Regional Ensemble Forecasts Using the Ensemble Transform Technique

CRAIG H. BISHOP, TEDDY R. HOLT, JASON NACHAMKIN, SUE CHEN, JUSTIN G. MCLAY, JAMES D. DOYLE, AND WILLIAM T. THOMPSON

Naval Research Laboratory, Monterey, California

(Manuscript received 15 February 2008, in final form 31 May 2008)

ABSTRACT

A computationally inexpensive ensemble transform (ET) method for generating high-resolution initial perturbations for regional ensemble forecasts is introduced. The method provides initial perturbations that (i) have an initial variance consistent with the best available estimates of initial condition error variance, (ii) are dynamically conditioned by a process similar to that used in the breeding technique, (iii) add to zero at the initial time, (iv) are quasi-orthogonal and equally likely, and (v) partially respect mesoscale balance constraints by ensuring that each initial perturbation is a linear sum of forecast perturbations from the preceding forecast. The technique is tested using estimates of analysis error variance from the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System (NAVDAS) and the Navy's regional Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS) over a 3-week period during the summer of 2005. Lateral boundary conditions are provided by a global ET ensemble. The tests show that the ET regional ensemble has a skillful mean and a useful spread–skill relationship in mass, momentum, and precipitation variables. Diagnostics indicate that ensemble variance was close to, but probably a little less than, the forecast error variance for wind and temperature variables, while precipitation ensemble variance was significantly smaller than precipitation forecast error variance.

1. Introduction

Edward Lorenz's (1969) seminal work on the predictability of flows that possess many scales of motion shattered the notion that perfect weather forecasts could be obtained if only we could accurately solve key equations and better observe the atmosphere. The fact that moist convection amplifies miniscule errors to large amplitudes in a matter of hours (Zhang et al. 2003) while nonlinearity transfers these errors to larger scales means that, at best, accurate observations and accurate solutions to the governing equations define a *distribution* of future states. To obtain the distribution of truth given error-prone initial conditions, an error-prone forecast model, and error-prone boundary conditions, one needs to account for the effects of these uncertainties on forecast errors. The focus of this paper is the problem of accounting for the forecast uncertainty due to the initial condition uncertainty in limited-area models of the atmosphere.

A distinguishing requirement of ensemble perturbations in variable resolution models is that the scale of the perturbations needs to vary in accordance with the scales of variability resolved by the model. Methods of creating initial perturbations for limited-area models that are based on ensembles of global model states (Grimit and Mass 2002; Marsigli et al. 2005; Walser et al. 2006) are a priori incapable of meeting this requirement. Limited-area adjoint approaches (Xu et al. 2001; Homar et al. 2006) have not included resolution-dependent initial condition perturbations but easily could if the adjoint of a nested model was available. The breeding ensemble generation technique (Toth and Kalnay 1993, 1997) when applied to the mesoscale (Stensrud et al. 1999) creates mesoscale analysis perturbations from mesoscale forecast perturbations and hence, in principle, provides initial condition perturbations at all scales resolved by the limited-area model (LAM). However, as pointed out by Wang and Bishop (2003), the covariance matrices associated with breeding ensembles have an unrealistically small number of significant eigenvalues. Recasting breeding within the con-

Corresponding author address: Craig H. Bishop, Naval Research Laboratory, Marine Meteorology Division, 7 Grace Hopper Ave., Stop 2, Building 702, Room 212, Monterey, CA 93943-5502.
E-mail: bishop@nrlmry.navy.mil

DOI: 10.1175/2008MWR2559.1

text of an ensemble transform Kalman filter (ETKF) ensemble generation technique, removes this gross deficiency of breeding (Wang and Bishop 2003). The ETKF ensemble generation technique would be optimal if the ensemble perfectly described forecast error covariances and the data assimilation scheme used ensemble covariances as its model of forecast error covariance. When neither of these conditions are met an ad hoc method is required to ensure that the ensemble variance is consistent with differences between observations and forecasts.

The Naval Research Laboratory (NRL) Atmospheric Variational Data Analysis System (NAVDAS) provides estimates of the error variance of the analyses it produces (Daley and Barker 2001). Bishop and Toth (1999) showed how such variance estimates can be used to constrain the magnitude of initial perturbations that represent *transformations* or linear combinations of ensemble forecast perturbations. We call the initial perturbations obtained in this way, ensemble transform (ET) perturbations. Because each ET analysis perturbation is just a particular sum of forecast perturbations, each analysis perturbation contains all of the scales resolved by the forecast model and hence is well suited to variable resolution models. Here, for the first time, we document the performance of an ET ensemble in a LAM.

The LAM model employed for this study is the Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS;[1] Hodur 1997) while the ensemble of lateral boundary conditions required for the LAM ensemble is obtained from the global ET ensemble described in McLay et al. (2008).

Section 2 briefly reviews the theoretical basis and properties of the ET technique. Section 3 describes a 3-week test of the ET LAM ensemble's ability to produce a skillful mean and a useful spread–skill relationship. Conclusions follow in section 4.

2. The ET technique

There are a variety of techniques by which one can use estimates of forecast error covariance and observation error covariance to estimate the covariance of the errors of an analysis (Fisher and Courtier 1995; Cohn 1997; Daley and Barker 2001). These estimates are generally based on crude approximations to the known equations for analysis error covariance because the unapproximated equations are prohibitively expensive to solve given the large numbers of observations in today's observational network. For example, NAVDAS produces a flow-*independent* estimate $\mathbf{B}^a$ of the covariance matrix of errors $\varepsilon^a = \mathbf{x}^a - \mathbf{x}^t$ of the n-vector analysis $\mathbf{x}^a$ ($\mathbf{x}^t$ is the truth) using (i) its flow-independent forecast error covariance matrix $\mathbf{B}^f$, (ii) its estimate of the observation error covariance matrix $\mathbf{R}$, and (iii) the observation operator matrix $\mathbf{H}$. These matrices are used in an approximation to the following equation:

$$\mathbf{B}^a = \mathbf{B}^f - \mathbf{B}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}^f, \tag{1}$$

for the analysis error covariance of the minimum error variance estimate that is based on contiguous observation volumes. This approximation involves dividing the domain into subdomains called *observation volumes* and then assuming that the forecast errors of variables within each observation volume are uncorrelated with forecast error variables outside of the observation volume. Furthermore, to save storage costs only the diagonal components of $\mathbf{B}^a$ are computed. To ameliorate edge effects, the computation is performed on two distinct contiguous sets of observation volumes that are constructed so that one set of observation volumes overlaps the edges of the other and vice versa. The results from each set are then averaged. The estimate of $\mathbf{B}^a$ provided by NAVDAS is useful in that it represents an estimate of analysis error variance that is consistent with the assumptions of the data assimilation scheme that produced the analysis, but it is inadequate in that (i) it is not flow dependent, and (ii) it is not multivariate and, hence, does not respect quasigeostrophic and other types of balance.

