Short-Range Ensemble Forecasting:
Concepts for Application to Severe Thunderstorms

Charles A. Doswell III
NOAA/ERL National Severe Storms Laboratory
Norman, Oklahoma (USA)

1. INTRODUCTION

Numerical weather prediction (NWP) has evolved a tradition of being mostly centralized within large national centers, running ever more sophisticated NWP models on the latest supercomputing machines the forecasting agency can afford. The products of the NWP models are then issued as guidance materials for their forecasting offices. With each incremental improvement in computing capability, it has been common to use that enhanced capability to provide an incremental enhancement to the NWP models, usually involving enhanced vertical and horizontal resolution, but also such things as improved numerical schemes, more complex physical parameterizations, etc. It also has been common to retain previous versions of the opera-tional models, for continuity and for comparison purposes. Moreover, there are now many players in this game, so many nations or groups of nations have their own different NWP systems, and access to the output from different centers has grown to the point where there may be 5-10 different sets of model output available for consideration in a forecast office. Re-viewing and comparing all of these might well represent a major component of a forecaster's work time.

This proliferation of models, even two decades ago, suggested to Thompson (1977) that it might be possible to use the known biases of each model to arrive at a sort of consensus forecast among the different models. In Thompson's view, it would be useful to take advantage of the strengths and weaknesses of the different models in developing a weighted consensus, rather than simply averaging them. It has long been known that a simple average of a number of different forecasting systems is statistically guaranteed to outperform any of the individual systems on the average over a set of forecasts. This is a sort of statistical artifact, but an inescapable one, nevertheless.

A different but clearly related concept was suggested by Epstein (1969), in-volving "stochastic-dynamic" models, wherein the forecasts were to include a sto-chastic component. This notion was car-ried further, to a full incorporation of the time evolution of the probability density function associated with the uncertainties; resulting in the so-called Liouville equation (Ehrendorfer 1994). As noted by Molteni et al. (1996), this exercise is interesting but apparently of only abstract interest due to some practical limitations.

A similar notion has been developed over a number of years by several different contributors: **ensemble** methods. These have mostly been developed for large scale, medium-range forecasts (out to 10 days or so). An excellent summary of the notions upon which medium-range ensemble forecasting is based can be found in Molteni et al. (1996). An interesting example of the idea is shown by Mullen and Baumhefner (1994), using the a reduced-resolution im-plementation of the same model to create a set of forecasts, wherein each model run was started with slightly different initial conditions. It has become widely accepted that in some situations, the model's perfor-mance seems to be situation-dependent. Moreover, it is known that there are many components to the errors in a numerical forecast, including model error, discretiza-tion error, truncation error in the finite computations, errors due to sampling or dynamics or instruments in the initial con-ditions, errors in the boundary conditions (if applicable), and so on.

These notions were being considered when the full impact of the work originally published by Lorenz (1963) began to be felt within the forecasting community during the

1980s. That the weather can be thought of as a nonlinear dynamical system had been known, of course, but the implications of "sensitive dependence on the initial conditions" began to have a profound effect on NWP. "Prediction of predictability" became something like a fad. Mullen and Baumhefner's results indicate that the consensus among the predictions from an "ensemble" of reduced-resolution model forecasts would consistently outperform (again, on the average) the single "control" run of the full-resolution model.

At the same time, it began to be observed that using a sequence of forecasts produced by the same model, all valid at the same time, this set of forecasts constituted an ensemble of sorts, referred to as the "lagged-average" of the forecasts all valid at the same time (see e.g. Hoffman and Kal-nay 1983). If the model was consistent in predicting a certain evolution, that was considered to infer that the particular evolu-tion was more likely to happen than if successive model runs had considerable variance.

Finally, it was observed that the constantly enlarging suite of models itself constitu-ted yet another form of an ensemble, this time with a wide variety of numerical me-thods, initial conditions, model physics, discretization, etc. It was recognized again (see, e.g., Wobus and Kalnay 1995) that if the models "agreed" on a certain evolution, that evolution was more likely than when the models "disagreed." This can be ex-ploited to attempt a forecast of forecasting skill in a particular situation, but it also can be considered from an ensemble viewpoint.

