Ensemble Forecasting in the Short to Medium Range: Report from a Workshop

Thomas M. Hamill,[*,+] Steven L. Mullen,[#] Chris Snyder,[*] Zoltan Toth,[@] and David P. Baumhefner[*]

1. Introduction

A workshop on ensemble forecasting in the short to medium ranges (0–14 days forecast lead time) was held at the National Center for Atmospheric Research in Boulder, Colorado, 9–11 September 1999. Approximately 45 people attended this workshop, with approximately a quarter joining us from outside the United States. The purpose of the workshop was to discuss the current state of knowledge of ensemble forecasting, to define the most important research problems for the next few years, and to seek common evaluation methods and tools. The sessions in this workshop were organized around three general topics: 1) use of ensemble forecasts for data assimilation, 2) issues related to model error in ensemble forecasts, and 3) the use, utility, and interpretation of ensemble forecasts.

In this meeting summary, we first provide some background on ensemble forecasting. We also compare the current state of our knowledge to what was known at the last U.S. ensemble forecasting workshop in 1994 (Brooks et al. 1995). Next, we provide summaries and recommendations from each of the three workshop sessions and end with a brief conclusion.

2. Background and recent progress in ensemble forecasting

Ensemble forecasting (EF) has been embraced as a practical way of estimating the uncertainty of a weather forecast. Since Lorenz (1963, 1969) it has been recognized that perfect numerical weather forecasts will always be unattainable; even the smallest of errors in the initial conditions will grow inexorably, eventually rendering any single deterministic forecast useless. Rather than pinning unrealistic hopes upon the accuracy of a single numerical forecast, EF adopts an alternative approach: generate multiple, individual numerical forecasts from different initial conditions and/or different numerical model configurations (Leith 1974). Probabilistic forecasts of the weather may then be generated from the relative frequencies of events in the en-

*National Center for Atmospheric Research,[1] Boulder, Colorado.
[+]Current affiliation: NOAA Climate Diagnostics Center, Boulder, Colorado.
[#]University of Arizona, Tucson, Arizona.
[@]National Centers for Environmental Prediction and General Sciences Corporation, Camp Springs, Maryland.
[1]The National Center for Atmospheric Research is sponsored by the National Science Foundation.
Corresponding author address: Dr. Thomas M. Hamill, NCAR/MMM, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: hamill@cdc.noaa.gov
In final form 5 June 2000.

Contributors of meeting summaries to the *Bulletin* now have an option to have their summaries published within a quicker time frame than has previously been offered. To take advantage of this expedited publication process, these articles must be brief (no more than 24 manuscript pages), tightly written, and cannot contain tables, figures, or displayed mathematics. Furthermore, the meeting summary must be externally reviewed by at least one individual who attended the same meeting. This reviewer will be of the author's choosing; this represents a departure from the conventional peer-review process.

All candidates for this expedited process should be sent to AMS electronically, as a Microsoft Word, WordPerfect, rich text format, or ASCII text attachment, to iabrams@ametsoc.org.

semble. Medium-range EFs have been produced operationally in the United States and Europe since late 1992 (Toth and Kalnay 1993, 1997; Palmer et al. 1993; Molteni et al. 1996). At the Canadian Meteorological Centre (CMC), EFs have been produced operationally since January 1996 (Houtekamer et al. 1996a,b; Houtekamer and Lefaivre 1997). Each of the centers produces the ensembles using different forecast models and different ensemble construction techniques. EFs are also now produced operationally at several other centers around the world as well (e.g., Rennick 1995; Kobayashi et al. 1996).

Encouraging results from medium-range EFs motivated the previous workshop on ensemble forecasting, which focused on the potential utility of short-range ensemble forecasts (SREFs; Brooks et al. 1995). It was believed that SREFs might provide useful information for short-range forecast problems such as severe storms forecasting (Brooks et al. 1992) and precipitation forecasting. At that workshop, issues of how to generate ensemble forecasts were considered, as well as considerations of model error and ways of dealing with the data overload of the multiple weather forecasts. It was recognized that the problems for generating useful SREFs were potentially much more difficult than for medium-range EFs; many systematic errors corrupt our current mesoscale numerical weather forecasts, owing to insufficient model resolution, use of physical parameterizations, insufficient knowledge of the land surface condition, and other such problems. As a result of this workshop, the National Centers for Environmental Prediction (NCEP) launched a pilot project to generate a small ensemble of SREFs using a reduced-resolution version of the Eta model (Black 1994) and the Regional Spectral Model (Juang and Kanamitsu 1994). These EFs were initialized with both interpolated bred initial conditions from NCEP's medium-range EF and from a variety of inhouse analyses.

