

The Pennsylvania State University

The Graduate School

Department of Meteorology

**ENSEMBLE FORECASTING
WITH THE ENSEMBLE TRANSFORM KALMAN FILTER**

A Thesis in

Meteorology

by

Xuguang Wang

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2004

20050121 075

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 08/31/2004		2. REPORT TYPE Final		3. DATES COVERED (From - To) 01/01/2000-08/31/2004	
4. TITLE AND SUBTITLE Ensemble Forecasting with the Ensemble Transform Kalman Filter				5a. CONTRACT NUMBER N00014-00-1-0106	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Wang, Xuguang (thesis)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Pennsylvania State University Office of Sponsored Programs 110 Technology Center University Park, PA 16802				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for Public Release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The ensemble transform Kalman filter (ETKF) initial ensemble perturbation generation scheme is introduced and compared with the simple and masked breeding schemes. Instead of directly multiplying each forecast perturbation with a rescaling factor to generate the initial perturbations as in the breeding schemes, the ETKF generates initial perturbations by postmultiplying the forecast perturbations by a transformation matrix. This matrix is chosen to ensure that the ensemble-based analysis error covariance matrix would be equal to the true analysis error covariance if the covariance matrix of the raw forecast perturbations were equal to the true forecast error covariance matrix and the data assimilation scheme were optimal. For small ensembles (~100), the computational expense of the ETKF ensemble generation is only slightly greater than that of the masked breeding scheme.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)

The thesis of Xuguang Wang was reviewed and approved* by the following:

George S. Young
Professor of Meteorology
Thesis Advisor
Chair of Committee

Michael J. Fritsch
Distinguished Professor of Meteorology

Sukyoung Lee
Associate Professor of Meteorology

David Pollard
Senior Research Associate

William H. Brune
Professor of Meteorology
Head of the Department of Meteorology

Craig H. Bishop
Senior Scientist
Naval Research Laboratory
Monterey, California
Special Member

*Signatures are on file in the Graduate School

ABSTRACT

The ensemble transform Kalman filter (ETKF) initial ensemble perturbation generation scheme is introduced and compared with the simple and masked breeding schemes. Instead of directly multiplying each forecast perturbation with a rescaling factor to generate the initial perturbations as in the breeding schemes, the ETKF generates initial perturbations by postmultiplying the forecast perturbations by a transformation matrix. This matrix is chosen to ensure that the ensemble-based analysis error covariance matrix would be equal to the true analysis error covariance if the covariance matrix of the raw forecast perturbations were equal to the true forecast error covariance matrix and the data assimilation scheme were optimal. For small ensembles (~100), the computational expense of the ETKF ensemble generation is only slightly greater than that of the masked breeding scheme.

Version 3 of the community climate model (CCM3) developed at National Center for Atmospheric Research (NCAR) is used to test and compare the ETKF and breeding schemes. The NCEP/NCAR reanalysis data for the boreal summer in 2000 are used for the initial analysis and the verifications. The ETKF ensemble variances at initial time can reflect the geographical variations of the initial condition uncertainty better than the breeding scheme. The ETKF maintains comparable amounts of variance in all orthogonal and uncorrelated directions spanning its ensemble perturbation subspace at 12-h forecast lead time, whereas both breeding techniques maintain variance in few directions. The maximal energy-norm perturbation growth within the ETKF ensemble perturbation subspace calculated with linear dynamics assumption significantly exceeds

that within the breeding perturbation subspace at short forecast lead times. The ETKF ensemble mean has lower root mean square errors than that of the breeding ensemble. The ETKF estimates of forecast error variance are considerably more accurate than those of the breeding schemes.

A new method to center the initial ensemble perturbations on the initial analysis is introduced and compared with the commonly used centering method of positive/negative paired perturbations. In the new centering method, called spherical simplex centering, one linearly dependent perturbation is added to a set of linearly independent initial perturbations to ensure that the sum of the new initial perturbations equals zero; the covariance calculated from the new initial perturbations is equal to the analysis error covariance estimated by the independent initial perturbations; and all the new initial perturbations are equally likely. When the number of uncertain directions is larger than the ensemble size, which is the case for numerical weather prediction, the spherical simplex centering has the advantage of allowing almost twice as many uncertain directions to be spanned as the symmetric positive/negative paired centering. Both centering schemes are applied to the CCM3 ETKF ensemble. Tests are performed on the accuracy of the ensemble means, the accuracy of predictions of forecast error variance and the ability of the initial ensemble variance to resolve inhomogeneities in the observational network. In all of these test categories, the spherical simplex ETKF ensemble is found to be superior to the symmetric positive/negative paired ETKF ensemble. The computational expense for generating spherical simplex ETKF initial perturbations is about as small as that for the symmetric positive/negative paired ETKF.

Because all current ensemble techniques partially misrepresent the effects of initial condition error and model error on forecast accuracy, it is inevitable that members of dynamic ensembles are not drawn from the same distribution as the distribution of truth given an ensemble forecast. To address this deficiency, a new ensemble postprocessing method that reduces seasonally averaged second moment errors of the ensemble forecasts is introduced. The method involves adding independent sets of N random 4-dimensional "dressing" perturbations to each of the K members of a dynamical ensemble forecast to obtain an $N \times K$ dressed ensemble. The new method mathematically constrains the stochastic process used to generate the statistical "dressing" perturbations so that it entirely removes seasonally averaged errors in the second moment measures. ETKF ensembles that were dressed with perturbations satisfying this constraint were found to give more accurate probabilistic forecasts than corresponding undressed ETKF ensembles. A random number generator experiment and an experiment with the CCM3 ETKF ensemble show that the previously proposed "best member" dressing method fails to reliably predict the second moment of the distribution of forecast errors whereas the new dressing method reliably predicts this second moment.

The CCM3 ETKF ensemble postprocessed with the new dressing method is applied for probabilistic forecasts of cooling degree days (CDD) for Boston. It is shown that the new kernel accounting for temporally correlated forecast errors results in ensemble forecasts of CDDs with reliable spread whereas the best member method leads to an underdispersive ensemble of CDD forecasts.

TABLE OF CONTENTS

LIST OF FIGURES.....	vii
ACKNOWLEDGEMENTS.....	viii
Chapter 1 Introduction.....	1
Chapter 2 A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes.....	7
Chapter 3 Which Is Better, an Ensemble of Positive-Negative Pairs or a Centered Spherical Simplex Ensemble?.....	8
Chapter 4 Improvement of Ensemble Reliability with a New Dressing Kernel.....	9
4.1 Introduction.....	9
4.2 Limitations of the best member dressing: the random number generator experiment.....	13
4.3 Dressing with second moment constraint.....	18
4.4 Further test with nonlinear CCM3 ETKF ensemble.....	25
4.4.1 Numerical experiment design.....	26
4.4.1.1 Dynamic ensemble, verification data, and variables of interest.....	26
4.4.1.2 Identification of the best member.....	27
4.4.1.3 Training and forecasting processes.....	28
4.4.2 Experiment results.....	30
4.5 Application on Cooling degree days forecasts at Boston.....	42
4.5.1 CDD definition.....	43
4.5.2 Application of dressing.....	43
4.5.3 Results on the reliability of the dressed CDD ensemble spread.....	45
4.6 Conclusion.....	48
Chapter 5 Concluding remarks and remaining challenges.....	52
Bibliography.....	56
Appendix A Derivation on the new dressing kernel.....	64
A1 Derivation on equation (4.6).....	64
A2 Derivation on equation (4.9a).....	65
Appendix B A list of basic concepts in data assimilation and ensemble forecasting.....	67

