J3.5 ENSEMBLE-BASED DATA ASSIMILATION AT A COASTLINE

Altug Aksoy *, Fuqing Zhang, John W. Nielsen-Gammon, Craig Epifanio
Texas A&M University, College Station, Texas

Chris Snyder
National Center for Atmospheric Research †, Boulder, Colorado

1. INTRODUCTION

Perhaps one of the most prominent dynamical features at coastlines is the *sea breeze* which is a thermally-induced diurnal boundary layer circulation. The numerical modeling of such circulations imposes interesting challenges, both from a perspective of interactions of larger-scale synoptic features with local inertia-gravity wave circulations and sensitivity to boundary layer parameterization schemes.

The sea breeze is perhaps one of the most extensively studied phenomena in atmospheric dynamics. Walsh (1974), Anthes (1978), and Rotunno (1983) provide a good review and a comprehensive list of references on the theoretical aspects of the phenomenon, while Wakimoto and Atkins (1994) and Atkins et al. (1995) give excellent reviews on the observational characteristics of the sea breeze. Simpson (1994) and, more recently, Miller et al. (2003) also provide excellent "primers" mostly on the observational features and some theoretical aspects of the sea breeze and other local wind phenomena.

Nielsen-Gammon (2001) pointed to some interesting features of the land/sea breeze observed along the coast of Texas, based on the results obtained from the Texas 2000 Air Quality Study (TexAQS-2000) which was a major field program that ran from August 15, 2000 to September 15, 2000 with the goal of understanding the formation and transport of ozone and particulate matter in eastern Texas, particularly around Houston. During TexAQS-2000, while measurements at several buoys showed that the land/sea breeze existed with some expected general characteristics (strong diurnal variation and the clockwise turning of the wind with time), measurements also revealed characteristics in support of Rotunno's (1983) linear theory, some predictions of which were counter-intuitive compared to the common experience (for instance, a buoy located 35 km offshore normally experienced a sea breeze during the night and a land breeze during the day).

* *Corresponding Author Address:* Altug Aksoy, Texas A&M University, Dept. of Atmospheric Sciences, College Station, TX 77843-3150.
E-mail: aaksoy@ariel.met.tamu.edu.

† The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Although some of these interesting features of the sea breeze were explained with linear theory (Rotunno, 1983) with success, the theory's limitations, especially in the areas of variability of vertical stratification and diffusivity, resulted in discrepancies with the observations of the phenomenon, such as the symmetric nature of land- and sea-breeze circulations and the absence of a sea-breeze front which is a strongly nonlinear feature (Simpson, 1994). Due to these limitations, the linear model should only be treated as a general framework for the land/sea breeze phenomenon while the description of the detailed local structure (which is mostly governed by the nonlinear features like the sea breeze front) can be carried out more effectively by the help of numerical models.

Because of its strong sensitivity to small-scale variations and its highly nonlinear nature, progress in the accurate sea-breeze modeling has been slower than the development of the linear theory. Among the major difficulties that are involved, successful integration of parameterization schemes with the dynamical model turns out to be one the most critical aspect of the sea-breeze modeling. Some of the factors which affect the land and sea breeze circulation are (Simpson,1994): (1) diurnal variation of the ground temperature, (2) diffusion of heat, (3) spatial and diurnal variation of static stability, (4) Coriolis force, (5) diffusion of momentum, (6) topography, and (7) prevailing winds. Clearly, while most of these are variables/processes that are usually treated within the context of parameterization schemes, they are rarely subject to direct observation or real-time estimation. Both of these facts (the importance of parameterization schemes in the context of sea-breeze modeling, and the lack/insufficiency of observations critical for the parameterization schemes involved) contribute to the complexity of the problem and warrant the integration of a well-designed *parameter estimation procedure* into the numerical model.

One very robust approach to parameter estimation is the utilization of already existing data assimilation schemes. Conventional Kalman filter techniques have long been used successfully in engineering for this purpose [see Gelb (1994) for a brief but useful discussion on this subject and further references]. Problems encountered in such engineering applications, however, are usually of relatively small degrees of freedom and the use of full-Kalman-filter-type techniques are well warranted under these circumstances. With the introduction of the ensemble Kalman filter (EnKF) by Evensen (1994) for geophysical

applications, the feasibility of the Kalman filter approach has been extended to problems with much larger degrees of freedom typically found in atmospheric sciences (Burgers et al., 1998; Houtekamer and Mitchell, 1998; Mitchell and Houtekamer, 2000; Houtekamer and Mitchell, 2001; Snyder et al., 2001; Snyder and Zhang, 2003; Zhang et al., 2003). This recent interest has emerged because of a number of appealing properties the filter offers (Zhang et al., 2002): no adjoints are required of the forecast model, estimates of forecast uncertainty are produced at no extra cost, it is highly parallel, and it is largely independent of the forecast model. While these studies have mostly focused on the direct application of the filter to a specific atmospheric modeling environment to improve the initial state of the system, more recently, Anderson (2001) applied the filter's statistical properties to the problem of model error by utilizing sample probability distributions of some of model parameters given observations. Although the idea of inserting these parameters into the state vector and using assimilation to estimate values for unknown model parameters appears to be a very appealing technique, Anderson (2001) argued that "it remains an open question whether there is sufficient information in available observations to allow this approach in current-generation operational models".