Now consider the properties of an ensemble comprised of perturbations $\mathbf{x}_j'^a$ that are linear combinations or transformations of K ensemble forecast perturbations $\mathbf{x}_i'^f = (\mathbf{x}_i^f - (1/K)\Sigma_{k=1}^K \mathbf{x}_k^f)$ such that $\mathbf{x}_j'^a = \Sigma_{i=1}^K \mathbf{x}_i'^f t_{ij}$; in other words,

$$\mathbf{X}^a = \mathbf{X}^f\mathbf{T}, \tag{2}$$

where $\mathbf{X}^a = [\mathbf{x}_1'^a, \mathbf{x}_2'^a, \ldots, \mathbf{x}_K'^a]$, $\mathbf{X}^f = [\mathbf{x}_1'^f, \mathbf{x}_2'^f, \ldots, \mathbf{x}_K'^f]$ and $(\mathbf{T})_{ij} = t_{ij}$ is the $K \times K$ transformation matrix. Note that $\mathbf{x}_j'^a$ and $\mathbf{x}_j'^f$ are both n vectors listing all of the variables defining the model state. It has long been known that a property of numerical simulations of the weather is that they adjust to "dynamically balanced" states as the simulation proceeds (Errico et al. 1988). Typically, the domain-averaged square of the surface pressure tendency undergoes significant fluctuations in the first few hours of the simulation, but these fluctuations disappear after about 6 h. Thus, to the extent that the members of a 6-h ensemble forecast are dynamically balanced, their perturbations represent deviations about the mean of a set of dynamically balanced perturbations. Now suppose one linearized a balance equation

[1] COAMPS is a registered trademark of the Naval Research Laboratory.

about a member close to the mean then, provided the forecast perturbations were small enough,[2] they would each be solutions to this linearized balance equations. Furthermore, any linear combination of the perturbations would be a solution to the linearized balance equation. Each perturbation about the ensemble mean is a linear combination of perturbations about a member close to the mean. Hence, to the extent that the aforementioned assumptions are satisfied, any covariance model based on transformed forecast perturbations will respect dynamical balance. This means that if we can find a transformation matrix $\mathbf{T}$ that provides transformed perturbations that have variance characteristics similar to the approximate estimate of the analysis error variance $\mathbf{B}^a$ then a model of analysis error covariance will be obtained that respects dynamical balance and the analysis error variance constraints provided by the data assimilation scheme. As shown in McLay et al. (2008), the required transformation matrix is given by

$$\mathbf{T} = \mathbf{C}\boldsymbol{\Gamma}^{-1/2}\mathbf{C}^{\mathrm{T}}, \tag{3}$$

where $\mathbf{C}$ is a $K \times K$ orthonormal matrix such that

$$\frac{\mathbf{X}^{fT}(\mathbf{B}^a)^{-1}\mathbf{X}^f}{n} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{\mathrm{T}} \tag{4}$$

gives the eigenvector decomposition of $\mathbf{X}^{fT}(\mathbf{B}^a)^{-1}\mathbf{X}^f/n$. (Recall that n is the number of state variables.) The diagonal matrix $\boldsymbol{\Lambda}$ lists the eigenvalues corresponding to the orthonormal eigenvectors listed in the columns of $\mathbf{C}$. One of these eigenvalues must be equal to zero because the sum of the forecast perturbations is zero. The diagonal matrix $\boldsymbol{\Gamma}$ in (3) is identical to the eigenvalue matrix $\boldsymbol{\Lambda}$ except that the zero eigenvalue has been replaced by the number 1. Hence, $\boldsymbol{\Gamma}$ has an inverse, but $\boldsymbol{\Lambda}$ does not (see McLay et al. 2008 for the justification for this replacement).

With Eqs. (3) and (4) in hand, the steps to creating an ET ensemble are as follows:

1) Create K small linearly independent initial perturbations that sum to zero. (In our case, these initial perturbations were the global model ET initial perturbations interpolated to the regional model's grid.) Add each of these to some initial condition to create K-independent initial conditions and then perform K-independent forecasts over a period long enough for each member to reach dynamical balance.
2) Compute the mean and perturbations $\mathbf{X}^f$ of the ensemble forecast generated by step 1 at the time of the next analysis.
3) Retrieve the analysis $\mathbf{x}^a$ and the estimate $\mathbf{B}^a$ of its error variance from the data assimilation scheme.
4) Compute the $K \times K$ matrix $\mathbf{X}^{f\,\mathrm{T}}(\mathbf{B}^a)^{-1}\mathbf{X}^f/n$.
5) Identify its eigenvalues $\boldsymbol{\Lambda}$ and eigenvectors $\mathbf{C}$ as indicated by Eq. (4). Create $\boldsymbol{\Gamma}$ by replacing the zero eigenvalue in $\boldsymbol{\Lambda}$ by 1.
6) Create the transformation matrix $\mathbf{T}$ using Eq. (3) and then use Eq. (2) to create the analysis perturbations $\mathbf{X}^a$.
7) Add the analysis perturbations to $\mathbf{x}^a$ to create a K-member ensemble of analyses. Use these to create a K-member ensemble forecast.
8) While making the K-member ensemble forecast be sure to write out the forecast perturbations at the next analysis time. These are needed to create $\mathbf{X}^f$ for the next ensemble forecast cycle.
9) Repeat step 2.

The fact that the ET creates analysis perturbations from linear combinations of forecast perturbations means that structures that, on average, grow faster than other structures are selected or "bred" in the same manner that the breeding technique creates growing perturbations. Because the model determines the shape of the selected structures, the approach provides a natural means of creating perturbations at the scales resolved by variable resolution LAM. This fact is graphically illustrated in Fig. 1 that compares the structure of initial ET perturbations over Tokyo at 45-km horizontal resolution with corresponding initial ET perturbations at 1.67-km nest horizontal resolution. The figure was generated as part of the (Tokyo) urban dispersion study of Holt et al. (2008) that employed the ET ensemble generation technique using nests down to a horizontal resolution of 1.67 km.