Since the 1994 workshop much has been learned about EFs and SREFs. Results from the pilot SREF project were described in Hamill and Colucci (1997, 1998) and Stensrud et al. (1999). Based on these and other recent results, NCEP plans to implement a semioperational SREF starting in 2000. Other work on EFs since the 1994 workshop includes, for example, case studies of intense cyclogenesis (Du et al. 1997; Leslie and Speer 1998; Hamill 1998; Mullen et al. 1999) and blocking (Colucci and Baumhefner 1998); the study of the performance of EFs, for precipitation forecasting (Hamill and Colucci 1997, 1998; Du et al. 1997; Eckel and Walters 1998; Buizza et al. 1999a; Mullen and Buizza 2000); a study of the benefit of postprocessing EFs (Eckel and Walters 1998); a synoptic evaluation of the NCEP medium-range ensemble (Toth et al. 1997); exploration of issues related to the choice of perturbation methodologies using perfect models (Houtekamer and Derome 1995; Anderson 1997; Hamill et al. 2000); a comparison of the relative effects of model and initial condition errors in the presence of convection (Stensrud et al. 2000); effects of increasing the ensemble size (Buizza and Palmer 1998) and the resolution of member forecasts (Buizza et al. 1998); the effects of domain size and lateral boundaries (Du and Tracton 1999); methods for evaluating EFs (Anderson 1996; Smith and Gilmour 1999; Wilson et al. 1999); examinations of spread-skill relationships in ensembles (Buizza 1997; Whitaker and Loughe 1998); and examinations of the potential utility of ensembles from multiple models and/or multiple initial conditions (Krishnamurti et al.1999; Harrison et al. 1999; Evans et al. 2000; Hou et al. 2000, Ziehmann 2000; Richardson 2000b, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*). Ehrendorfer (1997) and Palmer (2000) provide nice reviews of EF concepts.

Research in the use of ensemble forecasts for improving data assimilation has also blossomed. A crucial part of any data assimilation methodology is the specification of error statistics for the first-guess or "background" forecast. These statistics determine how much to weight the background relative to the observations and how to spread the influence of the observations away from the actual observation location. The accuracy of analyses and subsequent forecasts can potentially be improved greatly if background error covariances are better estimated. Older data assimilation methods such as optimum interpolation (Gandin 1963; Schlatter 1975; Lorenc 1981) and three-dimensional variational analysis (3D-Var; Lorenc 1986; Parrish and Derber 1992) use simple statistics for describing the background errors, which may not vary in time or space. Recent results suggest that it may be possible to generate more accurate spatially and temporally varying background error statistics from a set of EFs. Articles by Evensen (1994), Houtekamer and Mitchell (1998), Burgers et al. (1998), van Leeuwen (1999), Mitchell and Houtekamer (2000), and Hamill and Snyder (2000) discuss the use of an ensemble of forecasts using a technique called the "ensemble Kalman Filter" (EnKF). The EnKF is a special case of the nonlinear filter discussed by Anderson and Anderson (1999). Also, the use of ensembles in a reduced-rank extended Kalman filter is discussed by Fisher and Courtier (1995) and Fisher (1998), the use of bred mode information for improving analyses is discussed in Pu et al. (1997), and the use of ensemble forecast statistics for specifying improved stationary background error statistics to four-dimensional variational analysis (4D-Var) is described in Buizza and Palmer (1999).

Despite many advances in our understanding of how to use and construct ensembles, questions abound. We judged the three most important issues to be 1) how or how best to use ensemble information to improve data assimilation strategies, 2) how to address model errors in EFs, and 3) how to appropriately use and interpret the voluminous information from ensembles.

The workshop was organized into three sessions, each dealing with one of these issues. In each session, there was a set of invited, longer talks, a larger number of shorter presentations, extensive discussion, and a final working group tasked with summarizing the current state of the art and recommending areas requiring further research. A report from each session follows.

3. Session 1: Ensemble forecasting and data assimilation

A common theme ran through the presentation and discussion in this session: *ensemble forecasting and data assimilation are two aspects of the same problem,* namely, describing the evolution of the probability distribution for the atmospheric state given available observations and a forecast model. The presentations here reinforced the supposition that properly constructed ensembles may generate probabilistic information in the very short range that may be used to estimate background error statistics for data assimilation (DA) schemes, consequently improving the accuracy of analyses. Similarly, the generation of an ensemble of initial conditions for purposes of data assimilation should incorporate probabilistic information on analysis errors provided by the data assimilation scheme, which are affected by dynamically constrained errors and model errors (from the background) and by random errors (from assimilating imperfect observations). This general approach to the construction of ensemble initial conditions differs from those used in present operational ensemble systems at the European Centre for Medium-Range Weather Forecasts and NCEP. There, the ensemble of initial conditions are designed for forecast applications and project upon features that will grow or have grown rapidly, respectively.

a. *Summary of presentations*

An overview of the use of ensembles in data assimilation, with emphasis on the EnKF, was presented by P. Houtekamer (CMC). The EnKF is related to the Kalman filter (Kalman 1960), which provides the optimal estimate of the state of a linear dynamical system under the assumption that observational and background error statistics are precisely known and are Gaussian. The EnKF, however, differs from the Kalman filter in that the error covariance is estimated from an ensemble of short-range, nonlinear forecasts; at analysis times, each member is then updated in such a way that the ensemble perturbations approximate a random sample from the analysis error distribution. The effectiveness of the EnKF, even for small ensembles (~10 to 100 members) has been demonstrated in simple models, but it remains untested in operational numerical weather prediction (NWP).

Houtekamer discussed three practical difficulties related to operational implementation of the EnKF:

1) *Rank deficiency.* Since feasible ensembles in NWP are composed of far fewer members than the degrees of freedom in the forecast model, the ensemble perturbations cannot span the space of model solutions, and their sample covariance matrix is rank deficient. The resulting analysis corrects the background only in the subspace spanned by the ensemble members.
2) *Sampling errors.* Covariance estimates from a finite sample are subject to sampling errors that decrease only slowly with the size n of the ensemble (as $n^{-1/2}$; Casella and Berger 1990). Characteristic sampling errors include spurious correlations between widely separated locations and the overestimation of the leading eigenvalues of the covariance matrix.
3) *Model error.* Estimating background covariances solely from an ensemble of forecasts generated by the same imperfect model ignores the contribution of the error in the forecast model to the uncertainty of the background. Since a key source of model error is the omission or parameterization of unresolved scales, this problem can be expected to be worst at the smallest resolved scales and to lead to

a systematic underestimation of uncertainty at the small scales by the ensemble.