The purpose of this study is to investigate the potential of the EnKF in the area of model error estimation and reduction in a sea-breeze environment. Motivations for the selection of such a filter-dynamics combination are many-sided ranging from (1) better understanding of sea-breeze dynamics through uncertainty information produced by the EnKF and (2) existing lack of research in the area of EnKF integration to meso- and smaller-scale numerical models, especially with strong local forcing, to (3) improving our understanding and treatment of model error through the use of the EnKF. Our approach is to tackle the problem in an increasing-complexity setting. First, for the proof-of-concept demonstration, results from an application of the EnKF to a highly idealized harmonic oscillator system are presented. The second step involves integrating the EnKF to a more complicated two-dimensional nonlinear sea-breeze model and performing similar tests as the first step. The third and final step entails the investigation of whether findings of the first two steps can be generalized to an operational application, i.e. to a complex mesoscale numerical model like MM5. This report should be viewed as the summary of a work-in-progress. Our plan is to report more detailed results and discussions at the conference.

2. DESCRIPTION OF NUMERICAL MODELS AND THE ENSEMBLE KALMAN FILTER

2.1 The Stochastically-Forced Harmonic Oscillator Model

Some of the key characteristics of the sea breeze can be represented by a very simple harmonic oscillator system of the following form:

$$\ddot{x} + \mu\dot{x} + k^2 x = A\cos(\omega_0 t) + \varsigma(t)\,. \tag{1}$$

This is, of course, the equation for the relative displacement x of a linear spring with regular sinusoidal forcing. The parameters μ, k^2, A, and ω_0 are viscosity, elasticity, forcing amplitude, and forcing frequency, respectively. As this form of the spring system is linear in nature, differences between solutions with different initial conditions decay, rather than grow, with time, so the system is completely predictable (and observations are unnecessary). Thus, in order to introduce uncertainty to the system, a stochastic error term, $\varsigma(t)$, is included.

By setting $\dot{x} = v$ (thus defining v as the velocity), the single-equation system can be easily converted to a two-equation system of the following form:

$$\begin{pmatrix}\dot{x}\\ \dot{v}\end{pmatrix} = \underbrace{\begin{bmatrix}0 & 1\\ -k^2 & -\mu\end{bmatrix}\begin{pmatrix}x\\ v\end{pmatrix}}_{\text{Homogeneous dynamics}} + \underbrace{\begin{pmatrix}0\\ A\cos(\omega_0 t)\end{pmatrix}}_{\text{Forcing}} + \underbrace{\begin{pmatrix}0\\ \varsigma(t)\end{pmatrix}}_{\text{Stochastic error}}\,. \tag{2}$$

For linear systems, it is relatively easy to model this system by a transition matrix approach. Similar to equation (2) but ignoring model error, we can describe a dynamical system in the following general matrix form:

$$\dot{\mathbf{x}} = \mathbf{M}\cdot\mathbf{x} + \mathbf{g}(t)\,, \tag{3}$$

where $\mathbf{x}$ is the state vector (x;v), $\mathbf{M}$ is the dynamics matrix, and $\mathbf{g}(t)$ is the forcing in vector form. The transition-matrix $\mathbf{\Phi}_k$ for the above equation can be defined through the equation (Gelb, 1974)

$$\mathbf{x}_{m+1} = \mathbf{\Phi}_m\cdot\mathbf{x}_m + \int_{t_m}^{t_{m+1}} \mathbf{\Phi}_m\cdot\mathbf{g}(\tau)d\tau\,. \tag{4}$$

For a stationary (or even marginally stationary) transition matrix and a given time interval Δt, it can be shown that (Gelb, 1974)

$$\mathbf{\Phi}(\Delta t) = e^{\mathbf{M}\cdot\Delta t}\,. \tag{5}$$

As a result, our linear numerical model can be described as

$$\mathbf{x}_{m+1} = e^{\mathbf{M}\cdot\Delta t}\cdot\mathbf{x}_m + \int_{t_m}^{t_{m+1}} e^{\mathbf{M}\cdot\Delta t}\cdot\mathbf{g}(\tau)d\tau\,, \tag{6}$$

the final form of which used for the experiments presented in this study is as follows:

$$\begin{pmatrix} x_{m+1} \\ v_{m+1} \end{pmatrix} = e^{\mathbf{M}\cdot\Delta t}\begin{pmatrix} x_m \\ v_m \end{pmatrix} + e^{\mathbf{M}\cdot\Delta t}\frac{A}{\omega_0}\begin{pmatrix} 0 \\ \sin[\omega_0(m+1)\Delta t] - \sin[\omega_0 m\Delta t] \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_m \end{pmatrix}. \quad (7)$$

This version of the model has the forcing term directly integrated. The stochastic error term, ε_m, is explicitly added to the resulting form of equations (at every time step, a non-correlated error with a Gaussian distribution is computed and added on the forcing) and is related to the error term $\varsigma(t)$ through its variance, the dynamics matrix **M**, and Δt.

All of the assimilation experiments presented here were performed by observing only the variable x. Additionally, if not stated otherwise, parameter values have been fixed to facilitate comparison between experiments (k^2 = 5s^{-2}, μ = 1s^{-1}, A = 500ms^{-2}, and ω_0 = 4s^{-1}).