All of this began to suggest that using an ensemble of NWP model forecasts might be an avenue worth exploring (Brooks and Doswell 1993). In fact, several NWP centers (see Molteni et al. 1996; Tracton and Kalnay 1993) have already begun to explore the possibilities of ensemble forecasting for the "Medium Range" forecast problem (2-10 days or so). It has been discussed also for the "Short Range" forecast problem (6-48 h or so); see Brooks et al. (1995).

This rather lengthy Introduction has been needed because of the relative unfamiliarity of ensemble forecasting. I want to move toward addressing its potential uses in forecasting severe convection.

2. ENSEMBLE FORECASTS

It can be seen from the Introduction, that most forecasters actually are already familiar with many of the notions of en-semble forecasts. The multiplicity of mo-dels and the overlapping model forecasts produce a set of forecasts, all of which are valid at some future time, and differences as well as similarities always exist within that collection of "guidance" products. Given that disagreements are virtually certain, I know of no systematic way to select that single "model of the day" which is most likely to be correct, in spite of considerable forecaster folklore about how this *might* be accomplished. In fact, a considerable effort can be expended in what I view as thrashing about within the ensemble of con-flicting model guidance, attempting to puz-zle out which model to believe in a given situation. In my opinion, this is mostly a waste of time and effort.

As already noted elsewhere (Doswell 1996a), I believe that probabilistic fore-casting is the only sensible way to approach forecasting, and ensembles are an embodi-ment of the uncertainty associated with the NWP guidance output. The greatest effort in ensemble forecasting at the moment (at least for the medium-range problem) is focused on how to generate perturbations of the initial conditions in the most "efficient" way. What does this mean?

In some sense, you would like your ensemble to "span" some sort of abstract space of variability in the initial conditions. In other words, given that our knowledge of the initial state is uncertain, what is the range of that uncertainty? What are the ex-treme possibilities and how sensitive is the forecast to those extremes? What we want to know, in this sense, are the directions (in this abstract space of possible pertur-bations) of variability to which the model is most sensitive? If we have found those, then we can run an ensemble where we include members having variability in those variability "directions" and have some reasonable expectation that the collection of forecasts includes most of the plausible possible evolutions. Then, there is not likely

to be some wildly different outcome which was not contained within the ensemble. If nothing else, the Law of Murphy (not Allan Murphy, but the 'gremlin' Murphy!) says that the most likely outcome is almost certainly not included in the ensemble!

However, this is not the only source of variation in the forecasts. Perturbing the initial conditions is but one source of error in the forecast, as already noted. Including the different models within the ensemble might be a useful exercise. Basically, a focus on the initial conditions presumes that model errors are not the most important component of the forecast. However, we all know of systematic errors associated with the model that might well be a critical component if the uncertainty (noted by Molteni et al. 1996). Hence, I believe that as capabilities of new computers expand, we might well want to run separate perturbations of each of the models in the model suite to develop a wider range of possible outcomes.