Houtekamer also emphasized that the use of the full nonlinear forecast model and forward operators in the covariance calculations both simplifies the scheme and appears to make it more robust, even though the Kalman filter is formally applicable only for linear dynamics and forward operators.

Presumably, however, there is a point at which a system's dynamics become sufficiently nonlinear, and the distributions of interest sufficiently non-Gaussian, that a more general approach is required. J. Anderson described how one might build an EF/DA scheme in this regime (Anderson and Anderson 1999). As it turns out, the framework remains similar to the EnKF. An ensemble of short-term forecasts is used to estimate the prior distribution of the atmospheric state; when observations are available, this estimated distribution is updated given the new observations by Bayes' rule. Finally, a new ensemble of initial conditions is generated as a random sample from the updated distribution. Unlike the EnKF, where all distributions are assumed Gaussian and the ensemble simply provides an estimate of the background covariances, the distributions in this general case are approximated through nonparametric density-estimation techniques (Silverman 1986), which make minimal assumptions on the form of the distribution. Another presentation by M. Berliner further discussed the statistical concepts behind this approach.

Shorter presentations were made on ensemble Kalman filtering approaches (T. Hamill, J. Whitaker, C. Bishop, J. Hansen), on the improvement of 3D-Var using information from bred modes (D. Barker), on singular vectors (R. Gelaro, J. Ahlquist), and on the limits of linearity assumptions in the construction of ensemble perturbations (I. Gilmour).

b. Working group on ensemble forecasting and data assimilation

The working group began with a discussion of the fundamental problems posed by schemes that combine EF and DA. A successful ensemble-based strategy should provide accurate estimates of background error statistics despite imperfect forecast models, despite the imperfect knowledge of the errors in the observations and their relation to the forecast variables, and despite the requirement to use an ensemble whose size is small compared to the degrees of freedom in the forecast model. Opinions varied widely on the feasibility of this enterprise and on which aspects were, in fact, the most problematic. In the end, there was agreement that experiments in simple models had suggested that schemes combining EF and DA were both feasible and useful, but that success of such schemes in more realistic environments such as in operational NWP remained uncertain.

The working group identified three areas where further research could pave the way for tests of such schemes with operational NWP models.

1) *Comparing sampling techniques for estimating the background covariance matrix.* Two main approaches have been proposed. The first directly applies Monte Carlo techniques to generate an ensemble that is (approximately) a random sample from the distribution at the analysis time, that is, the EnKF. The second approach seeks to obtain more accurate estimates with fewer members by populating the ensemble with those initial perturbations that will evolve into the leading eigenvectors of the background error covariance matrix (Ehrendorfer and Tribbia 1997; Barkmeijer et al. 1998; for a related method, see Pham et al. 1998). Both of these techniques use the same information (estimates of the observation error covariances and a previous ensemble-based estimate of the background error covariances). Experimentation is still required to determine the relative merits of each approach.
2) *Dealing with small sample sizes.* As discussed by Houtekamer in his presentation, small sample sizes give rise to many of the known difficulties of ensemble-based data assimilation. Several techniques for dealing with many of these difficulties are known. For example, one can ameliorate both rank deficiency and spurious long-range correlations by "localizing" the covariances, either implicitly by excluding distant observations when calculating the analysis at a point, or explicitly by multiplying the covariances by a decreasing function of distance. However, the relative importance of the known difficulties and the efficacy of these proposed solutions is unclear. There is the potential for further problems, such as a lack of balance in the ensemble perturbations, to arise as ensemble-based assimilation schemes are tested in more complex models.
3) *Evaluating different methods.* At present, several methods for EF/DA strategies have been proposed and tested in simple models. An important next

step will be to compare these methods. While the final test of any assimilation scheme is clearly whether it can improve operational analyses and forecasts, the consensus that emerged was that comparisons should begin with simpler models and evaluation techniques that are already available (see section 5). Further, more refined metrics, such as those that might facilitate the evaluation of distributions in several dimensions, should be developed. Finally, comparisons would be facilitated by the availability within the community of a standardized hierarchy of common models in which to test proposed methods. To facilitate such comparisons, the working group recommended that the community should agree upon, and disseminate, a hierarchy of models for testing ensemble-based DA schemes. This hierarchy should include at the least a low-order model (such as that of Lorenz 1965), a quasigeostrophic or other balanced model having many degrees of freedom, and a simplified primitive equation model. In addition, alternative versions of each model should be available, so that methods can be compared in the case of imperfect models as well. These alternative "imperfect" versions may include specified stochastic forcings or be run at substantially different resolution.

Data assimilation working group participants were J. Anderson, C. Bishop, R. Buizza, L. Fillion, R. Gelaro, I. Gilmour, J. Hansen, T. Hamill, P. Houtekamer, J. Whitaker, and C. Snyder (chair).