2.2 The Two-Dimensional Nonlinear Sea-Breeze Model

Similar to Rotunno (1983), consider a Cartesian coordinate system in which a flat surface is represented by $z = 0$ and the coastline is at $x = 0$ with land for $x > 0$ and sea for $x < 0$. The general equations of motion can be simplified to a two-dimensional nonlinear, non-rotational, incompressible, and hydrostatic set in the following manner:

$$\left.\begin{aligned} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} &= -\frac{\partial p}{\partial x} + \kappa_u\frac{\partial^2 u}{\partial z^2}, \\ b &= \frac{\partial p}{\partial z}, \\ \frac{\partial b}{\partial t} + u\frac{\partial b}{\partial x} + w\frac{\partial b}{\partial z} + N^2 w &= Q + \kappa_b\frac{\partial^2 b}{\partial z^2}, \\ \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} &= 0, \end{aligned}\right\} \quad (8)$$

where u and w are horizontal and vertical winds, respectively; κ_u and κ_b are vertical viscosity constants for momentum and heat, respectively; $b = g\theta'/\theta_0$ is the buoyancy where g is the acceleration of gravity and $\theta = \theta_0 + \theta_B() + \theta'(x, z, t)$ is the potential temperature where θ_0 and θ_B are reference and background potential temperatures, respectively; $N^2 = (g/\theta_0)(\partial\theta_B/\partial z)$ is the Brunt-Väisälä frequency; and the explicit heating term Q has the following form:

$$Q = A_0\left(\frac{\pi}{2} + \tan^{-1}\frac{x}{x_0}\right)e^{-z/z_0}\cos\omega_0 t\,. \quad (9)$$

where A_0 is a constant heating amplitude; x_0 and z_0 are horizontal and vertical length scales of heating, respectively; and ω_0 is the diurnal frequency.

Through a simple horizontal vorticity transformation, these equations can be converted into the following final two-equation form:

$$\left.\begin{aligned} \frac{\partial \eta'}{\partial t} + (\bar{u} + u')\frac{\partial \eta'}{\partial x} + w'\frac{\partial \eta'}{\partial z} + \frac{\partial b'}{\partial x} &= \kappa_u\frac{\partial^2 \eta'}{\partial z^2}, \\ \frac{\partial b'}{\partial t} + (\bar{u} + u')\frac{\partial b'}{\partial x} + w'\frac{\partial b'}{\partial z} + N^2 w' &= Q + \kappa_b\frac{\partial^2 b'}{\partial z^2}. \end{aligned}\right\} \quad (10)$$

In this form, the prognostic model variables are perturbation vorticity (η') and perturbation buoyancy (b'). Perturbations are obtained by partitioning variables as follows:

$$\left.\begin{aligned} &u = \bar{u} + u',\ w = w', \\ &\eta' = \frac{\partial u'}{\partial z}, \text{ and} \\ &b' = g\frac{\theta'}{\theta_0}. \end{aligned}\right\} \quad (11)$$

Winds u' and w' are diagnosed through a streamfunction in the following manner:

$$\left.\begin{aligned} &u' = \frac{\partial \psi'}{\partial z};\ w' = -\frac{\partial \psi'}{\partial x};\ \text{where} \\ &\psi = \bar{\psi} + \psi' \ \text{ is the streamfunction.} \end{aligned}\right\} \quad (12)$$

The numerical model is based on the equation set (10) - (12) has the following properties:

(i) A leapfrog advection scheme for both of the variables,
(ii) Crank-Nicholson (implicit trapezoidal) vertical diffusion scheme,
(iii) Second-order and lagged numerical horizontal diffusion scheme, and
(iv) Lagged Rayleigh-damping sponge layers at the sides and the top.

2.3 The Filter

The basic update equation of the Kalman filter is given as

$$\mathbf{x}^{\mathbf{a}} = \mathbf{x}^{\mathbf{b}} + \mathbf{P}^{\mathbf{b}}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{P}^{\mathbf{b}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right)^{-1}\left(\mathbf{y}^{\mathbf{0}} - \mathbf{H}\mathbf{x}^{\mathbf{b}}\right), \quad (13)$$

where **x** is the state variable and superscripts "a" and "b" denote analysis (posterior) and background (prior) quantities, respectively; **H** is the observational operator; **R** is the observational error matrix; $\mathbf{y}^{\mathbf{0}}$ is the observation vector; and $\mathbf{P}^{\mathbf{b}}$ is the flow-dependent background covariance matrix. The ensemble provides an estimate for the background covariance matrix through the sample covariance relationship (Evensen, 1994):

$$\hat{\mathbf{P}}^{\mathbf{b}} = \frac{1}{N_e - 1}\sum_i\left(\mathbf{x}_i^{\mathbf{b}} - \bar{\mathbf{x}}^{\mathbf{b}}\right)\left(\mathbf{x}_i^{\mathbf{b}} - \bar{\mathbf{x}}^{\mathbf{b}}\right)^{\mathbf{T}}. \quad (14)$$

where the hat denotes the estimate of a given quantity.

	Truth	Ensemble
Initial *x* (m)	0.95	1.00
Initial *v* (ms^{-1})	0.85	1.00
Initial μ (s^{-1})	4.00	1.00
Initial *x* variance (m^2)	--	5.00
Initial *v* variance (m^2s^{-2})	--	5.00
Initial μ variance (s^{-2})	--	0.25

Table 1: *Single unknown parameter case* - Initial states for the truth and the ensemble model.