LIST OF FIGURES

- Figure 4.1: Random number generator experiment results in testing the reliability of the spread of the ensemble dressed by (a) the best member method and (b) the new dressing kernel. Thin solid contours indicate over-dispersive ensemble. Dashed contours indicate under-dispersive ensemble. Thick solid contours mean the spread is reliable. Contour interval is 4.17
- Figure 4.2: Illustration for the idea of the new dressing kernel in 2-dimensional space. Please refer section 4.3 for detailed explanation.21
- Figure 4.3: 14 verification sites over eastern USA for the experiment in section 4.4.27
- Figure 4.4: Rank histograms for (a) undressed, (b) new kernel dressed, (c) RS-10d-globe dressed, (d) RS-1-id-east dressed and (e) RS-id-east dressed CCM3 ETKF 500hPa U ensembles over the 14 sites from 1-day to 10-day lead times.34
- Figure 4.5: Brier scores for the undressed, new kernel dressed, RS-10d-globe dressed, RS-1-id-east dressed and RS-id-east dressed CCM3 ETKF 500hPa U ensembles from 1-day to 10-day lead times. Brier score from the sample climatology is also shown. The vertical solid and dashed lines are the standard errors of Brier score calculation with given samples for the new kernel dressed and undressed ensembles respectively.36
- Figure 4.6: seasonal mean 500hPa U ensemble covariance averaged over 91 pairs of sites among the 14 verification sites (dashed) and seasonal mean 500hPa U ensemble mean error covariance averaged over 91 pairs of sites among the 14 verification sites (solid) as a function of forecast lead times for undressed ensemble, new kernel dressed ensemble, RS-10d-globe best member dressed ensemble, RS-1-id-east best member dressed ensemble and RS-id-east best member dressed ensemble.41
- Figure 4.7: Rank histograms for undressed, new kernel dressed, and the best member dressed 3-day accumulated CDD ensembles over Boston during 2001 summer.47
- Figure 4.8: Ignorance scores for the undressed (UNDR), the best member dressed (BEST) and the new kernel (NEW) dressed CDD ensembles.48
- Figure B.1: Cartoon for a typical operational data assimilation cycle.71
- Figure B.2: Cartoon for a typical operational ensemble forecast cycle.73

ACKNOWLEDGEMENTS

The completion of this thesis would never have been possible without the patient guidance and continuous encouragement from my research advisor, Dr. Craig H. Bishop. He not only was instrumental in my thesis research but also provided me opportunities to get exposed to and communicate with scientists in the related field. His foresights and constant enthusiasm would affect me for a lifetime.

I would also like to thank my academic advisor, Dr. George. S. Young for his caring and generous guidance on pursuing an academic career. I also owe my deep gratitude for Dr. Michael J. Fritsch, Dr. Sukyoung Lee and Dr. David Pollard for their helpful suggestions on this study and enlightenment in other fields of atmospheric sciences.

Finally, I thank my family, especially my husband, Xinyu Dai, for his love, his comfort and his support.

Chapter 1

Introduction

It is widely recognized that numerical weather forecasts have limited skill. Errors in numerical forecasts are attributed to the inevitably existing inaccuracies in initial conditions and deficiencies in numerical models. Due to the chaotic nature of the numerical model, a small error in the initial condition can grow exponentially and eventually make the forecast useless (Lorenz 1963; 1969). Incomplete knowledge of the dynamical and physical equations of the atmosphere, and further approximations in numerics make the model trajectory diverge from the true state even if the initial condition is perfect. Since numerical forecasts are inherently uncertain, a forecast is incomplete unless it is accompanied with a prediction about its uncertainty and forecasts are more appropriately expressed in a probabilistic framework. Such additional information significantly expands the usage of the forecast.

Probabilistic forecasts could be ideally generated by propagating the probability density function (pdf) of the state through model dynamics such as the Liouville and Fokker-Planck equations (e.g. Epstein 1969; Ehrendorfer 1994a, b). However, it is too computationally expensive for numerical weather prediction (NWP) models. A computationally feasible approach to estimate the evolution of the pdf is through *ensemble forecasting*, where an ensemble of forecasts can be generated by integrating a numerical forecast model from distinct initial conditions that are consistent with the uncertainties in the initial condition and/or by using multiple models or model

configurations to represent the uncertainties in the computational representation of the equations that govern atmosphere motion. The uncertainty of the forecast is represented either by the dispersion of the ensemble forecasts or the forecast probabilities generated by using the relative frequencies of events of interest in the resulting collection of forecasts. Since ensemble forecasting is recognized as a practical way to provide probabilistic forecasts (Leith 1974), ensemble forecasting has undergone dramatic development. It has been operationally implemented for medium-range numerical weather prediction (e.g., Toth and Kalnay 1993,1997; Molteni et al. 1996; Houtekamer et al. 1996) and is also being used for short-range weather prediction (e.g., Hamill and Colucci 1997, 1998; Du et al. 1997; Stensrud et al. 1999; Hou et al. 2001; Gritit and Mass 2002; Stensrud and Yussouf 2003). It is found in these studies that compared to a single deterministic forecast with relatively high resolution, ensemble mean forecast by averaging ensemble members with relatively low resolution can have smaller root mean square errors and the ensemble forecast can provide flow-dependent forecast uncertainty information in advance. Recent studies (e.g., Richardson 2000; Zhu et al. 2001; Palmer 2002, Roulston et al. 2003) have demonstrated that the economic value of ensemble forecasts is greater than a single deterministic forecast for a wide range of weather forecast users.

One active research topic in ensemble forecasting is, for given computing resources, how to initialize ensembles with limited samples to effectively represent the initial condition uncertainty. So far three strategies have been adopted in major operational meteorological centers. The European Centre for Medium-Range Weather Forecast (ECMWF) uses a singular vector method (Molteni et al. 1996) to generate initial

perturbations that will grow rapidly during the first few days of the forecast. The National Centers for Environmental Prediction (NCEP) uses a breeding method (Toth and Kalnay 1993,1997) where initial perturbations are generated in directions where forecast errors have grown rapidly over previous data assimilation cycles. Initial perturbations generated by the singular vector method and the breeding method are then added to the initial analysis to generate perturbed initial conditions. The Canadian Meteorological Centre (CMC) uses a perturbed observation approach (Houtekamer et al. 1996) where an ensemble of analyses is generated by updating sets of first-guess forecasts with distinct sets of observations. The first goal of this thesis is to introduce and test a new initial perturbation generation scheme, called the ensemble transform Kalman filter (ETKF). This scheme solves the error covariance update equation for a Kalman filter data assimilation scheme (e.g., Kalman 1960; 1961) within the subspace of ensemble perturbations. The ETKF was first introduced by Bishop et al. (2001) as an adaptive sampling technique. As an ensemble generation scheme it is similar to the breeding scheme in that it creates analysis perturbations from forecast perturbations and is inexpensive to run for small ensemble sizes (<100). Unlike the breeding scheme, it explicitly accounts for the effect of observations on analysis error variance and, in the limit of very large ensemble size, converges to the theoretically optimal error covariance update procedure.

In operational singular vector and breeding ensembles, initial perturbations are constructed to let half of the perturbations to be the negative of the other half, so that the sum of the perturbed initial conditions is equal to the analysis. This procedure of centering the initial perturbations about the analysis is desirable as one wants the

ensemble mean to always be equal to the minimum error variance estimate of the true state, which at initial time is the analysis. The second goal of this thesis is to introduce a new centering scheme, called spherical simplex centering, and compare it with the traditional positive-negative pair centering by applying both to the ETKF initial perturbation generation framework.