4. Session 2: Model errors and ensemble forecasts

When ensemble forecasting was first implemented at NCEP (Toth and Kalnay 1993, 1997) and ECMWF (Palmer et al. 1993; Molteni et al. 1996), the approach was to assess forecast uncertainty related to growth of errors in the initial conditions due to large-scale chaotic dynamics. However, forecast uncertainty also arises because imperfect numerical models are used to predict the behavior of the atmosphere. Ensemble forecast systems that simulate uncertainties due to both initial condition and model errors (e.g., Houtekamer et al. 1996a,b; Houtekamer and Lefaivre 1997) may improve the ensemble, providing a more realistic spread of forecast solutions. Hence, there has been much recent interest in designing new techniques for addressing model uncertainty.

Part of model error can be classified as *systematic* and another part as random or *stochastic*. Systematic errors are those that can be reproduced if the model is run many times over similar cases; these errors are commonly referred to as "model bias." Systematic errors are typically a consequence of model formation, such as inadequate parameterization of certain subgrid scale processes. In principle, if systematic errors are known, model forecasts can be corrected (e.g., Dee and Da Silva 1998); in practice, many of the errors may be conditional, dependent upon the occurrence of convection or other processes, making them hard to estimate with finite samples.

Stochastic errors are not reproducible; they arise at each integration time step due to numerical inaccuracies, the finite truncation at some arbitrary scale from the grid spacing, and other inaccuracies that act randomly. Stochastic errors, just like the initial errors, turn in time into the direction of fastest growing perturbation directions, increasing errors associated with atmospheric instabilities (Toth and Kalnay 1997).

a. *Summary of presentations*

In recent years, different groups developed various methods to account for model related uncertainty in ensemble forecasting. In his presentation, P. Houtekamer described the technique used at the CMC (Houtekamer et al. 1996a,b; Houtekamer and Lefaivre 1997), whereby several versions of an NWP model are developed and used in parallel with each other. These versions possibly differ from each other in horizontal resolution, treatment of orography, convection, radiation parameterization, etc. For each ensemble model integration started with unique and slightly different initial conditions, a different model version is used. The goal is to capture systematic differences or errors in model forecasts, though, as Houtekamer pointed out, the real atmospheric solution still differs more from the ensemble members than the individual forecasts differ from each other.

Next, R. Buizza described the approach recently implemented at ECMWF (Buizza et al. 1999b). There, after each time step within a model integration, stochastic multiplicative noise (described below) is applied to the net parameterized tendencies using a number chosen randomly in the [0.5, 1.5] interval. The goal is to represent stochastic errors in the parameterization of subgrid-scale processes.

P. Sardeshmukh emphasized that all current physical parameterization schemes return only the expected value of the subgrid-scale feedback on the resolved tendencies, not the full distribution (Fritsch et al. 1998). This can contribute to an error in the ensemble mean as well as the forecast spread. Stochastic noise can be introduced into ensemble parameterizations to sample the distribution of subgrid-scale feedbacks. This noise is typically introduced as *additive* noise or *multiplicative* noise. Additive noise is a separate additional noise term in the prognostic equations, while multiplicative noise is noise that is multiplied with an existing time tendency term(s). Special care must be exercised when using stochastic noise, especially multiplicative noise, as its use may strongly change the ensemble mean (Sardeshmukh et al. 2000).

L. Smith discussed whether the stated goals of ensemble forecasting are appropriate ones, given theoretical considerations about model errors (Smith 1999). He argued that because we do not and will never have a perfect model of the atmosphere, it is impossible to design a reliable, or "accountable" forecast system. Even if the initial uncertainty could be perfectly known and sampled in a statistical sense (which is not the case), without a perfect model, it is impossible to carry forward in time the initial uncertainty in a perfectly consistent manner. Smith then proposed an alternative criterion for evaluating the usefulness of ensembles: ensuring that the ensemble is constructed in such a manner that at least one member follows (or "shadows") the evolution of the real atmosphere (Gilmour and Smith 1997). A lively discussion followed as to whether the enterprise of ensemble forecasting truly was in the dire straits suggested by the talk.

Shorter presentations were made on a variety of topics: use of different convective parameterizations in SREFs (D. Stensrud), use of ensembles for hurricane track prediction (M. Ramamurthy), operational experimentation at the National Centers for Environmental Prediction (J. Du and I. Szunyogh), and multimodel EFs (C. Ziehmann and G. Pellerin).

b. Working group on model errors

Discussion started with the question, what is the main problem caused by the use of imperfect models in ensemble forecasting? Is it that the overall spread of the ensembles is too small (presumably because stochastic model errors are not accounted for), or is it that the spread in general would be adequate but certain parts of the atmospheric attractor are not covered by the model due to systematic errors? What are the relative contributions of each? No clear answers emerged to these questions. This underscores the need for more quantitative studies analyzing the role of model errors.

O. Talagrand suggested that 4D-Var offers a tool for assessing the role of model errors in short-range forecasts. The inability of the model to properly fit observational data within the range of observational errors (i.e., the extent of the misfit) can be a direct measure of model errors. Another obvious measure is a comparison between the spread of an ensemble and the ensemble mean forecast error. Ideally, these two should be similar. The deficiency in spread may thus be used as a measure of the extent of model-related uncertainties. The extent of the problem can also be diagnosed using rank histograms (also known as "Talagrand diagrams;" Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997) and their extension to multivariate domains by the minimal spanning tree method (Smith 1999).

Next, the theoretical issues previously raised by L. Smith were considered. Participants agreed that despite reservations about the ability of operational ensembles to meet strict theoretical measures of consistency, EFs still can provide valuable probabilistic information. Whatever method(s) can further contribute to the practical goal of improving probabilistic forecasts is worthy of consideration, including statistical postprocessing.