As previously mentioned, an advantage of the ET technique over breeding is that ET perturbations are guaranteed to be quasi-orthogonal at the analysis time (McLay et al. 2008). In addition, ET perturbations are (i) equally likely under an analysis error variance norm; (ii) sum to zero at the initialization time; (iii) as the ensemble size increases, the initial ET perturbations variance better matches the variances in $\mathbf{B}^a$; and (iv) the computations required for the ET are highly scalable (McLay et al. 2008).

3. A 3-week test of an ET LAM ensemble

a. Experimental setup

COAMPS is the Navy's operational LAM. In this study, we will use the atmospheric module of COAMPS

[2] Unfortunately, moisture forecast errors are often too large to be accurately simulated with a linearized equation. This fact limits the thermodynamic balance of the ET initial states.

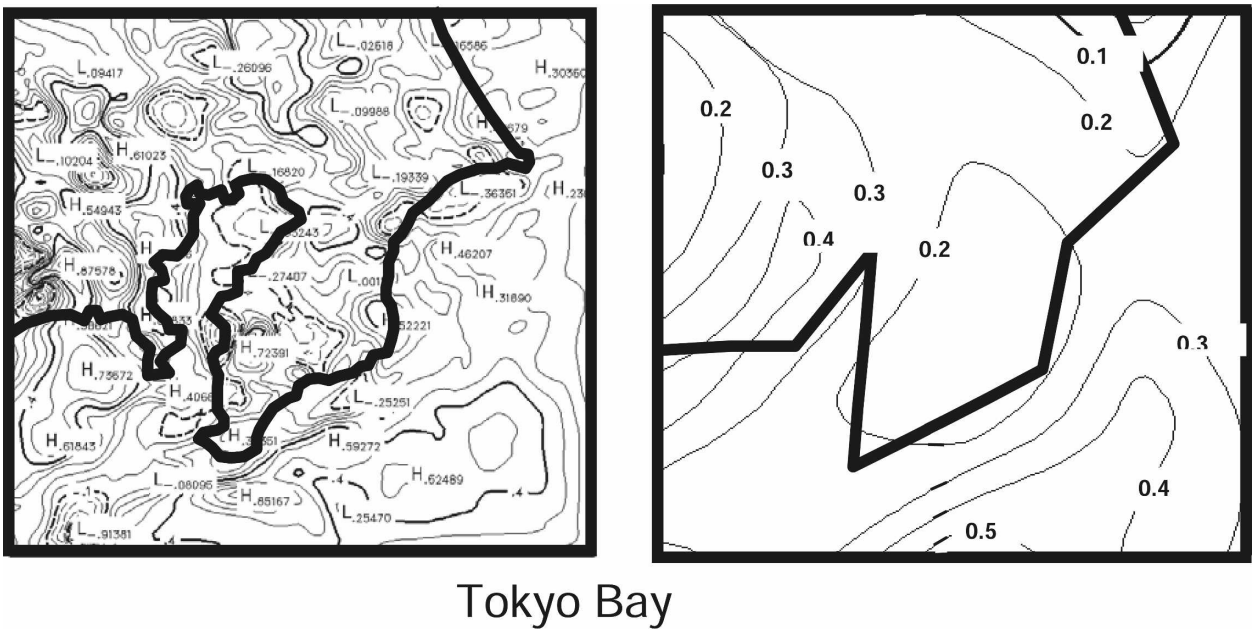

FIG. 1. This figure compares the structure of initial ET perturbations on (left) a 1.67-km horizontal resolution nest with the corresponding perturbation from (right) a 45-km horizontal resolution nest. The contour interval in both is 0.1 K. The domain covers Tokyo Bay, and pertains to the uncertainty of the COAMPS analysis on 1800 UTC 30 Jun 2005. The area of the domain shown is 167×167 km^2.

to examine the properties of an ensemble comprising 1 unperturbed control member and 28 perturbed members. The atmospheric portion of COAMPS is based on a finite-difference approximation to the fully compressible, nonhydrostatic equations and uses a terrain-following vertical coordinate transformation. In this application, the finite difference schemes are of second-order accuracy in time and space. A time-splitting technique with a semi-implicit formulation for the vertical acoustic modes is used to efficiently integrate the compressible equations (Klemp and Wilhelmson 1978). A prognostic equation for the turbulence kinetic energy (TKE) is used to represent the planetary boundary layer and free-atmospheric turbulent mixing and diffusion. The Louis (1979) and Louis et al. (1982) surface layer parameterization, which makes use of a surface energy budget based on the force–restore method, is used to represent the surface fluxes over land and a Coupled Ocean–Atmosphere Response Experiment version 2.6 (COARE2.6; Wang et al. 2002) parameterization is used over ocean. Subgrid-scale moist convection is represented using the Kain and Fritsch (1993) parameterization. The grid-scale evolution of the moist processes is explicitly predicted from budget equations for cloud water, cloud ice, raindrops, snowflakes, graupel, and water vapor (Rutledge and Hobbs 1983, 1984; Meyers et al. 1992; Cooper and Haines 1996; Hallett and Mossop 1974). The short- and longwave radiation processes are represented following Harshvardhan et al. (1987). The domain considered for this study is shown in Fig. 2.

Unfortunately, the LAM version of NAVDAS was not completed in time to be integrated into the initial testing of the mesoscale ensemble system presented here; indeed, for the simulations to be discussed here, a multivariate optimal interpolation (MVOI) scheme was employed. The MVOI analyses of upper-air sounding, surface, commercial aircraft, and satellite observations are quality controlled and blended with the 12-h COAMPS forecast fields (Barker 1992). An incremental update data assimilation procedure is used in conjunction with the MVOI, which enables mesoscale phenomena to be retained in the analysis increment fields. This scheme does not produce an estimate of the error variance of its analyses. However, as it will soon be replaced by a LAM NAVDAS and since its background error covariance model is similar to that of NAVDAS, we chose to crudely approximate its analysis error variance with that of the NAVDAS estimate of analysis error variance on a half-degree latitude–longitude grid with 60 vertical levels for the global model interpolated onto the COAMPS grid.