All right, suppose we have some sort of an ensemble of forecasts. What do we *do* with it? The most obvious thing is to form a *consensus* forecast, since it is statistically the most likely forecast over the long haul (although it might be terribly wrong on any given day!). This might be considered to be the "guidance" product for the day, but the availability of the ensemble means that we are not stuck with only that guidance. We also have information about the *variability* within the ensemble. Presum-ably, we would have higher confidence in forecasts when the within-ensemble varia-bility is relatively low, and vice-versa.
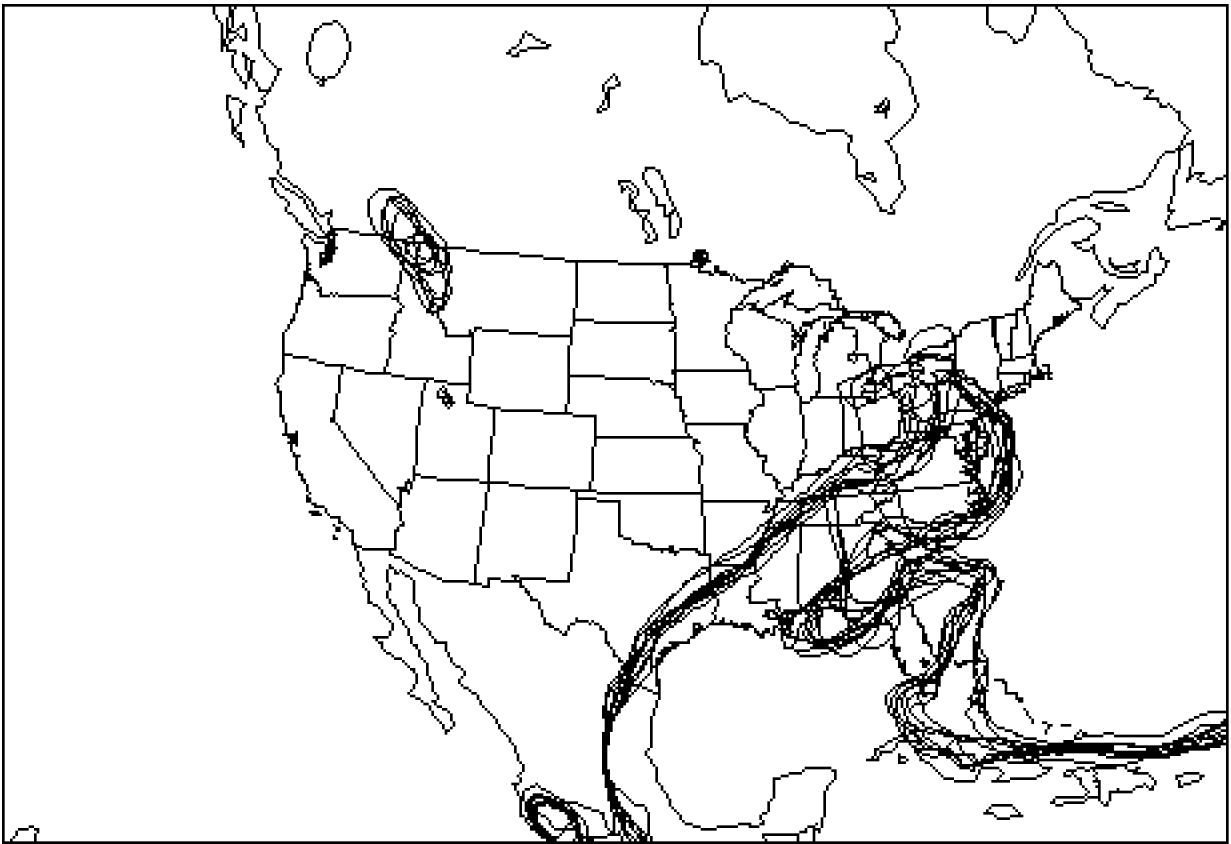

Figure 1. CAPE contour of 800 J kg-1 from all ten members of 36 h ETA ensemble forecast valid at 00 UTC on 11 May 1995. Each line represents the contour line from one ensemble member. [from Brooks et al. 1996]

But the ensemble's information content doesn't even stop there. An ensemble gives one a sense of the *most* probable outcome, and even an estimate of its probability, but it

also includes examples of low-probability outcomes that might be quite important. If we consider that in a statistical sense, hazardous weather events usually are low probability events, then we have an important benefit in the ensemble: it tells us what outcomes are *possible* within some reasonable range of variations. Although some events may be low probability events, that does not preclude their occurrence! Thus, the ensemble can alert us about that range of possibilities. It is that awareness of possibilities that I have been harping on for the last several years as a critical component in dealing with hazardous convective weather! *You are not likely to recognize an impending event for which you are not looking*.

Although we cannot say we know how to display the results of an ensemble forecast, it appears that something like Fig. 1 will be useful From this "spaghetti diagram," one can "eyeball" what the con-sensus forecast is, as well as seeing the variability and the extreme possibilities. Where the ensemble members tend to lie on top of one another, there is low variability about the consensus (implying relatively high certainty). Alternatively, large disper-sion of the ensemble members implies rela-tively low levels of confidence.