The relative merits of the CMC and ECMWF approaches to model uncertainty were discussed next. Each approach was judged useful, but when used independently, they were noted to be able to simulate only systematic or stochastic errors, but not both. Some hybrid of the different methods that can address both issues is clearly desirable. This is an area where further research is strongly recommended.

Because different forecast systems have different biases, a multimodel, multianalysis ensemble may have appeal (e.g., Harrison et al. 1999; Evans et al. 2000; Hou et al. 2000; Richardson 2000b, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*). Here, perturbations may be centered about different control forecasts and/or use different forecast models. A simpler approach is to form a smaller ensemble from different control forecasts (Krishnamurti et al. 1999; Ziehmann 2000). However, the number of control forecasts that can be combined is limited, and the prediction of rare events from a small ensemble of control forecasts is especially problematic (the rarer the event, the more ensemble members needed to accurately assess the probability of that event). The group agreed that combining ensembles generated operationally at different centers, preferably after their bias have been removed, shows promise as an approach to ensemble forecasting in the short term.

The closing discussion of the working group focused on a question raised by Z. Toth. In our efforts to ameliorate the problem related to accounting for model uncertainties, shall we try to build, maintain, and develop a host of models that can work well in different situations, ensuring that at least some solutions may have a chance of verifying well? Or, alternatively, shall we attempt to build one model that can possibly encompass all model versions by randomly varying the structure, parameters, and/or stochastic noise within the physics packages? In other words, shall we try to maintain different model versions that each work best under various conditions, or instead, try to build one model that is inclusive of all versions? No consensus was developed, but it was noted that within the United States, the current development of a single regional forecast modeling system (Dudhia et al. 1998) may offer a unique opportunity for the modeling community to address the issue of model uncertainties within the framework of a single model.

Participants at the model errors working group were J. Ahlquist, J. Du, D. Orrell, G. Pellerin, L. Smith, I. Szunyogh, O. Talagrand, and Z. Toth (chair).

5. Session 3: Use, utility, and interpretation of ensemble forecasts

a. *Summary of presentations*

First, O. Talagrand provided an overview of the verification of EFs. Ideally, a probabilistic prediction ought to have high reliability (i.e., exhibit low conditional bias for each issued forecast probability) and high resolution (ability of different forecasts to distinguish between different events; if the model is reliable, resolution is related to sharpness). Because qualities of probability distributions such as reliability are being evaluated, it is impossible to objectively assess the quality of an individual ensemble forecast; hence, EF systems must be verified over many cases. Talagrand explained how common scoring metrics (Brier Score, Ranked Probability Score, Relative Operating Characteristics curves, rank histograms, etc.) are contaminated by at least three sources of noise: improper estimates of probabilities from small-sized ensembles, insufficient variety and number of cases in the forecast evaluation, and imperfect observations.

D. Richardson then discussed the potential economic value of EFs (Richardson 2000a). Proper evaluation of the benefit of a forecast system to a particular user should involve not only the intrinsic skill of the forecasts, but also knowledge of the weather-sensitivity and decision-making process of the end user. Reliance on skill measures alone may give a misleading impression of forecast value. To illustrate the concept, he considered the impact of ensemble size on forecast value using a simple decision model (Murphy 1994; Katz and Murphy 1997) and output from the ECMWF Ensemble Prediction System (EPS). Probability forecasts derived from the EPS have greater benefit than a deterministic forecast produced by the same model, and for many users they can have more value than a shorter-range deterministic forecast by the same model or a deterministic forecast from a higher resolution model. The additional information in the EPS, which reflects only the uncertainty in the initial conditions, provides a benefit to users equivalent to many years development of the forecast model and assimilation system. While the difference in skill between 10 and 50 members appears relatively small, the larger ensemble size can yield substantial benefit to a range of users. It also appears that an increase in ensemble size to beyond 51 (the current size for the EPS) can provide additional value, especially for extreme (and unlikely) events.

H. Brooks closed the invited presentations with a discussion of the problems in designing ensemble forecasts for mesoscale weather prediction. Brooks started by noting that computing power has now reached the point where it is technologically feasible to run SREFs with mesoscale models (grid spacing of 30 km or less) in real time on local workstations. The construction of a mesoscale EF system with a limited-area model brings with it several basic questions: how do we even create such an ensemble? What can (and should) we infer about the weather from such an ensemble? How do we provide the output in a timely manner so that decision makers and forecasters can use it? And how do we evaluate the ensemble? The answers to these questions may not be the same as they are for longer-range ensembles, where the lead time and response time is greater. Many fields for which predictive information is desired are poorly observed at the appropriate scales; thus, the errors in mesoscale features can be close to saturated in initial analyses, and verification of them is a very uncertain proposition. Moreover, mesoscale weather phenomena (e.g., severe thunderstorms) can be strongly nonlinear and intermittent:

they quickly appear, amplify, then dissipate within the forecast period. Finally, the model error strongly depends on the phenomenon being forecast and likely becomes more important than initial data errors after a very short time (tens of minutes to a few hours) into the forecast.