The lateral boundary conditions for the 29 COAMPS forecasts were provided by 29 forecasts from a T119L30 NOGAPS global ET ensemble described in McLay et al. (2008). Inconsistencies between the temporal evolu-

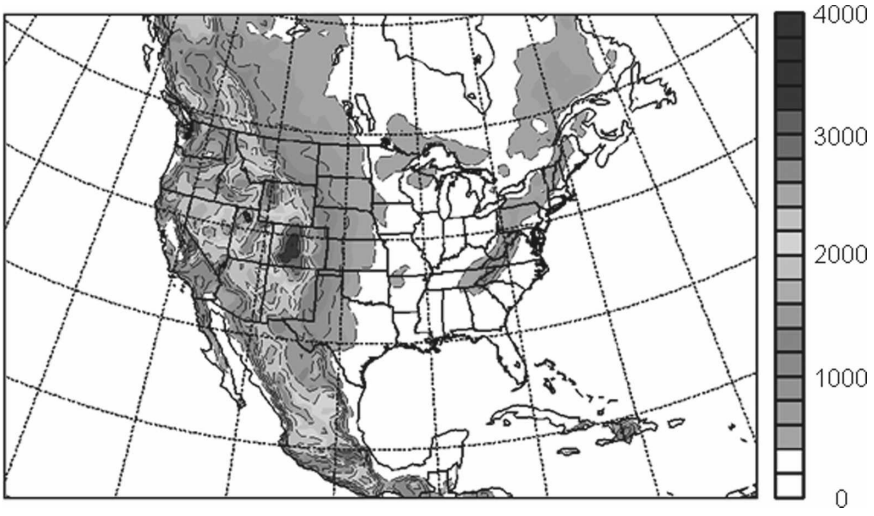

FIG. 2. COAMPS 45-km horizontal model domain (163 × 105 grid points) with terrain (m) shaded.

tion and dynamical balance of NOGAPS fields and COAMPS fields are handled by letting the tendencies of COAMPS fields be a weighted average of NOGAPS and COAMPS tendencies within seven grid points from the lateral boundaries. NOGAPS tendencies receive a weight of 1 at the boundary and decrease to a weight of 0 eight grid points away from the boundary using weights following Davies (1976).

To test the ET ensemble generation technique in a LAM we initialized a 29-member ET LAM ensemble cycle with 29 initial states from a T119L30 global NOGAPS ET ensemble at 0000 UTC 20 June 2005. One of these states (member 0) corresponded to the global analysis while the other 28 states (members 1–28) corresponded to perturbed states. These global model states were interpolated to a 45 km × 45 km horizontal resolution COAMPS model with 163 × 105 grid points in the horizontal and 30 vertical levels. The 6-h forecasts made from these initial states allowed us to complete step 1 of the ET procedure outlined above. To spin up the ET ensemble, steps 2–9 were then repeated using a 6-h analysis–forecast cycle until 0000 UTC 25 June 2005. Throughout the experiment, member 0 was the unperturbed or control forecast. These 5 days of cycling gave ample time for the ET cycle to spin up. From 0000 UTC 25 June 2005 to 1200 UTC 17 July 2005, the 6-h analysis–forecast cycle was continued in the same way except that at 0000 and 1200 UTC 48-h ensemble forecasts were performed. Thus, the 23-day experiment provided 46 ET ensemble forecasts comprised of 29 members each integrated to 48 h.

The time required to transform forecast perturbations into analysis perturbations using Eqs. (2)–(4) (2 min on a single CPU) was insignificant compared with the time required for integrating the forecasts.

b. Results

In this paper, we focus on the mean-square error (MSE) of the ensemble mean and the ensemble's spread–skill relationship. We largely refrain from examining measures that directly assess the extent to which the ensemble members appear to be drawn from the distribution of truth given the ensemble mean because ET perturbations are not constructed to represent this distribution. They are constructed to have a variance that, in a globally averaged sense, matches the analysis error variances listed in $\mathbf{B}^a$. To clearly see that this feature precludes the possibility that ET perturbations can, in general, represent a random sample, note that (as previously shown in McLay et al. 2008), Eqs. (3) and (4) imply that

$$\begin{aligned}\frac{\mathbf{X}^{a\mathrm{T}}(\mathbf{B}^a)^{-1}\mathbf{X}^a}{n} &= \frac{\mathbf{T}^{\mathrm{T}}\mathbf{X}^{f\mathrm{T}}(\mathbf{B}^a)^{-1}\mathbf{X}^f\mathbf{T}}{n}\\ &= \mathbf{C}\boldsymbol{\Gamma}^{-1/2}\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\Gamma}^{-1/2}\mathbf{C}^{\mathrm{T}}\\ &= \mathbf{C}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}\mathbf{C}^{\mathrm{T}} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}}{K}, \end{aligned} \tag{5}$$

where $\mathbf{I}$ is the $K \times K$ identity matrix and all the elements of the K-vector $\mathbf{1}$ are equal to 1. Hence, if the number of model variables was equal to 1 ($n = 1$) and there were just 2 ensemble members ($K = 2$), the square of each analysis perturbation divided by the

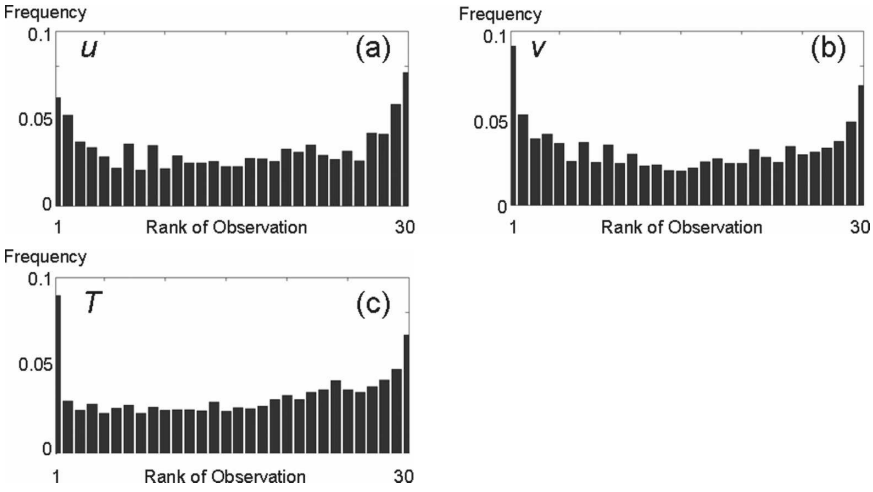

FIG. 3. Frequency rank histograms (Talagrand diagrams) for 850 mb u, v, and T. To account for the effect of radiosonde observation error, random numbers with std dev equal to estimated observation error variance (1.9 m s^{-1} for the zonal and meridional wind components u and v, 1 K for temperature T) were added to each of the forecasts before the histogram was formed (Bowler 2006). The overpopulation of low ranks relative to high ranks shown in Fig. 3c is consistent with a known cold bias of the model at 850 mb. There are 30 ranks because the unperturbed control member was included in the analysis.