As the purpose of ensemble forecasting is to access the uncertainty associated with numerical weather prediction, ensemble forecast members should be realistically diverse so that the true atmospheric state acts just like one of the ensemble members. However, it is often observed that observations fall outside the range of the ensemble members with a margin and frequency that cannot be explained by the estimates of observation errors. Presumably, this is because all current ensemble techniques partially misrepresent the effects of initial condition error and model error (e.g., Orrell et al. 2001; Smith 2001, Buizza et al. 1999; Palmer 2001) on forecast accuracy. Thus it is inevitable that members of dynamic ensembles are not drawn from the same distribution as the distribution of truth given an ensemble forecast.

To improve the reliability of the ensemble, one can try to further develop the initial perturbation generation scheme, improve the model, incorporate stochastic effects (e.g., Buizza et al 1999), adopt the multi-model/parameterization/configuration method (e.g., Evans et al. 2000; Fritsch et al 2000; Krishnamuri et al. 2000; Mylne et al. 2002; Richardson 2001; Wandishin et al. 2001; Houtekamer et al 1996; Stensrud et al. 2000; Gritmit and Mass 2002), and statistically adjust the output of ensemble forecasts (Du et al. 2000; Hamill and Colucci 1997, 1998; Eckel and Walters 1998; Atger 1999, 2003; Krzysztofowicz and Sigrest 1999; Wilks 2002; Hamill et al 2004; Raftery et al. 2003;

Roulston and Smith 2003). Of all these options, statistically postprocessing ensemble forecasts is of particular interest of this thesis. Motivated by the previously proposed best member dressing method by Roulston and Smith (2003), the third goal of this thesis is to introduce and test a new ensemble augmentation method to improve the reliability of the spread of the ensemble in the postprocessing.

Given the large amount of information from statistically calibrated ensemble forecasts, how should customers use them? For weather-related commercial users, probabilistic forecasts of meteorological weather elements are not directly useful. The weather-related quantities that the end-users are interested in may not depend linearly on just one meteorological variable, but nonlinearly on a number of meteorological variables in general. Therefore, ensemble forecasts should be fed in a quantitative user application model and the resulting output can be used to form probabilistic forecasts of the user relevant economic variable (Palmer 2002). Associated with testing the new dressing kernel, the ETKF ensemble augmented by the new dressing kernel is applied for probabilistic forecasts of cooling-degree-days (CDD), a frequently used quantity for weather derivative and insurance users.

In chapter 2, the ETKF initial perturbation method is introduced and compared with the breeding method. This work is published in the *Journal of Atmospheric Sciences*, Vol. 60, Issue 9, May 2003. In chapter 3, the spherical simplex initial perturbation centering method is introduced and compared with the positive-negative paired centering by using the ETKF framework. This work is published in the *Monthly Weather Review*, Vol. 132, Issue 7, July 2004. In chapter 4, a problem with the best member method (Roulston and Smith 2003) is revealed and a new dressing method to

statistically augment ensemble forecasts is introduced and tested with the ETKF ensemble. The new dressing method is further tested by applying it to probabilistic forecasts of CDD for Boston. Concluding remarks and future work are discussed in chapter 5. Basic concepts for data assimilation and ensemble forecasting are listed in appendix B.

Chapter 2

A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes

Reprint found in pocket.

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.

Chapter 3

Which Is Better, an Ensemble of Positive-Negative Pairs or a Centered Spherical Simplex Ensemble?

Reprint found in pocket.

Wang, X., C.H. Bishop and S. J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble. *Mon. Wea. Rev.*, **132**, 1590-1605.

Chapter 4

Improvement of Ensemble Reliability with a New Dressing Kernel

4.1 Introduction

During the last decade, ensemble forecasting has become an important part of numerical weather prediction (NWP). It has been operationally implemented for medium-range NWP (e.g., Molteni et al. 1996; Toth and Kalnay 1993,1997; Houtekamer et al. 1996) and is being incorporated to short-range NWP (e.g., Hamill and Colucci 1997, 1998; Du et al. 1997; Stensrud et al. 1999; Hou et al. 2001; Gritmit and Mass 2002; Stensrud and Yussouf 2003). Compared to a single deterministic forecast with high resolution, ensemble mean forecasts with relatively low resolution for each ensemble member can produce smaller root mean square errors. Moreover, ensemble forecasts can provide *flow-dependent* estimates of forecast errors depicted by ensemble spread or expressed in forecast probabilities (e.g., Toth et al. 2001; Whitaker and Loughé 1998). Studies by Richardson (2000); Zhu et al. (2001); Palmer (2002) and Roulston et al. (2003), amongst others, have demonstrated that the economic value of ensemble forecasts is greater than a single deterministic forecast for a wide range of weather forecast users.

Managers of weather sensitive activities can benefit from probabilistic forecasts that accurately represent the probability distribution of the verifications given the ensemble forecast (e.g., Palmer 2002). However, because of the sub-optimal initial perturbation generation techniques and the lack of consideration of model errors, it is

typically shown by rank histogram (e.g., Hamill and Colucci 1997; 1998) that outputs from raw ensembles may be biased and under-dispersive, which limits the predictive power of the ensemble. Hence, developing postprocessing methods to calibrate the outputs of ensemble forecasting systems has also been of interest. Some postprocessing studies involve directly calibrating the forecast probabilities. Methods include reliability diagram statistics (e.g., Zhu et al. 1996; Toth et al. 2001; Krzysztofowicz and Sigrest 1999; Atger 2003), verification rank histogram statistics (Hamill and Colucci 1997, 1998; Eckel and Walters 1998), Bayesian averaging (Kass and Raftery 1995; Raftery et al. 2003), and the logistic regression technique with an ensemble mean as predictor (Hamill et al. 2004). There are also studies to directly postprocess the spread of the ensemble (e.g., Atger 1999; Roulston and Smith 2003).

Of all these postprocessing techniques, the dressing method (Roulston and Smith 2003) is of particular interest in this paper. In the dressing method, statistical perturbations are added to each member of the dynamic ensemble in the postprocessing for the purpose of augmenting the spread of the ensemble. The dressing method provides an alternative to Wilks' (2002) approach to smooth the raw ensemble as one can easily add many dressing perturbations to each member of the dynamic ensemble to produce ensembles with as many members as 10^5 . It also tends to reflect all sources of residual errors that the dynamic ensemble has not yet accounted for. Another advantage of the dressing method relative to other methods that postprocess the spread of the ensemble directly (e.g., Atger 1999) is that the dressing procedure maintains all information of the flow-dependent forecast uncertainty information in the dynamic ensemble. Compared to the calibrated forecast probabilities, the dressed ensemble members can be more

conveniently applied to different types of user application functions, such as the accumulated cooling degree days for weather derivative users show in section 4.5.

In the “best member” dressing method proposed by Roulston and Smith (2003), the best member out of each historical ensemble forecast is first identified and the difference between the best member and the verification, i.e., the best member error, is stored. The archive of the best member errors is built from all historical ensemble forecasts available. When dressing, the statistical perturbations are drawn from the archived historical best member errors. The best member dressing perturbations are straightforward to construct and easy to apply to one- or multi-dimensional variables of interest. Roulston and Smith (2003) demonstrated the superiority of best-member dressed ensembles relative to ensembles constructed by dressing the control forecast with the archived control forecast errors.

To yield reliable probabilistic forecasts of verifying observations, a dressed ensemble should be drawn from the same distribution as the verifying observations given an ensemble forecast (Hereafter “reliable” means when an event is forecast to occur with 40% probability, this event is verified 40% of the time. See also Wilks 1995 p236 for the general definition of reliability.). While Roulston and Smith (2003) demonstrated that best member dressing had some useful properties, this type of dressing approach does not appear to mathematically constrain the distribution of dressed ensemble members to be indistinguishable from the distribution of verifying observations under *any* measure. Notably, if the spread of a dynamic ensemble of finite size were correct, the best member dressing would still dress them thus rendering the dressed ensemble overdispersive.