One practical approach to the more complicated basic EnKF is the sequential technique (Snyder and Zhang, 2003) which is an algorithm that assimilates each observation sequentially. In equation (13), consider the following sample estimate of the term $\mathbf{P^bH^T}$ [similar to equation (14)]:

$$\hat{\mathbf{P}}^{\mathbf{b}}\mathbf{H}^{\mathbf{T}} = \frac{1}{N_e - 1}\sum_i \left(\mathbf{x}_i^{\mathbf{b}} - \overline{\mathbf{x}}^{\mathbf{b}}\right)\left(\mathbf{H}\mathbf{x}_i^{\mathbf{b}} - \mathbf{H}\overline{\mathbf{x}}^{\mathbf{b}}\right)^{\mathbf{T}}. \qquad (15)$$

For a single observation, this matrix becomes a column vector $\hat{\mathbf{c}}$ while the estimate of the term ($\mathbf{HP^bH}^T$ + $\mathbf{R}$) becomes a scalar $\hat{s}$.

One final modification to the original version of the EnKF is implemented through the ensemble square-root filter (EnSRF) that was introduced by Whitaker and Hamill (2002). The idea behind this formulation was to modify the update equation in such a way to eliminate the necessity to perturb observations. In this formulation, updating of the ensemble mean is carried out in the traditional manner while the perturbations are updated with the following formula:

$$\left.\begin{aligned} \tilde{\mathbf{x}}_i^{\mathbf{a}} = \mathbf{x}_i^{\mathbf{a}} - \overline{\mathbf{x}}^{\mathbf{a}} = \left[\mathbf{I} - \beta\left(\frac{\hat{\mathbf{c}}}{\hat{s}}\right)\mathbf{H}\right]\left(\mathbf{x}_i^{\mathbf{b}} - \overline{\mathbf{x}}^{\mathbf{b}}\right), \\ \text{where } \beta = \left\{1 + \sqrt{\frac{\mathbf{R}}{\hat{s}}}\right\}^{-1}. \end{aligned}\right\} \qquad (16)$$

For parameter estimation experiments described in this paper, while the original form of the EnKF has been used for the harmonic oscillator system, the modified sequential square-root version of the EnKF has been implemented for the two-dimensional sea breeze model.

3. RESULTS

3.1 Parameter Estimation For The Harmonic Oscillator System

Parameter estimation experiments for the stochastically-forced harmonic oscillator system have produced very promising results. All of the experiments described in this section had a fixed member number of 50. The model was integrated with a time step of 0.01s for about 4000 time steps. Every result presented corresponds to an outcome averaged over 4 randomized runs.

One important aspect of parameter estimation is the fact that while parameters can be treated as model variables during the analysis step, they remain unchanged between analysis steps during model integration. The result of this is that the ensemble variance itself remains unchanged during model integration while consistently decreasing at every analysis step. This situation typically results in a phenomenon commonly known as *filter divergence*: as the parameter variance becomes smaller, the filter begins to "trust" the parameter estimate more and makes only minor adjustments to the background quantity at later analysis steps. Another important complication related to a decaying parameter variance is that there exists a possibility for the observation/parameter correlations to become exceedingly small relative to the inherent sampling error associated with a finite ensemble. This paper briefly addresses the filter divergence issue while sampling error is left to future investigation.

Two sets of experiments were performed to demonstrate the effects of filter divergence on the estimation of a parameter. In the first case, the parameter variance was left unmodified throughout the experiment, while in the second, parameter variance was inflated back to its initial value following every analysis step. These two ways to treat parameter variance represent two extremes of a variety of possibilities that can be implemented in more realistic applications. For comparison purposes, the results from both types of variance treatment are shown here.

For the first test, in addition to imperfect knowledge of the initial mode state, model error is assumed to originate from a single unknown parameter (the viscosity μ). Initial values of the state for the "truth" and the ensemble that have been used for this test are summarized in Table 1 ("truth" represents an independent model run with specified properties) and the results are shown in Figure 1. As explained previously, two different experiments were performed. In the first case, parameter variance is left unmodified

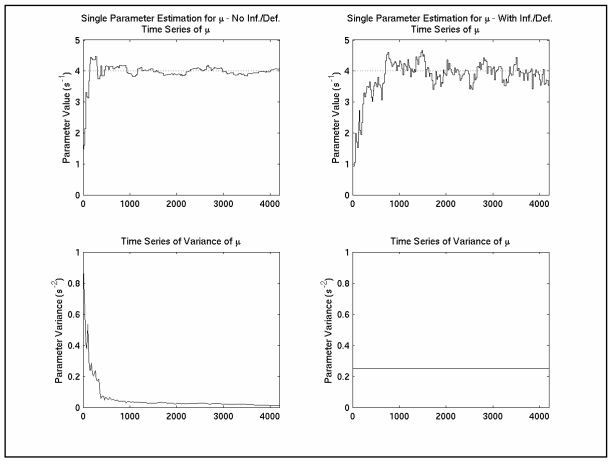

Figure 1: *Constant unknown viscosity case* – Time series of the parameter viscosity and its variance. The results with the unmodified variance experiment are shown in the left panel while the results with the inflated variance are shown in the right panel.