analysis error variance $(\mathbf{B}^a)_{11}$ would not be a random variable; it would always be equal to 0.5. [This implies that the sum of the square of the $(K = 2)$ perturbations divided by $(K - 1)$ would be $(\mathbf{B}^a)_{11}$.] To obtain estimates of the distribution of truth given ensemble mean, ensemble postprocessing techniques would need to be applied (Raftery et al. 2005; Wang and Bishop 2005). Even without such postprocessing, Fig. 3 shows that the frequency rank histograms (Hamill 2001) for this ET ensemble at 850 hPa are reasonably uniform given that, apart from bias correction, no attempt has been made to represent the effect of model error on forecast uncertainty. Corresponding histograms for 500 and 300 hPa were found to be similar or slightly flatter than those shown in Fig. 3.

It was found that the mean square distance (MSD) from verifying radiosonde observations and the 48-h ensemble mean forecast of the observation was smaller than that of the unperturbed control forecast in all variables examined (e.g., wind components, temperature, and humidity) and at all vertical levels. The significance of some of these differences was evaluated using parametric hypothesis testing. To do so, the daily (MSD) of the control forecast and of the ensemble-mean forecast were calculated for a given variable for each of the 46 distinct ensemble forecasts made during the test period. The two resulting series of MSD (one for the control forecast and one for the ensemble-mean forecast) each consisted of 46 different values. Figure 4 displays the difference between these two time series for the 300-hPa meridional wind. It shows that in all but 3 of the 46 ensemble forecasts the ensemble mean was closer to verifying radiosonde observations than the control forecast. In considering this diagram, it is worth remembering that it is nonlinearity and nonlinearity alone that causes the ensemble mean to differ from the control forecast. Presumably, fluctuations in the differences of the two MSD time series are partly related to fluctuations in the importance of nonlinear processes in the forecasts.

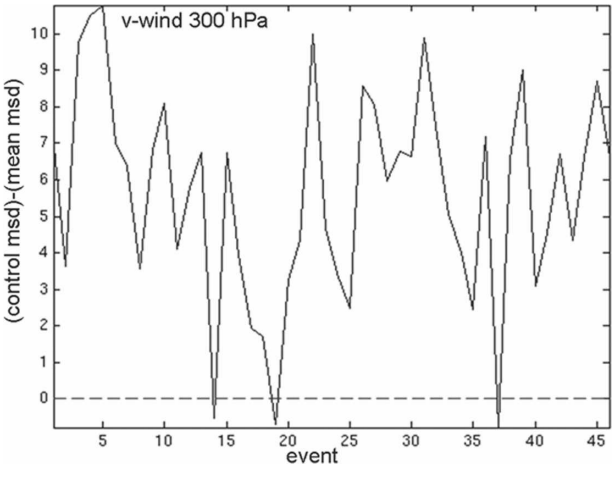

FIG. 4. The MSD of the ensemble mean from radiosondes observations subtracted from the MSD of the control from radiosonde observations for 48-h forecasts of 300-hPa meridional wind as a function of the forecast event.

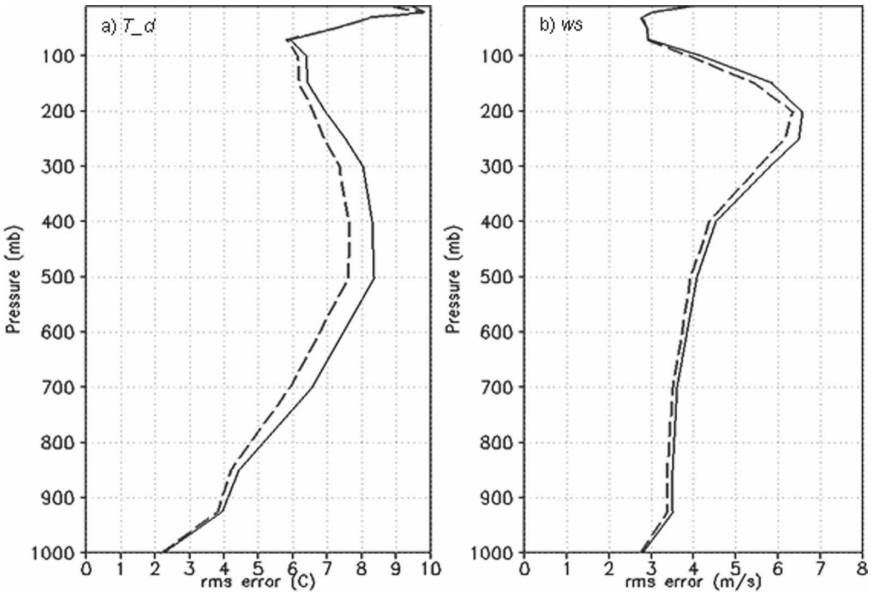

FIG. 5. Vertical profile of RMSD from radiosonde observations for COAMPS control forecast (solid line) and ensemble mean (dashed) of (a) dewpoint temperature (°C) and (b) wind speed (m s^{-1}).

A formal parametric hypothesis test was constructed so as to evaluate whether the average of these 46 differences was significantly greater than 0 (such would be expected if the ensemble-mean forecast has systematically lower MSD than the control forecast). In addition, the methodology of Katz (1982) was implemented to account for possible autocorrelation in the series of MSD differences. Autocorrelation can lead to an overconfident hypothesis test, by producing a spuriously small sample variance and, hence, an overly large test statistic (Wilks 2006). The methodology of Katz (1982) employs autoregression modeling to appropriately adjust the sample variance in the presence of autocorrelation.