In sections 4.2 and 4.4 of this paper, we explicitly demonstrate how best member dressing results in distributions that are different from the distribution of verifying observations under second moment measures. In particular, we show that the best-member dressed ensemble may be over-dispersive or under-dispersive depending on, for example, the size of the undressed ensemble, how under-dispersive the undressed ensemble is (section 4.2) and the subspace from which the best member is identified (section 4.4). The prototype test in section 4.2 is based around ensembles generated with a random number generator in which the difference between the distribution of undressed ensemble members and the distribution of verifying observations can be controlled. The test in section 4.4 is based around an ensemble generated using the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004).

In section 4.3, we give the theoretical basis of a new dressing technique that overcomes the limitations of the best member dressing technique and illustrate it using the ensemble generated with a random number generator. In section 4.4, the performance of the new dressing technique is compared against the best member dressing technique for improving the reliability of the 500mb U wind ensemble forecasts from the ETKF ensemble made with the Community Climate Model Version 3 (CCM3; Jeffery et al. 1996). In section 4.5, both dressing techniques are further tested by applying them for the forecasts of a user-specific weather index, the cooling degree days at Boston. Concluding remarks follow in section 4.6.

4.2 Limitations of the best member dressing: the random number generator experiment

In this section we use a simple random number generator experiment to identify the limitations of the best member dressing technique. Let us assume that for each case, a verifying observation y is drawn from a normal distribution with zero mean and standard deviation σ_t ; in other words, assume that $y \sim N(0, \sigma_t)$. As a proxy for an undressed K member ensemble, let us draw K random numbers $x_k, k=1,2,\dots,K$ where each x_k represents a random draw from a normal distribution with a correct mean but an incorrect standard deviation σ_e , in other words we assume that $x_k \sim N(0, \sigma_e)$. For under-dispersive ensembles, $\sigma_e^2 < \sigma_t^2$. For this experiment, we let $\sigma_e^2 = 20$ and let σ_t^2 be greater than σ_e^2 by d , that is, $\sigma_t^2 = \sigma_e^2 + d$.

Training statistics for the best member dressing perturbations are built in the following manner for a given K and d . Step 1: Draw a verification from $N(0, \sigma_t)$ and a K -member undressed ensemble from $N(0, \sigma_e)$. Step 2: Find the ensemble member that is closest to the verification and find its distance from the verification. Step 3: Store this "best member error" in an archive. Step 4: Repeat steps 1-3 M times to obtain an archive of the M best member errors for M cases. Step 5: Compute the sample variance σ_b^2 of the archive of the best member errors. Note that since we required that the undressed ensemble be drawn from a distribution with the same mean as the verifying observations, in this simplified case, the mean of the M archived best member errors is zero when M approaches infinity.

Having obtained this archive of errors, we then generate K independent N -member statistical ensembles of best member errors by either randomly sampling from the archive or by drawing K independent sets of N random numbers $\varepsilon_{kn}, n=1,2,\dots,N; k=1,2,\dots,K$ where $\varepsilon_{kn} \sim N(0, \sigma_b)$. The statistical ensembles are then combined with the dynamical ensemble to create a $N \times K$ member dressed ensemble $\psi_{kn}, k=1,2,\dots,K; n=1,2,\dots,N$ using

$$\psi_{kn} = x_k + \varepsilon_{kn}, k=1,2,\dots,K; n=1,2,\dots,N, \quad (4.1)$$

for each case.

Now note that if the verification were drawn from the same probability distribution as the ensemble then the average square distance between any two randomly selected dressed ensemble members ought to be the same as the average square distance between randomly selected ensemble members and the verification. Consequently, we can test whether the best member dressing results in an ensemble that appears to be drawn from the distribution of the verification for a particular K and d using the following steps. Step 6: Draw a verification from $N(0, \sigma_v)$ and a K -member undressed ensemble from $N(0, \sigma_e)$. Step 7: Using data from the archived best member errors create a dressed $N \times K$ member ensemble as Eq. (4.1). Step 8: repeat steps 6-7 M times to collect M cases. Step 9: Compute the averaged square distance between each distinct pair of dressed ensemble members. Note that since the total number of dressing perturbations is different from the number of undressed ensemble members, from Eq. (4.1) this

quantity is calculated by $\left\langle \left\langle (x_{mk} - x_{mi})^2 \right\rangle_{i \neq k} \right\rangle_m + \left\langle \left\langle (\varepsilon_{mkn} - \varepsilon_{mil})^2 \right\rangle_{kn \neq il} \right\rangle_m$, where subscript

m denotes the m th case of the M cases, $\langle \rangle_{i \neq k}$ is the average over all combinations of distinct undressed ensemble members for the m th case, $\langle \rangle_{kn \neq il}$ is the average over all combinations of distinct dressing perturbations for the m th case, and $\langle \rangle_m$ is the average over all M cases. Step 10: Compute the mean square distance between the verifying observations and each ensemble member by $\left\langle \left\langle (\psi_{mkn} - y_m)^2 \right\rangle_{kn} \right\rangle_m$ where $\langle \rangle_{kn}$ is the

average over all dressed ensemble members for the m th case. Step 11: Compare the difference (denoted as *DIFF*) of the quantities in steps 9 and 10, i.e., calculate

$$DIFF = \left\langle \left\langle (x_{mk} - x_{mi})^2 \right\rangle_{i \neq k} \right\rangle_m + \left\langle \left\langle (\varepsilon_{mkn} - \varepsilon_{mil})^2 \right\rangle_{kn \neq il} \right\rangle_m - \left\langle \left\langle (\psi_{mkn} - y_m)^2 \right\rangle_{kn} \right\rangle_m. \quad \text{Then}$$

repeat the previous steps for different choices of K and d .

Figure 4.1 (a) shows *DIFF* as a function of K and σ_e^2 / σ_i^2 for $M = 10000$, and $N = 100$. Negative (positive) *DIFF* indicates that the dressed ensemble is under-dispersive (over-dispersive). The figure shows that for $K = 1$, *DIFF* is equal to zero for all σ_e^2 / σ_i^2 . When K is larger than 1, for any given σ_e^2 / σ_i^2 , there is only one value of K that renders the best member dressing method reliable. The best member dressed ensemble is either over-dispersive or under-dispersive beyond that regime depending on the undressed ensemble size K and how under-dispersive (measured in σ_e^2 / σ_i^2) the undressed ensemble is. We also observe that when σ_e^2 is equal to σ_i^2 ($\sigma_e^2 / \sigma_i^2 = 100\%$), i.e., the undressed ensemble has the correct spread, the best-member dressed ensemble is

over-dispersive for any finite K . In the next section we introduce a new dressing kernel that does not suffer from these limitations.