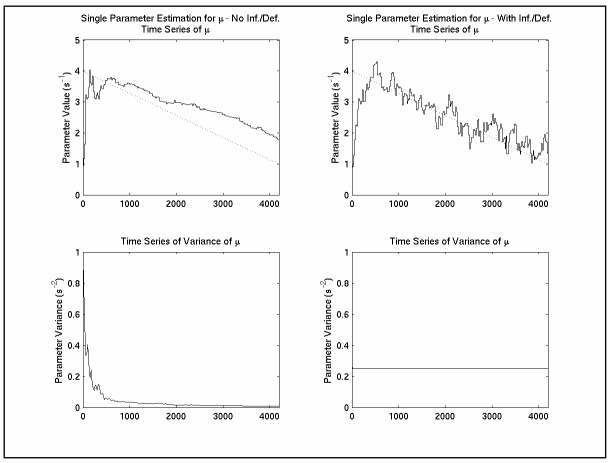

Figure 2: *Time-dependent unknown viscosity case* – Time series of the parameter viscosity and its variance. The results with the unmodified variance experiment are shown in the left panel while the results with the inflated variance are shown in the right panel.

throughout the experiment. The results can be seen in the left panels. We see that the filter quickly corrects the parameter value toward its true value (top panel) while the variance converges toward zero (bottom panel). In the second case, parameter variance is inflated at every analysis step, the results of which are shown in the right panels. Similar to the first case, parameter correction is quite satisfactory yet response is slower. By looking at these results, it may seem that the usefulness of a variance inflation technique may be questionable, yet the issue of filter divergence becomes much more apparent when the parameter is allowed to vary in time. A second test is thus carried out for the time-dependent parameter case. For this purpose, the true viscosity value is let to decrease linearly in time from a value of $4s^{-1}$ to $1s^{-1}$ while the ensemble mean value is initiated at $1s^{-1}$. The results of these experiments are summarized in Figure 2. Similar to the format of Figure 1, left panels and right panels show results with unmodified and inflated parameter variance, respectively, while top panels and bottom panels display time series of the parameter and its variance, respectively. Signs of filter divergence are clearly visible for the unmodified variance tests. We see that while the filter is able to update the parameter value quite well initially, as the variance begins to decrease, update magnitudes become smaller and the mean diverges from the true value. The usefulness of variance inflation is very evident at this point. When parameter variance is not allowed to decrease through an artificial inflation technique, enough ensemble spread is maintained at all times to accommodate for the time-varying character of the parameter. It is also interesting to note that parameter response in the inflated variance case is now comparable to the unmodified variance case. It is believed that a technique that combines the two extreme approaches in a more efficient manner, perhaps by alternating them at certain time steps or even by constructing an on-line monitoring scheme to detect filter divergence and adjust variance accordingly, has the potential to improve the parameter estimate even further. One alternative approach could be to include an autoregressive model for the evolution of the parameter. Such a model would be able to grow the variance of the parameter in the absence of observations, while a combination of observations and a strong correlation with the parameter would result in the decrease of the parameter variance.

	Truth	Ensemble
Initial *x* (m)	0.95	1.00
Initial *v* (ms^{-1})	0.85	1.00
Initial *μ* (s^{-1})	4.00	1.00
Initial *A* (ms^{-2})	750.00	500.00
Initial *x* variance (m^2)	--	5.00
Initial *v* variance (m^2s^{-2})	--	5.00
Initial *μ* variance (s^{-2})	--	0.25
Initial *A* variance (m^2s^{-4})	--	10000.00

Table 2: *Two unknown parameter case* - Initial states for the truth and the ensemble model.

Experiments with two unknown parameters have produced very similar results to the experiments with a single unknown parameter. Only results from experiments with constant true parameters and variance inflation are presented here. Time-dependent parameter estimation results are similar to the ones of constant-parameter estimation. For this purpose, in addition to viscosity, forcing amplitude is assumed to be unknown. Table 2 summarizes initial values of the "truth" and ensemble states used for the experiment. The results of these experiments, in terms of time series, are presented in Figure 3. We see that the filter is able to update both of the parameter values quickly toward their true values.

A final result that will be shown here pertains to the concept of model error treatment. The idea is whether, in a situation when it is not feasible to know or estimate model parameters specifically, it is possible to

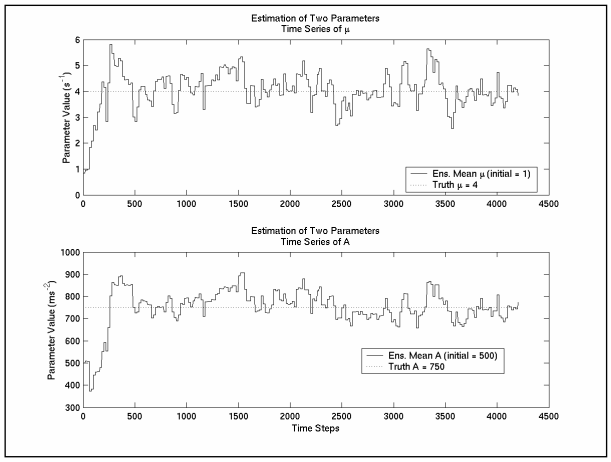

Figure 3: *Two constant unknown parameter case* – Time series of viscosity (top panel) and forcing amplitude (bottom panel).