The hypothesis testing was applied for seven different 48-h lead-time variables: air temperature at 850 hPa, geopotential height at 200 and 500 hPa, zonal and meridional wind at 300 and 925 hPa, and relative humidity at 700 hPa. For all variables, the null hypothesis that the ensemble mean and control forecasts had the same accuracy as the ensemble mean forecast could be rejected with greater than 99.98% confidence. This indicates that the MSD of the ensemble-mean forecast from verifying observations was significantly less than that of the control forecast, on average. These results are consistent with the fact that the values in the series of MSD differences were nearly always positive for each of the variables. The testing showed that the most significant MSD differences were those associated with the 300-hPa meridional wind, the 700-hPa relative humidity, the 300-hPa zonal wind, and the 850-hPa air temperature, in that order. The root-mean-square (RMS) distance from radiosonde observations as a function of vertical level for the wind speed and dewpoint temperature is shown in Fig. 5. Corresponding diagrams for geopotential height and temperature showed somewhat smaller improvements.

While most ensemble studies have been able to demonstrate ensemble means with smaller RMS errors than the control forecast, many have been unable to demonstrate a significant spread–skill relationship. To be specific, Stensrud et al. (1999) and Leslie and Speer (2000) note the absence of spread–skill relationships in LAMs while Scherrer et al. (2004) and Hamill et al. (2004) comment on the absence of such spread–skill relationships in global models. Grimit and Mass (2007) have also commented on the somewhat disappointing spread–skill results reported in the literature. They explain some of the fundamental reasons for these results drawing on the work of Houtekamer (1993) but then go on to demonstrate that calibration of spread–skill information based on appropriate binning of data (along the lines of Atger 1999) can add significant value to the ensemble forecast.

To test the extent to which large and small ET ensemble variances distinguish large forecast error variance from small forecast error variance in a LAM model, we follow Majumdar et al.'s (2001) binning approach. The expected slope and ordinate inter-

cept of the spread–skill curves that would be obtained from this binning approach for a perfect ensemble forecasting system and an infinite amount of verification data can be understood from the following detailed description of how the spread–skill curves are computed.

1) Choose a variable of interest; such as the zonal component of the wind u at 850 mb.
2) Estimate the bias of each forecast and remove it. (For simplicity, here we assumed that the bias was a function of lead time only and not a function of observation location or of ensemble member)
3) For each verifying radiosonde observation y_u of 850 mb u, compute the difference $y_u - \overline{x}^f_u$ between it and the 48-h bias-corrected ensemble mean forecast $\overline{x}^f_u$ of the observation of u at 850 mb ($\overline{x}^f_u = (1/K)\Sigma_{j=1}^K x^f_{uj}$).
4) Compute the ensemble sample variance $s^2 = [(1/K - 1)\Sigma_{i=1}^K (x^f_{ui} - (1/K)\Sigma_{j=1}^K x^f_{uj})^2]$ for each of the 850-mb u observations.
5) Form the data pair $[s^2, (y_u - x^f_{u0})^2]$ for each observation site where x^f_{u0} denotes a key reference forecast such as the ensemble mean or the control forecast. Note that while s^2 is a sample variance, $(y_u - x^f_{u0})^2$ is just a squared innovation (i.e., they are different statistical quantities).
6) Order the data pairs from smallest ensemble variance to largest ensemble variance.
7) Group the ordered data pairs into m equally populated bins.
8) Compute mean sample variance and MSE within each bin to form the m data pairs $[\overline{(s^2)}_i, \overline{[(y_u - x^f_{u0})^2]}_i]$, $i = 1, 2, \ldots, m$. Note that if y^t_u represents the true state of the zonal wind and the angle bracket equals the expectation operator then

$$\langle \overline{[(y_u - x^f_{u0})^2]}_i \rangle = \langle \overline{\{[(y_u - y^t_u) - (x^f_{u0} - y^t_u)]^2\}}_i \rangle = \langle \overline{[(\varepsilon^o - \varepsilon^f)^2]}_i \rangle = \overline{[\langle \varepsilon^{o2} \rangle - 2\langle \varepsilon^o \varepsilon^f \rangle + \langle \varepsilon^{f2} \rangle]}_i$$
$$= \overline{[\langle \varepsilon^{o2} \rangle + \langle \varepsilon^{f2} \rangle]}_i \quad \text{assuming} \quad \langle \varepsilon^o \varepsilon^f \rangle = 0 = \overline{[\langle \varepsilon^{f2} \rangle]}_i + \overline{[\langle \varepsilon^{o2} \rangle]}_i, \tag{6}$$

where ε^f and ε^o denote forecast and observation errors. Hence, if forecast errors were uncorrelated with observation errors, the expected value of the bin-averaged square of the innovation is the sum of the bin-averaged forecast error variance $\overline{[\langle \varepsilon^{f2} \rangle]}_i$ and the bin-averaged observation error variance $\overline{[\langle \varepsilon^{o2} \rangle]}_i$.

9) Plot the conditioned mean-square innovation $\overline{[(y_u - x^f_{u0})^2]}_i$ as a function of $\overline{(s^2)}_i$. Ideally, the bin-averaged sample variance $\overline{(s^2)}_i$ should be equal to the bin-averaged forecast error variance $\overline{[\langle \varepsilon^{f2} \rangle]}_i$ and, hence, a plot of $\overline{[(y_u - x^f_{u0})^2]}_i$ as a function of $\overline{(s^2)}_i$ should yield a line with a slope of 1 and an ordinate intercept of $\overline{\langle \varepsilon^{o2} \rangle}$, where $\overline{(s^2)}_i = 0$. We shall refer to plots of $\overline{[(y_u - x^f_{u0})^2]}_i$ as a function of $\overline{(s^2)}_i$ as spread–skill plots.

Figure 6 shows the spread—skill plots at 850 mb for temperature, dewpoint temperature, and the zonal and meridional wind components. In all cases, a robust spread–skill relationship is obtained. In particular, the slopes of the lines pertaining to the innovation variance with respect to the ensemble mean are close to unity. Navy data assimilation schemes estimate radiosonde observation error variance to be about 1 K^2 and 4 m^2 s^{-2} for 850-mb temperature and wind observations, respectively. Hence, the intercepts of the ordinate axis displayed in Fig. 6 are somewhat larger than these estimates of radiosonde observation error variance. However, the observation error variance that contributes to the innovation variance $\overline{[(y_u - x^f_{u0})^2]}_i$ is contributed to by both instrument error variance and the error variance of representation. The error variance of representation that accounts for the fact that the point observation provided by a radiosonde is unrepresentative of the temporally and spatially averaged values represented by a numerical model, which is flow dependent, depends on the grid resolution of the model and is difficult to quantify. Hence, the hypothesis that the ensemble is not underdispersive can only be rejected with limited confidence.