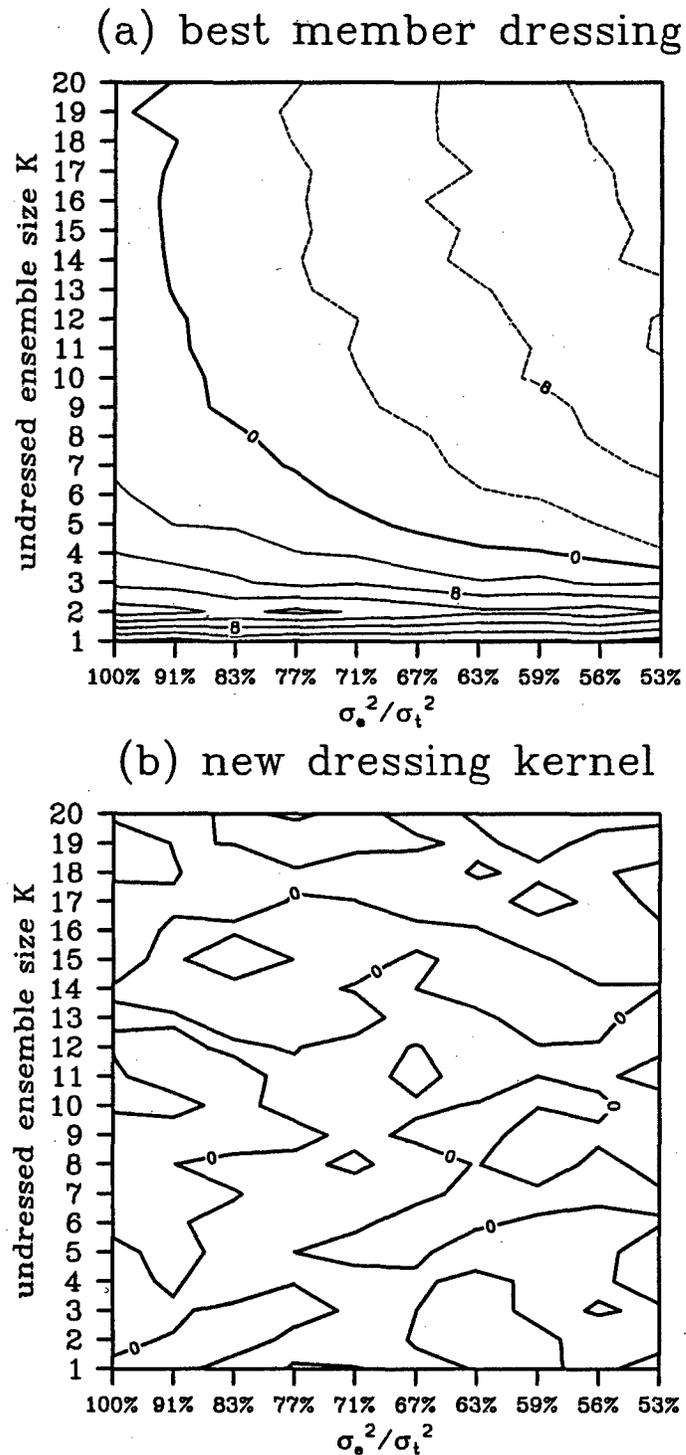


Figure 4.1: Random number generator experiment results in testing the reliability of the spread of the ensemble dressed by (a) the best member method and (b) the new dressing kernel. Thin solid contours indicate over-dispersive ensemble. Dashed contours indicate under-dispersive ensemble. Thick solid contours mean the spread is reliable. Contour interval is 4.

4.3 Dressing with second moment constraint

We seek a mathematical constraint on a dressing kernel that will render the distribution of dressed ensemble members indistinguishable from the distribution of verifying observations given an ensemble forecast on seasonally averaged basis. To measure the differences between the two distributions we will focus on the second moment measure. The new dressing kernel is first built from historical ensemble forecasts and verifications and then applied for the current ensemble forecasts. For each case of forecasts over a season, let \mathbf{y} contain a list of verifications that we wish to predict and let \mathbf{x} contain the corresponding list of forecast variables from one member of the dynamic ensemble. Assume the dynamic ensemble, after removing the seasonally averaged bias, is drawn from an infinite number of realizations of a stochastic process,

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{x}', \quad (4.2)$$

where $\langle \mathbf{x} \rangle = \bar{\mathbf{x}}$ and $\langle \mathbf{x}' \rangle = \mathbf{0}$. The covariance of Eq. (4.2) is denoted as

$$\Sigma^2 = \langle \mathbf{x}' \mathbf{x}'^T \rangle. \quad (4.3)$$

When dressing, statistical perturbations $\boldsymbol{\varepsilon}$ are added to each dynamic ensemble member. Let $\boldsymbol{\psi}$ list the corresponding dressed forecasts. Written in the similar format of Eq. (4.2), the dressed ensemble members are drawn from an infinite number of realizations of a stochastic process

$$\boldsymbol{\psi} = \mathbf{x} + \boldsymbol{\varepsilon} = \bar{\mathbf{x}} + \mathbf{x}' + \boldsymbol{\varepsilon}, \quad (4.4)$$

where $\langle \boldsymbol{\varepsilon} \rangle = \mathbf{0}$, $\langle \boldsymbol{\varepsilon} \mathbf{x}' \rangle = \mathbf{0}$. Note that the mean of the dressed ensemble is still $\bar{\mathbf{x}}$. Also note that we have assumed the seasonally averaged bias of $\bar{\mathbf{x}}$ has been removed. The

basic idea of the new dressing kernel is to chose ϵ to make ψ indistinguishable from the verification, y , under second moment measurement on a seasonally averaged basis.

Mathematically, we require

$$\left\langle \left\langle (\psi_{ii} - \psi_{ij})(\psi_{ii} - \psi_{ij})^T \right\rangle_{i \neq j} \right\rangle_l = \left\langle \left\langle (\psi_{ii} - y_l)(\psi_{ii} - y_l)^T \right\rangle_i \right\rangle_l, \quad (4.5)$$

where subscript l denotes the l th case over a season and subscripts i and j denote any two different dressed ensemble members from Eq. (4.4), $\langle \cdot \rangle_l$ is the averaging over all cases over a season, $\langle \cdot \rangle_{i \neq j}$ denotes averaging over all combinations of any two different dressed ensemble members for the l th case, and $\langle \cdot \rangle_i$ is the averaging over all choices of i for a particular case. Substituting Eq. (4.3) and Eq. (4.4) into Eq. (4.5), one can show (see appendix A) that Eq. (4.5) is satisfied provided that

$$\langle \epsilon \epsilon^T \rangle = \langle (\bar{x}_l - y_l)(\bar{x}_l - y_l)^T \rangle_l - \langle \Sigma^2_l \rangle_l, \quad (4.6)$$

where \bar{x}_l and Σ^2_l are the mean and covariance of the underlying distribution from which the undressed ensemble is drawn for the l th case. Note that the covariance of the dressing perturbations $\langle \epsilon \epsilon^T \rangle$ is the same for all ensemble members for all cases. So, we put no subscript on this term.

To understand the new dressing kernel $\langle \epsilon \epsilon^T \rangle$ given by Eq. (4.6), we use a two-dimensional figure (Figure 4.2) to illustrate the idea. Assume the number of variables that we are interested in forecasting is two, that is, x , y and ψ contain two elements each. Each black dot in Figure 4.2 (a) represents the difference between one of the

members of one ensemble forecast made during a season and the corresponding underlying ensemble mean. Thus, since there are L K -member forecasts made each season, the number of dots present in Figure 4.2 (a) is equal to $K \times L$ and the covariance of these points corresponds to the $\langle \Sigma^2_i \rangle_i$ term in Eq. (4.6). The 1-sigma ellipse associated with this covariance is shown by the black line in Figure 4.2 (a). Each of the L grey dots in Figure 4.2 (b) gives the difference between a verification and a corresponding underlying ensemble mean. The covariance of these dots gives the first term in Eq. (4.6). Since the seasonally averaged bias of undressed ensembles has been removed, the grey dots in Figure 4.2 (b) as well as the black dots in Figure 4.2 (a) center at (0,0). Note that the 1-sigma ellipse for the grey dots is larger than the ellipse for the black dots indicating that the undressed ensemble is under-dispersive. In Figure 4.2 (c) we show how a black dot from Figure 4.2 (a) can be dressed with perturbations drawn from a distribution with a covariance matrix given by the difference between the covariance matrices associated with Figure 4.2 (b) and Figure 4.2 (a). After we dress each black dot of Figure 4.2 (a), in Figure 4.2 (d) we get the scattered stars that are the differences of the dressed ensemble members from the corresponding underlying ensemble mean. The corresponding 1-sigma ellipse associated with the stars is also shown in Figure 4.2 (d). The idea of Eq. (4.6) is to constrain the second moment of the dressing perturbations, the first term on the left side of Eq. (4.6) (i.e., the 1-sigma ellipse in Figure 4.2 (c)), so that the 1-sigma ellipse associated with the covariance of the dressed ensemble perturbations (stars) in Figure 4.2 (d) is identical to the 1-sigma ellipse corresponding to the distribution of the verifications (grey dots) about the underlying ensemble mean in Figure 4.2 (b). Note in Figure 4.2 (d)

that $\langle \langle \Sigma^D \rangle \rangle_t$ denotes the seasonally averaged covariance of the dressed ensemble perturbations. In this way, the dressed ensemble members are indistinguishable from the verification under the second moment over a season.