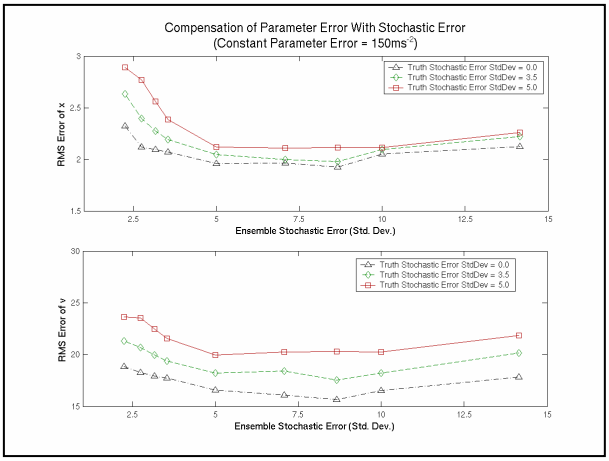

Figure 4: *Compensation for parameter error with stochastic error* – Root-mean square error of model variables x (top panel) and v (bottom panel) as a function of stochastic error magnitude (standard deviation, ms^{-2}) of forcing amplitude in the presence of a forcing amplitude error of $150ms^{-2}$. Different lines represent differing magnitudes of truth stochastic error.

compensate for model error by "simulating" it with stochastic error. The setup of our experiments is ideal for such a test. For this purpose, the magnitude of the forcing amplitude had a fixed $150ms^{-2}$ difference between the truth and forecast runs (thus resulting in a parameter error). The forecast run is designed so that only model variables *x* and *v* are updated by the EnKF (hence parameter error does not change during the forecast run). A stochastic error is then introduced to the forcing. Similar to previous experiments, only model variable *x* is observed. Figure 4 shows the results and consists of two plots; the root-mean square error of model variable *x* (top panel) and model variable *v* (bottom panel) as a function of the magnitude of forecast stochastic error (in terms of its standard deviation). The experiments are performed for multiple values of truth stochastic error which is reflected by the lines with differing styles. The immediate result that strikes the eye is the fact that introduction of a small stochastic error to the forecast run actually reduces the root-mean square error of both of the model variables which is an indication that stochastic error does have the ability to compensate for parameter error. Another result worth mentioning here is that with increasing stochastic error in the truth, relative improvements in the model variable estimates become more pronounced.

3.2 Preliminary Results From Two-Dimensional Sea-Breeze Model Runs

The sea breeze model described here and the associated sequential square-root EnKF are still in their developmental stage. As of writing of this manuscript, the model was producing stable results while the EnKF was still being tuned for best performance. As a result, in this section, some very preliminary results are presented.

Time Interval (s)	150
Horizontal Grid Spacing (m)	4000
Vertical Grid Spacing (m)	50
Mean Horizontal Wind (ms^{-1})	0
Mean Static Stability (s^{-1})	10^{-4}
Coefficients of Vertical Diffusion (m^2 for Vorticity, ms for Buoyancy)	0.25 (none for linear)
Heating Amplitude (ms^{-3})	$5x10^{-7}$ – linear $5x10^{-6}$ – nonlinear
Horizontal Length Scale of Heating (m)	10000
Vertical Length Scale of Heating (m)	500
Time of maximum heating	18:00Z

Table 3: *Key properties of sea-breeze model runs* .

To begin with, a comparison between nonlinear and linear versions of the model is demonstrated. Both of the experiments were run using the same set of model parameters with the exceptions that, for the linear run, (a) the nonlinear terms in model integration were explicitly turned off, (b) coefficients of vertical diffusion were set to zero, and (c) heating amplitude is weaker (as this is known to contribute to nonlinearity). Some of the key properties of these model runs are summarized in Table 3 while key model variables at 76 hour into model integration are plotted in Figure 5 (perturbation vorticity and buoyancy) and Figure 6 (perturbation winds and streamfunction). In both of the figures, linear results are shown in top panels while nonlinear results are shown in bottom ones.

Our linear run produced results comparable to Rotunno's (1983) analytical solution for latitudes less than 30°. The flow response is in the form of internal-inertial waves that extend to infinity along shallow rays emanating from the coastline. At plotted model time (76 hours or a 00:00Z phase), buoyancy is at its peak on land (by a 6-hour lag after maximum heating) and surface winds at the coastline are onshore depicting the onset of a sea-breeze scenario (the peak phase of the sea breeze occurs roughly 2 hours after the peak phase of buoyancy which is not shown here). The addition of nonlinearity introduces interesting variations to the linear response. The most prominent feature is the sea-breeze front that is in the form of a shallow "shock wave" – a strong gradient in both vorticity and buoyancy fields at the location of the front. We also see that the front has penetrated inland. Wind response is also in accordance with the frontal structure: strong convergence is present at the surface and lifting occurs along the front. Outside the frontal zone, the flow no longer fully exhibits the uniform linear internal-inertial wave structure but contains higher-harmonic signature typical for nonlinear sea breeze circulations.