Shorter presentations were given on SREFs in the United States (S. Tracton), SREFs in Europe (K. Mylne), postprocessing of EFs (F. Eckel), and the economic value of EFs (Z. Toth).

b. *Working group on the use and interpretation of ensemble forecasts*

The group first discussed the need for researchers, forecasters, and end users to understand the benefits and limitations of ensemble forecasting. In particular, numerical experimentation suggests there may be a time limit to the ultimate predictability of the atmosphere (Lorenz 1969, 1982). This limit is presumed to vary with the forecast variable in question (e.g., 500-hPa geopotential heights are predictable for longer than cloud cover), the scale of the phenomenon (e.g., Van den Dool and Saha 1990), and the spatial and temporal averaging that are performed (Islam et al. 1993; Vannitsem and Nicolis 1998). Ensemble techniques probably cannot change this limit but can, through ensemble averaging, improved initial conditions, and estimation of case-dependent forecast uncertainty, bring the level of forecast performance somewhat closer to this limit.

Predictability limits are uncertain because they can only be estimated from imperfect numerical models. These estimates are strongly dependent on the dispersion properties of the chosen model and on how close the modeled variance is to the observed variance. If a model lacks variance for a particular scale of motion or cannot even resolve it, energy cascades across the scales will be treated improperly, and the predictability estimate can be expected to be unduly optimistic. In the future, it would be particularly useful to perform new predictability experiments using models with smaller grid spacings so a larger part of the spectrum of atmospheric motions and their interactions can be well resolved. Further research into how predictability estimates change with the scale of the phenomenon is also suggested.

Verification and diagnosis of ensemble forecasts were next discussed. A single verification score is generally inadequate for evaluating all of the desired information about the performance of an analysis/forecast. Each measure provides unique information on system performance. For that reason, a suite of verification measures, appropriate for the evaluation of probabilistic forecasts, should be used.

The group agreed for the need to establish a generally accepted, standardized suite of verification scores and diagrams to evaluate ensemble systems. The group's consensus was that, at a minimum, the following metrics should be used.

1) Probabilistic skill score measures such as the Brier Score, Brier Skill Score, Ranked Probability Score, and/or Ranked Probability Skill Score (Wilks 1995). These scores can provide an overall, single-number metric for judging the quality of probabilistic forecasts. Their very simplicity also prohibits them from being very informative about the nature of probabilistic forecast errors (Murphy and Winkler 1987; Murphy 1991). However, the Brier score can be decomposed into reliability, resolution, and uncertainty terms (Murphy 1973). A similar decomposition for a continuous ranked probability score was proposed recently (Hersbach 2000).
2) Reliability diagrams (Wilks 1995), plotted together with a decomposition of the Brier score and information on the distribution of forecasts issued (the sharpness). Reliability diagrams can provide information on conditional biases of ensemble forecasts. However, as noted in Wilks (1995), they can be noisy and uninformative unless populated over a large set of cases. Recently, Hamill (1997) demonstrated a multicategory reliability diagram that is less sensitive to the number of cases.
3) The Relative Operating Characteristic (ROC) (Swets 1973; Mason 1982; Stanski et al. 1989; Atger 1999). The ROC curve graphs probabilities of incorrect null and alternative hypotheses as each sorted ensemble member is used as a decision-making threshold. The ROC curve is based on stratification by observations; it is independent of reliability and instead provides a measure of resolution. It is particularly valuable for comparing the performance of ensemble systems against single deterministic forecasts at higher resolution, and the more general resource issue of ensemble size/configuration versus model resolution. Moreover, potential economic value for the simple decision model discussed by Richardson (2000a) is uniquely determined from ROC curves.
4) Rank histograms (Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997), as well as their extension to higher dimensions by the minimal spanning tree (Smith 1999). These diagnose

the ability of the ensemble to sample from the correct probability distribution. Model bias and under- or overvariability of the ensemble can detected from the shape of the rank histogram.

Other evaluative techniques (spread/skill relationships, cluster analysis, etc.) may prove useful depending on the research issue in question.

Since the parameters verified should be driven by the needs of the end users of the forecasts, the group suggested that ongoing verification efforts begin to place more emphasis on sensible weather instead of traditional fields such as the 500-hPa height field. There should also be a concerted effort to assess performance of ensemble forecasts for rare events and for high-impact weather (e.g., severe thunderstorms). Of course, verification assumes that accurate observed fields are available, which unfortunately is not always the case for many regions (e.g., over the ocean), weather features (e.g., clear-air turbulence), and scales (e.g., meso- and microscale). In fact, the question of how to verify mesoscale EFs and validate model variability, in the absence of complete observations and quantitative estimates of observational/analysis uncertainty, is one of the most challenging issues facing the community.

Allocation of computational resources is another important consideration for optimal implementation of any analysis-forecast system. Ongoing evaluation will always be required to determine the optimal tradeoff between grid spacing and the number of ensemble members; this can be expected to change as computational resources increase. The evaluation should also include the best ways to deal with model errors (see section 4).

The research community and forecast users would benefit from access to both the full ensemble forecast fields and some quick, useful summary information of the EFs. The full forecast fields would be useful to leverage the local forecast offices and universities to accelerate the development and improvement of ensemble forecast systems. These users could explore in full the potential benefits of ensembles and tailor ensemble products to their particular forecast problem, and would be provided the necessary initial analyses and lateral boundary conditions to run finescale, limited-area mesoscale ensembles on local workstations. Then again, in view of operational time constraints, it is impossible for a forecaster to examine and mentally synthesize every individual outcome from an ensemble, so synthesized products are desired in addition to raw model output. Further research into effective ways to condense, synthesize, and visualize ensemble output is suggested.