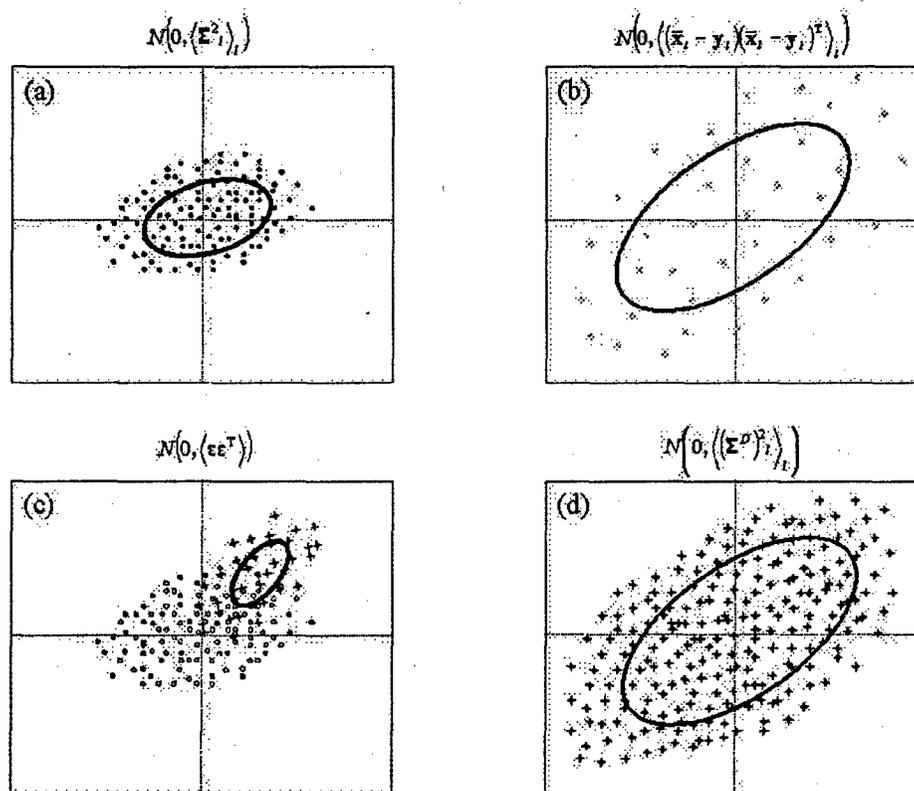


Figure 4.2: Illustration for the idea of the new dressing kernel in 2-dimensional space. Please refer section 4.3 for detailed explanation.

For a finite undressed ensemble size, the underlying ensemble mean and covariance $\bar{\mathbf{x}}_t$ and Σ_t^2 in Eq. (4.6) are estimated using a sample ensemble mean $\bar{\mathbf{x}}_t^s$ and a sample ensemble covariance $\Sigma_t^{s^2}$, namely,

$$\bar{\mathbf{x}}^s_l = \frac{1}{K} \sum_{m=1}^{m=K} \mathbf{x}_{lm}, \quad (4.7)$$

and

$$\Sigma^{s^2}_l = \frac{1}{(K-1)} \sum_{m=1}^{m=K} (\mathbf{x}_{lm} - \bar{\mathbf{x}}^s_l)(\mathbf{x}_{lm} - \bar{\mathbf{x}}^s_l)^T. \quad (4.8)$$

Seasonally averaged bias of the sample ensemble means are assumed to be removed from (4.7) and (4.8). As shown in the appendix A, Eq. (4.6) becomes

$$\langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle = \left\langle \left(\bar{\mathbf{x}}^s_l - \mathbf{y}_l \right) \left(\bar{\mathbf{x}}^s_l - \mathbf{y}_l \right)^T \right\rangle_l - \left(1 + \frac{1}{K} \right) \left\langle \Sigma^{s^2}_l \right\rangle, \text{ for } K \geq 2, \quad (4.9a)$$

in order to satisfy Eq. (4.5). Please see the appendix A for the derivation. In the situation wherein there is only one control forecast \mathbf{x}^c_l for the l th case, that is, $K = 1$, the new dressing kernel is

$$\langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle = \left\langle \left(\mathbf{x}^c_l - \mathbf{y}_l \right) \left(\mathbf{x}^c_l - \mathbf{y}_l \right)^T \right\rangle_l, \text{ for } K = 1. \quad (4.9b)$$

Note for $K = 1$, the new dressing kernel and the best member dressing kernel are the same. To test the new dressing kernel, we also adopt the random number generator experiment with the same procedures and the same measure as in section 4.2 except the 1-dimensional new dressing kernel is given by Eq. (4.9a) and Eq. (4.9b). The result is shown in Figure 4.1 (b). Figure 4.1 (b) demonstrates that the new dressing kernel Eq. (4.9a) and Eq. (4.9b) can provide a reliable ensemble spread for all K and σ_e^2 / σ_l^2 under the second moment measure given in step (11) of the random number generator experiment in section 4.2.

Also note that when K is greater than one but rather limited, the new kernel defined by Eq. (4.9a) only makes the dressed ensemble satisfy the second moment property that the seasonally averaged covariance of the differences between ensemble

members and the verifications is equal to seasonally averaged covariance of the differences between ensemble members. It does not make the dressed ensemble satisfy the second moment property that the seasonally averaged covariance of the differences of the ensemble from the *sample* ensemble mean is equal to the seasonally averaged error covariance of the *sample* ensemble mean. The latter property can be obtained by replacing Eq. (4.9b) with

$$\langle \mathbf{\varepsilon} \mathbf{\varepsilon}^T \rangle = \left\langle \left(\bar{\mathbf{x}}^s_l - \mathbf{y}_l \right) \left(\bar{\mathbf{x}}^s_l - \mathbf{y}_l \right)^T \right\rangle_l - \left(1 - \frac{1}{K} \right) \left\langle \mathbf{\Sigma}^{s^2}_l \right\rangle_l, \text{ for } K \geq 2. \quad (4.9c)$$

In other words, these two second-moment properties can not be satisfied simultaneously for smallish K . However, as K tends to infinity, both properties are simultaneously satisfied. When one's forecast application relies solely on the ensemble mean, using Eq. (4.9c) to define the new kernel is probably the best option. In contrast, when one's forecast application relies on a forecast probabilistic distribution, using Eq. (4.9a) to define the new kernel would be the best option. The random number generator experiment (not shown) demonstrates that when $K > 10$, one of the two properties can be satisfied precisely and the other can be satisfied approximately by the new dressing kernel either defined by Eq. (4.9a) or Eq. (4.9c). The best member dressing kernel, however, does not satisfy either second moment property. Note in Figure 4.1 the new dressing kernel is given by Eq. (4.9a,b) and the measure is based on the first second-moment property. Since in the ETKF ensemble experiments to be described in Section 4.4, $K = 16$, the results obtained with Eq. (4.9a) are very similar to those obtained with Eq. (4.9c).