Hamill et al. (2004) report that the ensemble archived as part of the multidecadal reanalysis described in Kalnay et al. (1996) did not show a spread–skill relationship for precipitation. To analyze the precipitation prediction performance of our ensemble, the stage IV regional multisensor precipitation analysis was interpolated to the COAMPS grid and treated as "observations." This stage IV analysis is generated at the National Centers for Environmental Prediction (NCEP). The components derive from the River Forecast Centers quality controlled precipitation analyses. (A reference describing these data can be found online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/.)

Figure 7 shows that a useful spread–skill relationship holds for the 48-h accumulated precipitation in our ET LAM. In contrast to the spread–skill relationships

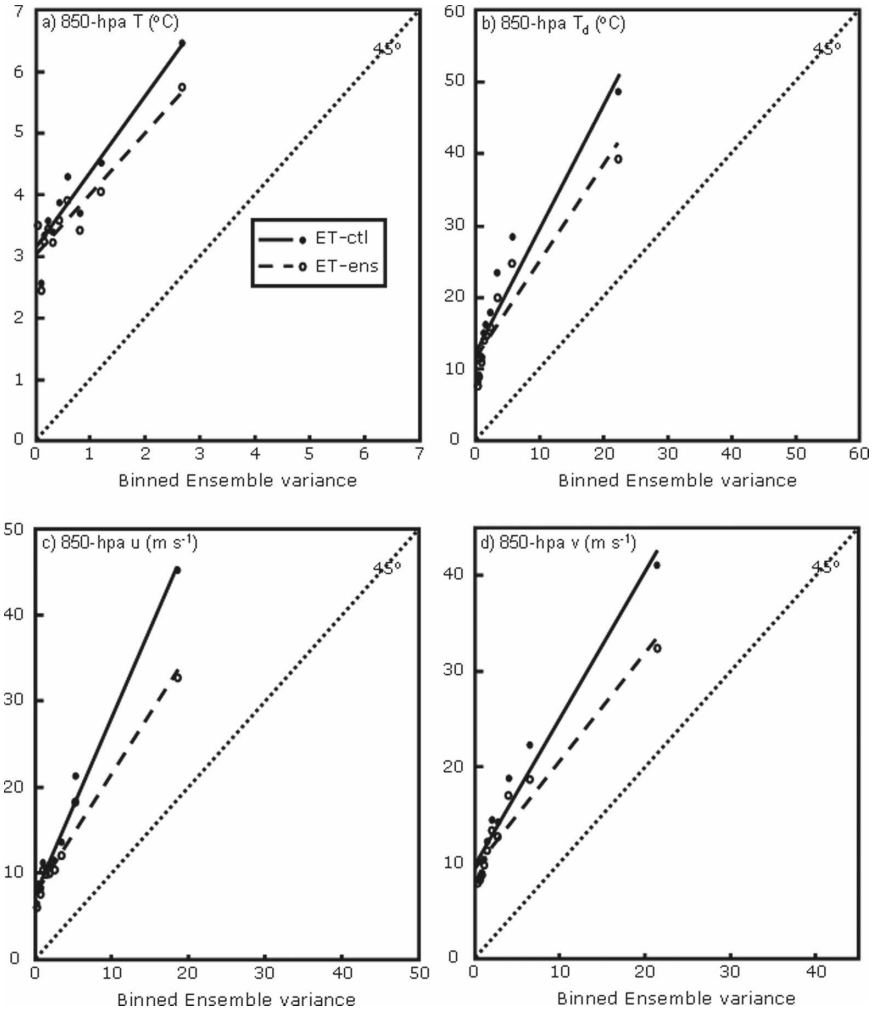

FIG. 6. Spread–skill plot for COAMPS simulations relative to the control mean (solid line) and ensemble mean (dashed) of 850-hPa (a) air temperature (°C), (b) dewpoint temperature (°C), (c) u-wind component (m s^{-1}), and (d) v-wind component (m s^{-1}).

shown in Fig. 6, however, here the slope of the spread–skill line is clearly too steep indicating that the ensemble is underdispersive. The horizontal lines in Fig. 7 show that the mean-square difference of the ensemble mean precipitation forecast relative to observations is 19 mm^2 less than that of the control forecast. We speculate that the lack of ensemble precipitation spread is because, in this paper, we have not attempted to account for the effects of model misrepresentations of subgrid-scale processes.

4. Conclusions

A new ET method of generating ensembles of initial conditions for LAM ensemble forecasting has been introduced. A 3-week experiment using the method demonstrated the following.

- The root-mean-square error of the ET ensemble mean is significantly smaller than the unperturbed ensemble control forecast.
- The ET ensemble provides a useful spread–skill relationship for u, v, T, T_d, and 48-h accumulated precipitation. Diagnostics indicate that ensemble variance was marginally smaller than the forecast error variance for wind and temperature variables, but the ensembles precipitation variance was significantly smaller than the precipitation forecast error variance.
- The method of generating initial ensemble perturbations is computationally inexpensive compared to a perturbed observation ensemble data assimilation scheme or any of the ensemble Kalman filter data assimilation schemes that use localization.
- The method can incorporate constraints from nonensemble data assimilation scheme.

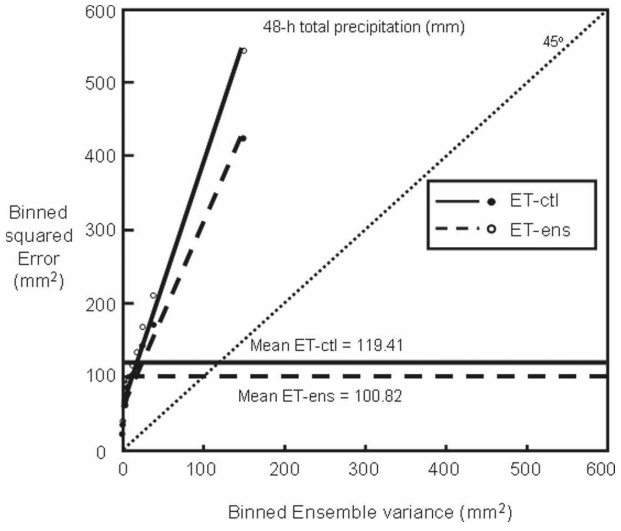

FIG. 7. Spread–skill plot for COAMPS simulations similar to Fig. 6, but for 48-h total accumulated precipitation (mm^2). The horizontal lines give the mean-square precipitation error of the control forecast (solid line) and of the ensemble mean (dashed line).