Closely related to the synthesis of ensemble output is its calibration. It is firmly established that statistical postprocessing of NWP output can significantly improve the skill of deterministic forecasts, primarily through the reduction of biases (e.g., Carter et al. 1989). The calibration of ensemble forecasts presents greater challenges than that of deterministic forecasts because the higher moments of the probability distribution may be mis-specified, not just the mean. Recent results indicate that techniques besides multiple linear regression can be successfully employed to calibrate ensemble output (e.g., Zhu et al. 1996; Hamill and Colucci 1998; Eckel and Walters 1998).

Discussion group participants included D. Baumhefner, T. Eckel, K. Mylne, M. Rennick, D. Richardson, S. Tracton, C. Ziehmann, and S. Mullen (chair).

6. Conclusions

Clearly, research on EFs and their use is growing. Despite the progress, ensemble forecasting is not yet used to its full potential in this country. We believe that this is due partly to a long institutional and societal inertia toward making and using deterministic rather than probabilistic forecasts, and partly because ensemble forecasting is still a relatively new endeavor. This workshop highlighted how far we have come in the last five years. Five years ago the potential usefulness of ensembles in data assimilation was not widely appreciated. Five years ago we knew model errors were problematic but had few ideas about how best to address them in an ensemble. Five years ago our knowledge of how to interpret ensemble forecasts and how to use them was minimal. Over the next five years, our community will be testing coupled ensemble forecast/data assimilation schemes in realistic numerical models; we will be researching and testing ways of addressing model errors and model uncertainty; and we will be looking for more effective ways to verify and communicate ensemble forecasts.

References

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9,** 1518–1530.

——, 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.,* **125,** 2969–2983.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,* **127,** 2741–2758.

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

Barkmeijer, J., M. van Gijzen, and F. Bouttier, 1998: Singular vectors and estimates of the analysis error covariance metric. *Quart. J. Roy. Meteor. Soc.,* **124,** 1695–1713.

Black, T. L., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Brooks, H. E., C. A. Doswell, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting,* **7,** 120–132.

——, M. S. Tracton, D. J. Stensrud, G. DiMego, and Z. Toth, 1995: Short-range ensemble forecasting: Report from a workshop, 25–27 July 1994. *Bull. Amer. Meteor. Soc.,* **76,** 1617–1624.

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* **125,** 99–119.

——, and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.,* **126,** 2503–2518.

——, and ——, 1999: Ensemble data assimilation. *Proc. 17th Conf. Weather Analysis and Forecasting,* Denver, CO, Amer. Meteor. Soc., 231–234.

——, and Coauthors, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **124,** 1935–1960.

——, A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999a: Probabilistic predictions of precipitation using the ECMWF ensemble prediction system. *Wea. Forecasting,* **14,** 168–189.

——, M. Miller, and T. N. Palmer, 1999b: Stochastic simulation of model uncertainty in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **125,** 2887–2908.

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

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

Casella, G., and R. L. Berger, 1990: *Statistical Inference.* Duxbury Press, 650 pp.

Colucci, S. J., and D. P. Baumhefner, 1998: Numerical prediction of the onset of blocking: A case study with forecast ensembles. *Mon. Wea. Rev.,* **126,** 773–784.

Dee, D. P., and A. M. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.,* **124,** 269–295.

Du, J., and M. S. Tracton, 1999: Impact of lateral boundary conditions on regional-model ensemble prediction. *Research Activities in Atmospheric and Oceanic Modelling,* H. Ritchie, Ed., CAS/JSC Working Group Numerical Experimentation Rep. 28, WMO/TD No. 942, 6.7–6.8.

——, S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.,* **125,** 2427–2459.

Dudhia, J., J. Klemp, W. C. Skamarock, D. Dempsey, Z. I. Janjic, S. G. Benjamin, and J. M. Brown, 1998: A collaborative effort towards a future community mesoscale model. Preprints, *12th Conf. on Numerical Weather Prediction,* Phoenix, AZ, Amer. Meteor. Soc., 42–43.

Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting,* **13,** 1132–1147.

Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: A review. *Meteor. Z.,* **6,** 147–183.

——, and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.,* **54,** 286–313.

Evans, R. E., M. S. J. Harrison, and R. J. Graham, 2000: Joint medium-range ensembles from The Met. Office and ECMWF systems. *Mon. Wea. Rev.,* **128,** 3104–3127.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

Fisher, M., 1998: Development of a simplified Kalman filter. Research Department Tech. Memor. 260. 16 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

——, and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. Research Department Tech. Memor. 220, 28 pp. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

Fritsch, J.M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the eighth prospectus development team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.,* **79,** 285–299.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields.* Gidrometeorologicheskoe Iszatelstvo; English translation, Israel Program for Scientific Translations, 1966, 242 pp.

Gilmour, I., and L. A. Smith, 1997: Enlightenment in shadows. *Applied Nonlinear Dynamics and Stochastic Systems near the Millenium,* J. B. Kadtke and A. Bulsara, Eds., AIP, 335–340.

Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting,* **12,** 736–741.

——, 1998: Comments on "Short-range ensemble forecasting of explosive Australian east-coast cyclogenesis." *Wea. Forecasting,* **13,** 1205–1207.

——, and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.,* **125,** 1312–1327.

——, and ——, 1998: Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

——, and C. Snyder, 2000: A hybrid ensemble Kalman filter—three-dimensional variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

——, ——, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.,* **128,** 1835–1851.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: Two transplant case studies. *Quart. J. Roy. Meteor. Soc.,* **126,** 2487–2515.