After the new dressing kernel is defined, we use a multi-dimensional random number generator to generate the dressing perturbations. First, note that the covariance matrix given by Eq. (4.9), denoted as \mathbf{Q} hereafter, is real and symmetric but not positive definite. We first perform an eigenvalue decomposition on \mathbf{Q} ,

$$\mathbf{Q} = \langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle = \mathbf{E} \boldsymbol{\Omega} \mathbf{E}^T, \quad (4.10)$$

where the columns of \mathbf{E} contain the eigenvectors and the diagonal matrix $\boldsymbol{\Omega}$ contains the corresponding eigenvalues. Positive eigenvalues indicate that on the directions of the corresponding eigenvectors the ensemble is underdispersive and thus dressing is necessary. In contrast, negative eigenvalues indicate that the undressed ensemble is overdispersive in the directions of the corresponding eigenvectors. Since dressing the ensemble in the overdispersive directions would make it even more overdispersive in this direction, we only dress in the directions corresponding to positive eigenvalues. Based on this argument, we define the new dressing perturbation generator as

$$\boldsymbol{\varepsilon} = x_1 \mathbf{e}_1^+ + x_2 \mathbf{e}_2^+ + \cdots + x_k \mathbf{e}_k^+, \quad (4.11)$$

where \mathbf{e}_i^+ , $i = 1 \cdots k$, are all eigenvectors corresponding to the positive eigenvalues. The coefficients x_i , $i = 1 \cdots k$, are univariate random variables which are parameterized as normal distributions with mean equal to zero and variance equal to the i th positive eigenvalue of \mathbf{Q} , denoted as ω_i^+ . Mathematically,

$$x_i \sim N\left(0, \sqrt{\omega_i^+}\right). \quad (4.12)$$

Note that (4.11) and (4.12) enable the new kernel to generate multi-dimensional dressing perturbations for the multi-dimensional variables of interest at small cost. Also

note that the new dressing kernel is designed not only to provide reliable spread for each individual variable but also to produce a reliable estimate of the error covariance between the variables of interest, which will be shown in sections 4.4 and 4.5. Depending on the variables of interest, the new dressing kernel can be constructed to consider both temporal and spatial correlations of the forecast errors. Thus, the method allows 4-dimensional dressing. The new dressing perturbations can also be drawn from an archive instead of a prescribed distribution. The method by which this can be done is discussed in section 4.6.

4.4 Further test with nonlinear CCM3 ETKF ensemble

The best member dressing method was first designed and tested by Roulston and Smith (2003) with the nonlinear ensemble prediction system of the European Centre for Medium Range Weather Forecasts (ECMWF). The error statistics of nonlinear systems on a given day are usually non-Gaussian. In the random number generator experiment of section 4.2 and 4.3 we assume a Gaussian error system. To check the performance of the best member dressing and the new dressing methods in the nonlinear system with non-Gaussian error statistics, we apply both dressing methods to the 1-10 day CCM3 ETKF nonlinear atmospheric ensemble forecasts (Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004). We also use this section to illustrate the sensitivity of best member dressing to the manner in which one defines the "best ensemble member".

4.4.1 Numerical experiment design

4.4.1.1 Dynamic ensemble, verification data, and variables of interest

The ensemble to be dressed is a 16-member spherical simplex ETKF ensemble of 10-day forecasts. The ensemble is run on the NCAR CCM3 (Jeffery et al. 1996) and the initial conditions for each control forecast are obtained from the NCEP/NCAR reanalysis (Kalnay et al. 1996). The observational network in the current experiment simulates both rawinsonde and satellite observations. For details on the construction of the spherical simplex ETKF ensemble, please refer to previous experiments in Wang and Bishop (2003) and Wang et al. (2004).

The verifications are NCEP/NCAR reanalysis data located on the reanalysis grids that are nearest to known rawinsonde sites. The variables that we are interested in dressing are 500-hPa U over 14 reanalysis grids over the eastern USA (Figure 4.3) at individual forecast lead times. The CCM3 ensemble outputs are interpolated to these grids during the training and validating phases of the experiment.

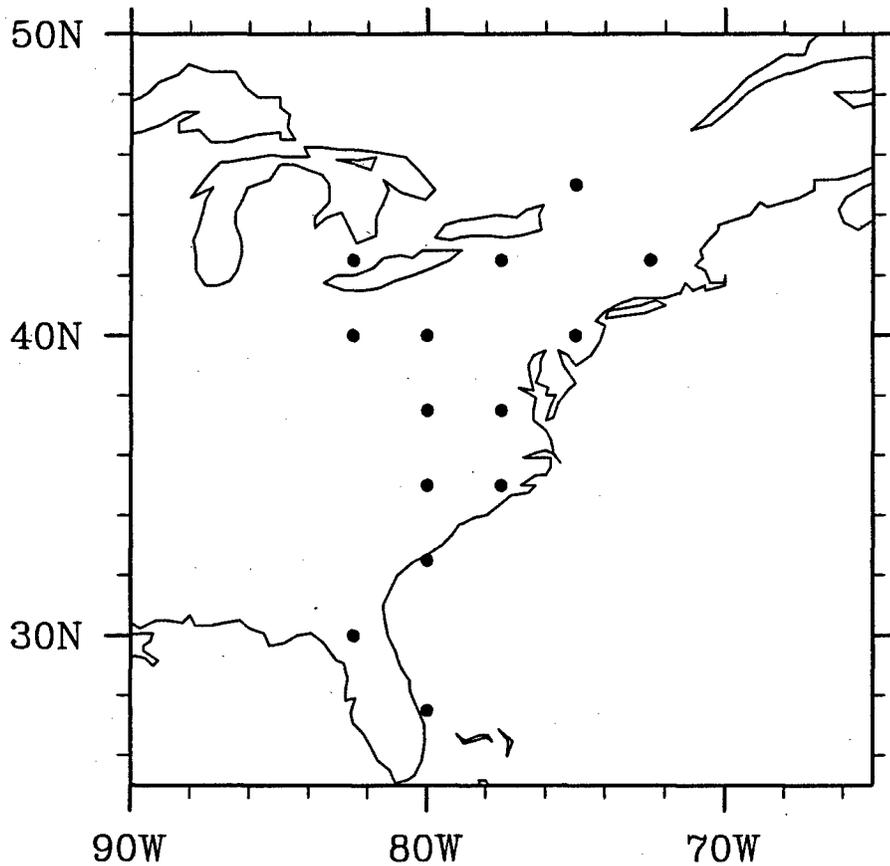


Figure 4.3: 14 verification sites over eastern USA for the experiment in section 4.4.

4.4.1.2 Identification of the best member

In Roulston and Smith (2003), the best member is defined as the closest to the verification in the *full* space including all spatial locations, all quantities and all forecast lead times. However, using the full space to make the identification is time consuming. Roulston and Smith (2003) tried to empirically determine the minimum number of variables that are unlikely to lead to misidentification. They suggested that if practically

feasible, high dimensional space should be used even if the variables that we are interested in dressing are only in a small subspace. To test whether the best-member error statistics with the best member identified in a high dimensional space can provide reliable spread, we use a quite high dimensional space, 500-hPa U over global verification sites throughout 1 to 10 day forecast lead times, to identify the best member, although we are only interested in 500-hPa U wind over the 14 sites for each individual lead time. This subspace for identifying the best member is denoted as RS-10d-globe.