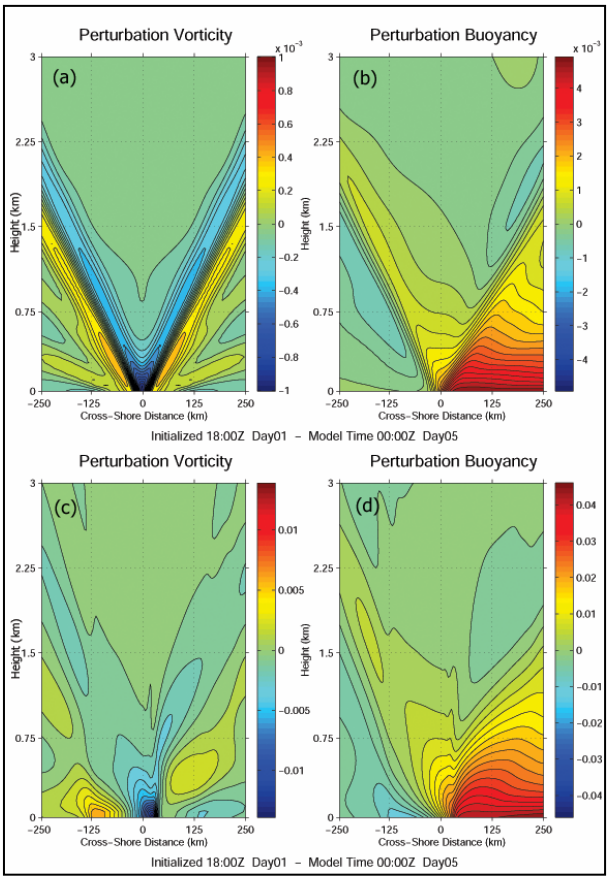

Figure 5: *Outputs from the sea-breeze model: Vorticity and Buoyancy* - Top panels are from the linear run while bottom panels are from the nonlinear run. The coastline is at $x = 0$ with $x > 0$ representing the land and $x < 0$ representing the sea. Output is plotted at 78 hours into model integration.

A preliminary test with the ensemble Kalman filter has also produced promising results. In this test, one analysis cycle has been performed after a 12-hour model run using 20 ensemble members and simulated buoyancy observations placed at 20km spacing in the horizontal and 250m in the vertical. Initial fields were obtained by adding random perturbations on a realization taken from an independent run. Figure 7 shows the prior and posterior distribution of the difference between mean and true buoyancy for the analysis step at 12 hours. We see that there is a domain-wide reduction of buoyancy error (measured as the difference of the ensemble mean from the truth). At this time step, the domain-averaged root mean square error drops from $0.048 ms^{-2}$ to $0.015 ms^{-2}$. One peculiarity about the results is that some noise is introduced at the surface, especially around the sea-breeze front where phase difference between the mean forecast field and the truth field is most pronounced. Although this noise does not influence the domain-wide performance of the filter, it does introduce some unbalanced structure, especially in the vicinity of the frontal zone, that could influence the propagation of the front in later model steps. This phenomenon is currently being investigated in terms of the covariance structure that leads to it and later model response.

4. FUTURE WORK

Preliminary evaluations of the EnKF for parameter estimation have been very promising and resulted in a variety of new ideas to be explored in the near future. Our plans are many-fold. As mentioned in previous sections, the development of the sea-breeze model and its associated EnKF is not yet fully completed. In this regard, further investigation of filter performance utilizing the covariance information produced by the ensemble is the next logical step. Issues related to the noise introduced by the filter will be scrutinized and the filter will be fine-tuned to eliminate these problems.

Application of the EnKF to the sea-breeze problem offers numerous paths of analysis that can potentially be followed. While our main focus will be on parameter estimation and model error, the covariance structure produced by the ensemble is also planned to be analyzed to improve our understanding of error growth mechanisms and predictability of thermally-induced local circulations.

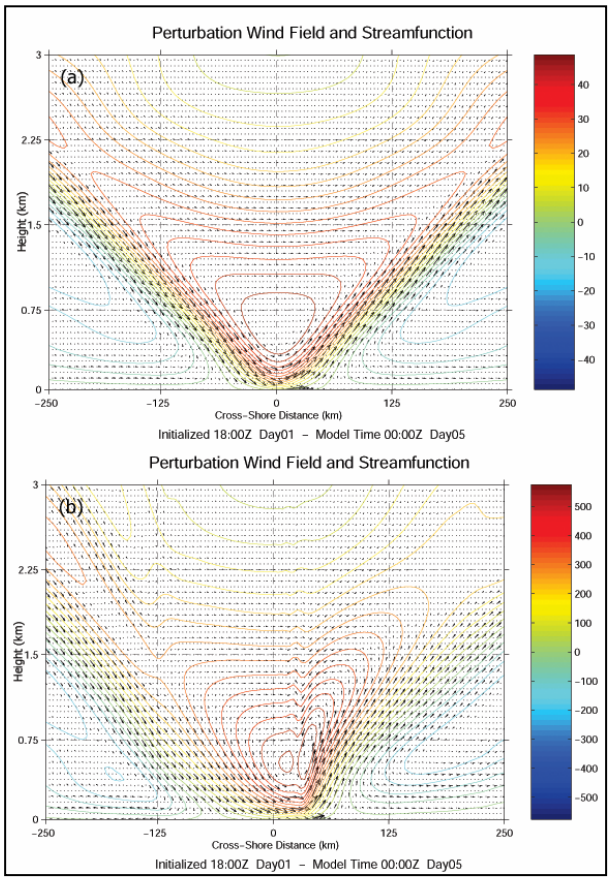

Figure 6: *Outputs from the sea-breeze model: Winds and Streamfunction* - Top panels are from the linear run while bottom panels are from the nonlinear run. The coastline is at $x = 0$ with $x > 0$ representing the land and $x < 0$ representing the sea. Output is plotted at 78 hours into model integration.