3. APPLICATION TO CONVECTION

With this lengthy preamble, I am now able to turn to the issue of how ensembles might be of considerable value in the forecast of convection and its related phenomena. Obviously, operational mo-dels do not now treat convection and its phenomena explicitly. The may not be able to do so for many years to come. Even if that capability develops faster than I anticipate it will, the *accuracy* of such forecasts is almost certainly going to be dubious (see Brooks et al. 1992). Thus, NWP has only an indirect role in the prediction of severe weather: NWP provides insight into the *possibilities* for concatenation of the ingredients for severe weather (see Doswell et al. 1996). Many issues often remain uncertain in such cases, even down to within a few hours of a major episode. Although it is possible to speak of "synoptically evident" severe weather

patterns (Doswell et al. 1993), forecasters always can find aspects over which to agonize; forecasting is never easy until the events have unfolded and the "wheelchair generals" then can see that it was "obvious."

It is not the synoptically evident situations that cause the most grief, how-ever. Rather, the worst situations are the "surprise" events that seemingly come out of nowhere, and the forecast simply has not accounted for that possibility. There may well always be some residual of such events that will remain mysterious and invite additional research, but some of them *might* have been anticipated and simply were not, for one reason or another. Having an ensemble of forecasts means the following sort of strategy might lead to an enhanced chance of anticipating rare events like severe thunderstorms.

Step 1: **Consider the severe weather potential in the consensus forecast**. This is a relatively straightforward task, simply involving application of all the stan-dard methods for diagnosing the potential for severe weather in a synoptic-scale NWP product. This would include doing diag-nosis of the model-predicted fields, etc. The object is to determine the conditional probability of hazardous weather, given that the most probable forecast is correct. This is denoted by $P(X|F_c)$, where X is the event and F_c is the consensus forecast, assumed to be correct.

Step 2: **Assess the uncertainty asso-ciated with that consensus**. This is also relatively straightforward in a situation where one has an ensemble of forecasts to consider. As already noted, it is possible to determine the probability of the consensus forecast by considering how many forecasts within the ensemble resemble the consen-sus. This might be done subjectively, or it is possible to do something like a *cluster analysis* and determine objectively (a) how closely the consensus resembles the maps within each cluster, and (b) if the consensus clearly belongs in one of the clusters, how many members of the ensemble are within that cluster. The more members in the en-semble that resemble the consensus, the

more likely is the consensus. The unconditional probability of severe weather, $P(X)$, then depends on the probability that the consensus is correct. To be specific,

$$P(X) = P(X|F_c)P(F_c)$$
$$+ \sum_{i=1}^{N-1} P(X|F_i^*)P(F_i^*)$$

where F_i^* denotes the ith cluster out of a total of N-1 "not F_c" forecast clusters. That is, there are N clusters, to one of which the consensus belongs and N-1 "other" clusters

Step 3. **Review the clusters and consider the probability for severe weather within each cluster**, even if there is only one member in any particular cluster. This is a key task in anticipating un-likely events. We are attempting to anticipate a situation that has low probability, but which contains the threat of a *significant* event. That is, we are seeking a situation where $P(X|F')$ is large, even though $P(F')$ is small, where F' is some low probability forecast out of the ensemble. For most (if not all) of the other forecasts, the $P(X|F_i)$ might well be small, no matter what the various $P(F_i)$ might be. If there is at least one possibility where $P(X|F')$ is large, and there may be more than one such possibility within any given ensemble, they may not be of sufficient likelihood even to mention in the forecast. Nevertheless, the idea is to be *aware* of those possibilities.

Step 4. **Determine the key aspects of the evolution in those forecasts of low probability that lead up to an important weather event**. Presumably, the evolution of the low probability forecast differs substantially from that of the most probable event. The forecaster should be aware of the signs in the observations that would indicate the atmosphere is indeed following the path leading to that low probability outcome. If the probability of the forecast F' turning out to be correct is increasing as the observations come in, the probability of X is thereby increasing. By watching the weather, it should be possible to detect when the atmosphere is indeed following a low-probability path and amend the probabilities of an important event accordingly.

4. DISCUSSION

With all the foregoing in mind, it seems that ensembles may have an important role to play in the future of weather forecasting, even down to questions of convection. As we move towards the new millennium, it is worth reflecting for a moment on how ensemble forecasting could alter our perspectives and ways of going about our business. Whereas the traditional "paradigm" for NWP means large computational horsepower centralized in a small number of major forecasting hubs with a staff of experts feeding this engine and a set of field forecasters to interpret the output in local terms, an ensemble approach offers some different concepts.