- The method ensures that the horizontal and vertical scale of the initial perturbations matches the scales of resolved by variable resolution models.

Presumably, the dispersion of the ensemble of precipitation forecasts would be increased by letting each ensemble member have differing settings of uncertain model parameters that are linked to precipitation. We hope to address this issue in future work.

Acknowledgments. CHB, TH, JN, WTT, SC, and JD acknowledge support from ONR Project Element 0602435N. Computational resources were supported in part by a grant of High Performance Computing (HPC) time from the Department of Defense (DoD) Major Shared Resource Center at NAVOCEANO in Stennis, MS, and the DoD HPC Distributed Center at FNMOC in Monterey, CA.

REFERENCES

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.,* **127,** 1941–1953.

Barker, E. H., 1992: Design of the Navy's multivariate optimum interpolation analysis system. *Wea. Forecasting,* **7,** 220–231.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.,* **56,** 1748–1765.

Bowler, N. E., 2006: Explicitly accounting for observation error in categorical verification of forecasts. *Mon. Wea. Rev.,* **134,** 1600–1606.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

Cooper, M., and K. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.,* **101** (C1), 1059–1077.

Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.,* **129,** 869–883.

Davies, H. C., 1976: A lateral boundary formulation for multilevel prediction models. *Quart. J. Roy. Meteor. Soc.,* **102,** 405–418.

Errico, R. M., E. H. Barker, and R. Gelaro, 1988: A determination of balanced normal modes for two models. *Mon. Wea. Rev.,* **116,** 2717–2724.

Fisher, M., and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo. 220, 35 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

Grimit, E. P., and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting,* **17,** 192–205.

——, and ——, 2007: Measuring the ensemble spread–error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Wea. Rev.,* **135,** 203–221.

Hallett, J., and S. C. Mossop, 1974: Production of secondary ice particles during the riming process. *Nature,* **249,** 26–28.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensembles. *Mon. Wea. Rev.,* **129,** 550–560.

——, J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.,* **132,** 1434–1447.

Harshvardhan, R. Davies, D. Randall, and T. Corsetti, 1987: A fast radiation parameterization for atmospheric circulation models. *J. Geophys. Res.,* **92,** 1009–1015.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean–Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.,* **125,** 1414–1430.

Holt, T. R., J. Pullen, and C. H. Bishop, 2008: Urban and ocean ensembles for improved meteorological modeling of the coastal zone. *Tellus,* in press.

Homar, V., D. J. Stensrud, J. J. Levit, and D. R. Bright, 2006: Value of human-generated perturbations in short-range ensemble forecasts of severe weather. *Wea. Forecasting,* **21,** 347–363.

Houtekamer, P., 1993: Global and local skill forecasts. *Mon. Wea. Rev.,* **121,** 1834–1846.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 46, Amer. Meteor. Soc., 165–170.

Katz, R. W., 1982: Statistical evaluation of climate experiments with general circulation models: A parametric time series modeling approach. *J. Atmos. Sci.,* **39,** 1446–1455.

Klemp, J., and R. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.,* **35,** 1070–1096.

Leslie, L. M., and M. S. Speer, 2000: Comments on "using ensembles for short-range forecasting." *Mon. Wea. Rev.,* **128,** 3018–3020.

Lorenz, E. N., 1969: Predictability of a flow which possesses many scales of motion. *Tellus,* **21A,** 289–307.

Louis, J. F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Bound.-Layer Meteor.,* **17,** 187–202.

——, M. Tiedtke, and J. F. Geleyn, 1982: A short history of the operational PBL-parameterization at ECMWF. *Workshop on Planetary Boundary Parameterization,* Reading, United Kingdom, ECMWF, 59–79.

Majumdar, S. J., C. H. Bishop, I. Szunyogh, and Z. Toth, 2001: Can an Ensemble Transform Kalman Filter predict the reduction in forecast error variance produced by targeted observations? *Quart. J. Roy. Meteor. Soc.,* **127,** 2803–2820.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: Validation of the methodology and verification. *Nonlinear Processes Geophys.,* **12,** 527–536.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.,* **136,** 1093–1108.

Meyers, M. P., P. J. DeMott, and W. R. Cotton, 1992: New primary ice-nucleation parameterizations in an explicit cloud model. *J. Appl. Meteor.,* **31,** 708–721.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

Rutledge, S. A., and P. V. Hobbs, 1983: The mesoscale and microscale structure of organization of clouds and precipitation in midlatitude cyclones. VIII: A model for the "seeder-feeder" process in warm-frontal rainbands. *J. Atmos. Sci.,* **40,** 1185–1206.

——, and ——, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cold-frontal rainbands. *J. Atmos. Sci.,* **41,** 2949–2972.

Scherrer, S. C., C. Appenzeller, P. Eckert, and D. Cattani, 2004: Analysis of the spread–skill relations using the ECMWF ensemble prediction system over Europe. *Wea. Forecasting,* **19,** 552–565.

Stensrud, D. J., H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433–446.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

Walser, A., M. Arpagaus, C. Appenzeller, and M. Leutbecher, 2006: The impact of moist singular vectors and horizontal resolution on short-range limited-area ensemble forecasts for two European winter storms. *Mon. Wea. Rev.,* **134,** 2877–2887.

Wang, S., Q. Wang, and J. Doyle, 2002: Some improvement of Louis surface flux parameterization. Preprints, *15th Symp. on Boundary Layers and Turbulence,* Wageningen, Netherlands, Amer. Meteor. Soc., 547–550.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60,** 1140–1158.

——, and ——, 2005: Improvement of ensemble reliability with a new dressing kernel. *Quart. J. Roy. Meteor. Soc.,* **131,** 965–986.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Xu, M., D. J. Stensrud, J. W. Bao, and T. T. Warner, 2001: Applications of the adjoint technique to short-range ensemble forecasting of mesoscale convective systems. *Mon. Wea. Rev.,* **129,** 1395–1418.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.,* **60,** 1173–1185.