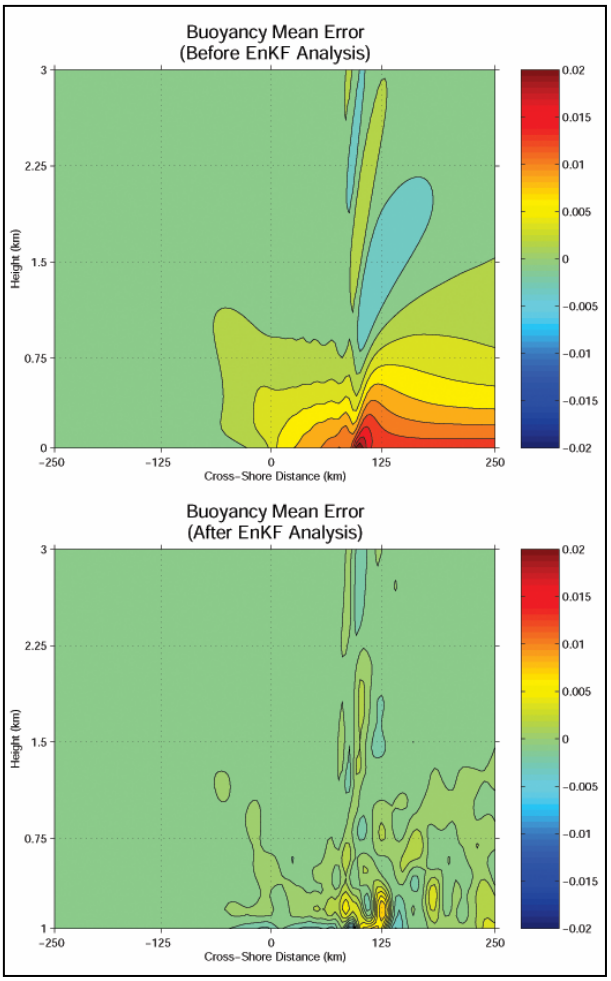

Figure 7: *Prior (a) and posterior (b) distribution of mean buoyancy error* – Analysis performed after a 12-hour model run.

Our ultimate goal is to apply the experience and information obtained from experiments with the harmonic oscillator and the two-dimensional sea-breeze model to a more complicated operational environment and study parameter estimation using MM5 in order to improve both our forecasting skill for local sea-breeze type of circulations and our understanding of model error behavior.

5. ACKNOWLEDGEMENTS

We thank to Lee Panetta for drawing our attention to the similarities between the dynamics of the sea breeze and the harmonic oscillator.

6. REFERENCES

Anderson, J.L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, 129, 2884-2903.

Anthes, R.A., 1978: The height of the planetary boundary layer and the production of circulation in a sea breeze model. *J. Atmos. Sci.*, 35, 1231-1239.

Atkins, N.T., R.M. Wakimoto, and T.M. Weckwerth, 1995: Observations of the sea-breeze front during CaPE. Part II: Dual-Doppler and aircraft analysis. *Mon. Wea. Rev.*, 123, 944-969.

Burgers, G., P.J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 126, 1719-1724.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 (C5), 10143-10162.

Gelb, A. (editor), 1974: Applied Optimal Estimation. The M.I.T. Press, Cambridge, Massachusetts, pp. 348-356.

Houtekamer, P.L. and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126, 796-811.

Houtekamer, P.L. and H. Mitchell, 2001: A sequential ensemble Kalman filter for data assimilation. *Mon. Wea. Rev.*, 129, 123-137.

Miller, S.T.K., B.D. Keim, R.W. Talbot, and H. Mao, 2003: Sea Breeze: Structure, Forecasting, and Impacts. *Rev. Geophys.*, 41 (3), 1011.

Mitchell, H. and P.L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, 128, 416-433.

Nielsen-Gammon, J.W., 2001: The subtropical sea breeze. *Preprints, 14th Conf. on Numerical Weather Prediction*.

Rotunno, R., 1983: On the linear theory of land and sea breeze. *J. Atmos. Sci.*, 40, 1999-2009.

Simpson, J.E., 1994: Sea breeze and local wind. Cambridge University Press, Cambridge, Great Britain.

Snyder, C., F. Zhang, J. Sun, and A. Crook, 2001: Tests of an ensemble Kalman filter at convective scales. *Preprints, 14th Conf. on Numerical Weather Prediction*, 444-446.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, 131, 1663-1677.

Wakimoto, R.M. and N.T. Atkins, 1994: Observations of the sea-breeze front during CaPE. Part I: Single-Doppler radar, satellite, and cloud photogrammetry analysis. *Mon. Wea. Rev.*, 122, 1092-1114.

Walsh, J.E., 1974: Sea breeze theory and applications. *J. Atmos. Sci.*, 31, 2012-2026.

Whitaker, J.S. and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130, 1913-1924.

Zhang, F., C. Snyder, and J. Sun, 2002: Impacts of initial analyses and observations on the convective-scale data assimilation with an ensemble-Kalman filter. *Preprints, 15th Conf. on Numerical Weather Prediction*, 296-299.

Zhang, F., C. Snyder, and J. Sun, 2003: Impacts of initial estimate and observations on convective-scale data assimilation with an ensemble-Kalman filter. *Mon. Wea. Rev.*, accepted.