observations in air quality modeling. Journal of the Atmospheric Sciences 60, 434–450.

Elbern, H., Schmidt, H., Talagrand, O., Ebel, A., 2000. 4-D variational data assimilation with an adjoint air quality model for emission analysis. Environmental Modelling and Software 15, 539–548.

Emanuel, K.A., Langland, R., 1998. FASTEX adaptive observations workshop. Bulletin of the American Meteorological Society 79, 1915–1919.

Emanuel, K.A., et al., 1995. Report of the first prospectus development team of the US Weather Research Program to NOAA and NSF. Bulletin of the American Meteorological Society 76, 1194–1208.

Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 53, 343–367.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society 125, 723–757.

Hamill, T.M., Snyder, C., 2002. Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. Monthly Weather Review 130, 1552–1572.

Hanea, R.G., Velders, G.J.M., Heemink, A., 2004. Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model. Journal of Geophysical Research 109, D10302.

Heemink, A.W., Segers, A.J., 2002. Modeling and prediction of environmental data in space and time using Kalman filtering. Stochastic Environmental Research and Risk Assessment 16, 240–255.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review 129, 123–137.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, New York.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 130–141.

Lorenz, E.N., Emanuel, K.A., 1998. Optimal sites for supplementary observations: simulation with a small model. Journal of the Atmospheric Sciences 55, 399–414.

Mendoza-Dominguez, A., Russell, A.G., 2001. Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. Atmospheric Environment 35, 2879–2894.

Navon, I.M., 1998. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dynamics of Atmosphere and Oceans 27, 55–79.

Morss, R.E., Emanuel, K.A., Snyder, C., 2001. Observational strategies for improving numerical weather prediction. Journal of the Atmospheric Science 58, 210–232.

Sax, T., Isakov, V., 2003. A case study for assessing uncertainty in local-scale regulatory air quality modeling applications. Atmospheric Environment 37 (25), 3481–3489.

Seigneur, C., 2005. Air pollution: current challenges and future opportunities. A.I.C.h.E. Journal 54 (2), 356–364.

Seinfeld, J.H., 2004. Air pollution: a half century of progress. A.I.C.h.E. Journal 50 (6), 1096–1108.

Snyder, C., Zhang, F., 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. Monthly Weather Review 131, 1663–1677.

Stuart, A.L., Jain, S., Libicki, S.B., 1996. The use of long-term meteorological information to predict impact probabilities resulting from toxic chemical releases. In: Proceedings of the International Topical Meeting of Probabilistic Safety Assessment. American Nuclear Society, October.

Tippett, M.K., Anderson, J.L., Bishop, C.H., M Hamill, T., Whitaker, J.S., 2003. Ensemble square root filters. Monthly Weather Review 131, 1485–1490.

Whitaker, J.S., Hamill, T.M., 2002. Ensemble data assimilation without perturbed observations. Monthly Weather Review 130, 1913–1924.

Zhang, F., Bei, N., Nielsen-Gammon, J.W., Li, G., Zhang, R., Stuart, A., Aksoy, A., 2006. Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. Journal of Geophysical Research. (in press).