Hersbach, H., 2000: Decomposition on the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2000: Objective verification of SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.,* in press.
Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.,* **123,** 2181–2196.
——, and L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.,* **125,** 2416–2426.
——, and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126,** 796–811.
——, L. Lefaivre, J. Derome, 1996a: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability,* Vol. II, Reading, United Kingdom, 121–146. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.]
——, ——, ——, H. Ritchie, and H. L. Mitchell, 1996b: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.
Islam, S., R. L. Bras, and K. A. Emanuel, 1993: Predictability of mesoscale rainfall in the Tropics. *J. Appl. Meteor.,* **32,** 297–310.
Juang, H.-M., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.,* **122,** 3–26.
Kalman, R., 1960: A new approach to linear filtering and prediction problems. *Trans. AMSE,* **82D,** 35–45.
Katz, R. W., and A. H. Murphy, 1997: *Economic Value of Weather and Climate Forecasts.* Cambridge University Press, 222 pp.
Kobayashi, C., K. Yoshimatsu, S. Maeda, and K. Takano, 1996: Dynamical one-month forecasting at JMA. Preprints, *11th Conf. Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 13–14.
Krishnamurti, T. N., and Coauthors, 1999: Improved weather and seasonal climate forecasts from multimodel superensemble. *Science,* **285,** 1548–1550.
Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.
Leslie, L. M., and M. S. Speer, 1998: Short-range ensemble forecasting of explosive Australian east-coast cyclogenesis. *Wea. Forecasting,* **13,** 822–832.
Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation system. *Mon. Wea. Rev.,* **109,** 701–721.
——, 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177–1194.
Lorenz, E. N., 1963: Deterministic nonperiod flow. *J. Atmos. Sci.,* **20,** 130–141.
——, 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus,* **17,** 321–333.
——, 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.
——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* **34,** 505–513.
Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.,* **30,** 291–303.
Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416–433.
Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.
Mullen, S. L., and R. Buizza, 2000: Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* in press.
——, J. Du, and F. Sanders, 1999: The dependence of ensemble dispersion on analysis-forecast system: Implications to short-range ensemble forecasting of precipitation. *Mon. Wea. Rev.,* **127,** 1674–1686.
Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.,* **12,** 595–600.
——, 1991: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.,* **119,** 1590–1601.
——, 1994: Assessing the economic value of weather forecasts: An overview of methods, results, and issues. *Meteor. Appl.,* **1,** 69–73.
——, and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.,* **115,** 1330–1338.
Palmer, T. N., 2000: Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.,* **63,** 71–116.
——, F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. ECMWF Seminar on Validation of models over Europe,* Vol. 1, Reading, United Kingdom, ECMWF, 21–66. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.]
Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.
Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Sys.,* **16,** 323–340.
Pu, Z.-X., E. Kalnay, D. Parrish, W.-S. Wu, and Z. Toth, 1997: The use of bred vectors in the NCEP global 3-D variational data assimilation system. *Wea. Forecasting,* **12,** 689–695.
Rennick, M. A., 1995: The ensemble forecast system (FFS). Models Department Tech. Note 2-95, Fleet Numerical and Oceanography Center, 19 pp. [Available from Models Department, FLENUMMETOCCEN, 7 Grace Hopper Ave., Monterey, CA, 93943.]
Richardson, D. S., 2000a: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **126,** 649–668.
——, 2000b: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.,* submitted.
Sardeshmuk, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate,* in press.
Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.,* **103,** 246–257.
Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.
Smith, L. A., 1999: Disentangling uncertainty and error: On the predictability of nonlinear systems. *Nonlinear Dynamics and Statistics,* Alistair E. Mees, Ed., Birkhauer Press, 31–64.
——, and I. Gilmour, 1999: Accountability and internal consistency in ensemble formation. *Proc. the ECMWF Workshop on Predictability,* Reading, United Kingdom, ECMWF, 113–128. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.]
Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. Environment Canada Research Rep. 89-5, 114 pp. [Available from Atmospheric Environment Service, Forecast Research Division, 4905 Dufferin St., Downsview, ON M3H 5T4, Canada.]

Stensrud, D. J., and Coauthors, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433–446.
——, J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077–2107.
Swets, J. A., 1973: The relative operating characteristic in psychology. *Science,* **182,** 990–999.
Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability,* Reading, United Kingdom, ECMWF, 1–26. [Available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.]
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.
——, ——, M. S. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting,* **12,** 140–153.
Van den Dool, H. M., and S. Saha, 1990: Frequency dependence in forecast skill. *Mon. Wea. Rev.,* **118,** 128–137.
van Leeuwen, P. J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique." *Mon. Wea. Rev.,* **127,** 1374–1377.
Vannitsem, S., and C. Nicholis, 1998: Dynamics of fine-scale variables versus averaged observables in a T21L3 quasi-geostrophic model. *Quart. J. Roy. Meteor. Soc.,* **124,** 2201–2226.
Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.,* **126,** 3292–3302.
Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.
Wilson, L. J., W. R. Burrows, and A. Lanzinger, 1999: A strategy for verification of weather element forecasts from an ensemble prediction system. *Mon. Wea. Rev.,* **127,** 956–970.
Zhu, Y., G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th AMS Conf. Weather Analysis and Forecasting,* Amer. Meteor. Soc., Norfolk, VA, j79–j82.
Ziehmann, C., 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus,* **52A,** 280–299.