McIntyre (1993) has suggested some interesting ideas for the role of humans in the future. Building on McIntyre's ideas, it seems to me that forecasters may well have the ability to recognize something of the variety of possibilities inherent in a synoptic situation. Perhaps an important role in the future for humans might include the opportunity to add some ensemble members to the objectively-generated suite in case they feel that the purely objective schemes are inadequate to span the space of possibilities. If the NWP of the future involves generating ensembles of forecasts from models that are limited versions of the most sophisticated models, this suggests that some decentralization of NWP capability might be both possible and desirable. That is, the local use of models having considerably reduced computational requirements than the NWP models at large NWP centers might be advantageous. Such models could be used to explore the mesoscale or convective scale possibilities in an ensemble mode, just as the large-scale models explore the possibilities on the synoptic and global scales with ensembles. This is an important issue at the moment because with limited fiscal resources, weather forecasting services need to evaluate where to put their limited finances for model development to do the most good:

push it at the traditional centralized facilities, or spread it among a wider group of facilities. Needless to say, the centralized facilities will be adamantly opposed to even debating such an issue, but it is not obvious that they are completely unbiased contributors to the debate. We need the debate.

Moreover, the implementation of ensemble forecasting is still in its infancy. What I have suggested for using the information contained in an ensemble is only a primitive start in the direction of exploring how to use this information. High-resolution NWP model development has been the primary focus in operational centers, but it is not obvious that this is necessarily our best strategy. Perhaps if some effort of a magnitude roughly comparable to that currently made on behalf of centralized model development were expended on giving ensemble forecasting a systematic basis for exploitation in operations, some important innovations in operational use of ensembles would ensue. If the number of papers on ensembles being presented at the upcoming 15th AMS Conference on Weather Analysis and Forecasting is any guide, it appears that a significant swing toward ensembles is underway.

Perhaps only considerable time will elapse before the issue is decided, but it certainly appears that ensemble forecasting is promising enough that it deserves a careful look. As a natural basis for developing means of quantifying uncertainty in weather forecasting, and for identification of potentially serious but low-probability events, it appears to have a distinct edge over the traditional NWP paradigm, if we can learn how to exploit the advantages it exhibits.

REFERENCES

Brooks, H. E., and C. A. Doswell III, 1993: New technology and numerical weather prediction: A wasted opportunity? *Weather*, **48**, 173-177

_____, _____ and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. and Forecasting*, **8**, 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.

_____, D.J. Stensrud and C.A. Doswell III, 1996: Application of short-range NWP model ensembles to severe storm forecasting. Preprints, *18th Conf. Severe Local Storms* (San Francisco, CA), Amer. Meteor. Soc., 372-380.

Doswell, C.A. III, 1996a: Severe Thunderstorm Warning Services: Responsibilities to the public. This volume.

_____, R.H. Johns and S.J. Weiss (1993): Tornado forecasting: A review (Invited paper). *The Tornado: Its Structure, Dynamics, Hazards, and Prediction* (Geophys. Monogr. 79), Amer. Geophys. Union, 557-571.

_____, H. E. Brooks, and R. A. Maddox, 1996: Flash-flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, (in press).

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skills. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703-713.

Epstein, E. S, 1969: Stochastic dynamic predictions. *Tellus*, **21**, 739-759.

Hoffman, R.N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte-Carlo forecasting. *Tellus*, **35A**, 100-118.

Lorenz, E.N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 130-141.

McIntyre, M. E., 1988: Numerical weather prediction: A vision of the future. *Weather*, **43**, 294-298

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**, 73-120.

Mullen, S.L., and D.P. Baumhefner, 1994: Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.*, **122**, 1548-1567.

Tennekes, H., 1988: Numerical weather prediction: Illusions of security, tales of imperfection. *Weather*, **43**, 165-170

Thompson, P. D., 1977: How to improve accuracy by combining independent forecasts. *Mon. Wea. Rev.*, **105**, 228-229

Tracton, M.S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379-398.

Wobus, R.L., and E. Kalnay, 1995: Three years of operational prediction of forecast skill at NMC. *Mon. Wea. Rev.*, **123**, 2132-2148.