To reveal that the spread of the best-member dressed ensemble may not be reliable due to the uncertainty in selecting the subspace to identify the best member, we also try the experiments where the best member is defined in two relatively low dimensional spaces. One is 500-hPa U over the 14 eastern USA sites for each individual verification lead time, denoted as RS-id-east. The other is 500-hPa U over the 14 sites from day-1 till the verification lead time, denoted as RS-1-id-east. Note the norm of the distance of an ensemble member and the verification used to identify the best member is defined the same way as in equation (1) of Roulston and Smith (2003).

4.4.1.3 Training and forecasting processes

The training statistics for bias and dressing perturbations are obtained from forecasts during the summer (June, July and August) of 1999. The 500-hPa U bias is obtained for each verification site for each forecast lead time by averaging the corresponding ensemble mean errors collected from 16-member ETKF runs during the 1999 summer. Before generating training statistics for the dressing perturbations for both

the new kernel and the best member method, the bias is first removed from each member of the 16-member ETKF ensemble for each verification site and at the 1, 2, ..., etc., 10 day lead times.

Since we are interested in 500hPa U forecasts over the 14 verification sites at individual lead times, the new dressing kernel is constructed for each forecast lead time independently. In Eq. (4.9), vector $\bar{\mathbf{x}}_l^s$ contains 14 elements corresponding to the 500-hPa ensemble mean U forecasts at the 14 sites of the l th case during 1999 summer for each particular lead time. Vector \mathbf{y}_l contains the corresponding verifications and $\Sigma_l^{s^2}$ is the 14×14 ensemble covariance matrix. The resultant \mathbf{Q} matrix is 14×14 .

For the best-member method, the best member out of each 16-member ETKF run during 1999 summer is selected first for the three subspaces. For the subspaces RS-id-east and RS-1-id-east, the best member errors corresponding to 500hPa U over the 14 verification sites are stored in a vector of 14 elements for each lead time. The archive of the best member errors is built by archiving these vectors for each lead time over all runs of 1999 summer. For the subspace RS-10d-globe, the index of the ensemble member that is the best member identified in the subspace of RS-10d-globe is the same for all lead times. In this case, the best member errors are stored in a vector of 140 elements for each 1 to 10-day run. The first 14 elements store the errors of the best member over the 14 sites for 1-day lead time and the second 14 elements store the errors of the same member for 2-day lead time, so on and so forth. The archive of the best member errors for RS-10d-globe is then built by collecting such vectors from all 10-day forecasts over the 1999 summer.

To perform an out-of-sample test of the dressing techniques, forecasts were made for the 2001 Northern Hemisphere summer. For each 16-member ETKF run during 2001 summer, the training bias is first removed from each ensemble member. Independently sampled dressing perturbations are then added to each of the 16 members. For the new dressing kernel, 14-dimensional vectors are randomly generated using Eq. (4.10)-(4.12) for each forecast lead time and added to each member of the 16-member 500hPa U forecasts over the 14 sites. For RS-id-east and RS-1-id-east methods, random 14-dimensional vectors are randomly drawn from the corresponding archives for each forecast lead time. For RS-10d-globe method, random vectors of length 140 are randomly drawn from the corresponding best-member error archive. As mentioned above, the 140 elements contain 10 sets of 14 dimensional vectors corresponding to 1-10 day lead times. The first set of 14 elements is added to the ensemble forecast over the 14 verification sites for 1-day lead time and the second set is added to the same ensemble forecast for the 2-day lead time, so on and so forth.

4.4.2 Experiment results

The performances of the dressed ensembles are measured by the rank histogram and probability scores. For each forecast lead time, samples are collected from all ensemble forecasts during the 2001 summer over the 14 verification sites. For the best-member method, if the dressing perturbations are drawn from the best-member error archive, the number of dressing perturbations to be added to each ETKF ensemble member is limited by the length of the time period during which the best-member error is

collected. As we built the archive from one season's forecasts, the number of best-member dressing perturbations is limited by $o(10)$ in order to make the dressing perturbations for each of the 16 ETKF ensemble members diverse enough. On the other hand, we want the number of dressing perturbations to be large enough so that the probability distribution derived from the dressed ensemble will be smooth and also the ensemble mean whose seasonal average bias is removed will not be shifted due to the sampling error of the dressing perturbations. In our experiment we tried to dress each member of the 16-member ETKF ensemble with 2, 8, 16, and 32 perturbations. We found that the results start to converge when the number of dressing perturbations approaches 16 and 32. The latter renders the sampling error of the dressing perturbation mean to be less than 5%. In the results shown in this section, each member of the 16-member ETKF ensemble has been dressed with 32 perturbations thus yielding 512-member dressed ensembles. For the best member method, the 32 perturbations are drawn from the best member error archive. For the new dressing kernel, the 32 perturbations are drawn from multi-dimensional Gaussian distribution following Eq.(4.10)-(4.12).

The first measurement of the reliability of the ensembles is applicable to scalar verifications and is called the rank histogram (Anderson 1996; Hamill 2001). The rank histogram is constructed by first sorting the K ensemble members for one forecast from the smallest to highest value in the scalar forecast variable of interest – in this case 500hPa U. The values of the sorted ensemble members then define $K + 1$ bins or categories for each case ranked from the lowest to highest. Then record the bin number of the bin that the verification falls in. Repeat the above procedures, for example, for a season's forecasts. The rank histogram gives the frequency with which the verifications

fall into each of the ranked bins. If the dressed ensemble members are being drawn from the same distribution as the verifications, then the verification and the ensemble value defining an edge of a bin would be statistically interchangeable. Therefore in this case the verifications should fall in each bin with equal frequency and the rank histogram is flat. Recall that the sizes of the undressed ensemble and the dressed ensemble are 16 and 512 respectively. Because the number of verifications to construct the rank histogram is limited relative to the rank of 512, and also because we want the y axis of the histogram to have the same scale for the dressed and undressed ensembles, instead of constructing the histogram for the dressed ensemble by using all 512 dressed members we randomly choose 16 members out of 512 members for each sample. Figure 4.4 (a) is the result for the undressed 16-member ensemble for the 2001 summer after removing the bias from the 1999 summer. The undressed ensemble is under-dispersive especially for longer forecast lead times. The χ^2 test for the uniformity of the rank histogram (Anderson 1996; Wilks 1995; Hamill 2001) rejects the null hypothesis that the rank histogram is flat with confidence level much higher than 99% (the P value is equal to 7.1×10^{-4} for day 1 and much smaller than 10^{-10} for 2 to 10 day lead times). After dressing with the new kernel shown in Figure 4.4 (b), the rank histogram becomes much flatter throughout 1 to 10 forecast lead times, which indicates a more reliable ensemble spread. The χ^2 test can not reject the null hypothesis that the rank histogram is flat even with confidence as low as 88% (the P values greater than 0.12). For the RS-10d-globe dressed ensemble in Figure 4.4 (c), the rank histogram is over-dispersive through the 1 to 10 day forecast lead times. The χ^2 test confirms this impression of non-uniformity. The P value is nearly zero

(much smaller than 10^{-10}) for all lead times, indicating the null hypothesis of uniform rank histogram can be rejected with high confidence level (much higher than 99%). For the RS method, where the best member is identified by RS-1-id-east shown in Figure 4.4 (d), the histogram is over-dispersive for 1 to 7 lead times and the χ^2 test rejects the null hypothesis that the rank histogram is flat with confidence level much higher than 99% (the P value is much smaller than 0.0001). Figure 4.4 (e) is the result corresponding to RS-id-east. The rank histogram is over-dispersive for 1 to 2 day lead times and under-dispersive for 8 to 10 day lead times. The χ^2 test confirms the non-uniformity for these 5 lead times by rejecting the hypothesis of uniformity of rank with confidence level much higher than 99% (the P value is much smaller than 0.01).

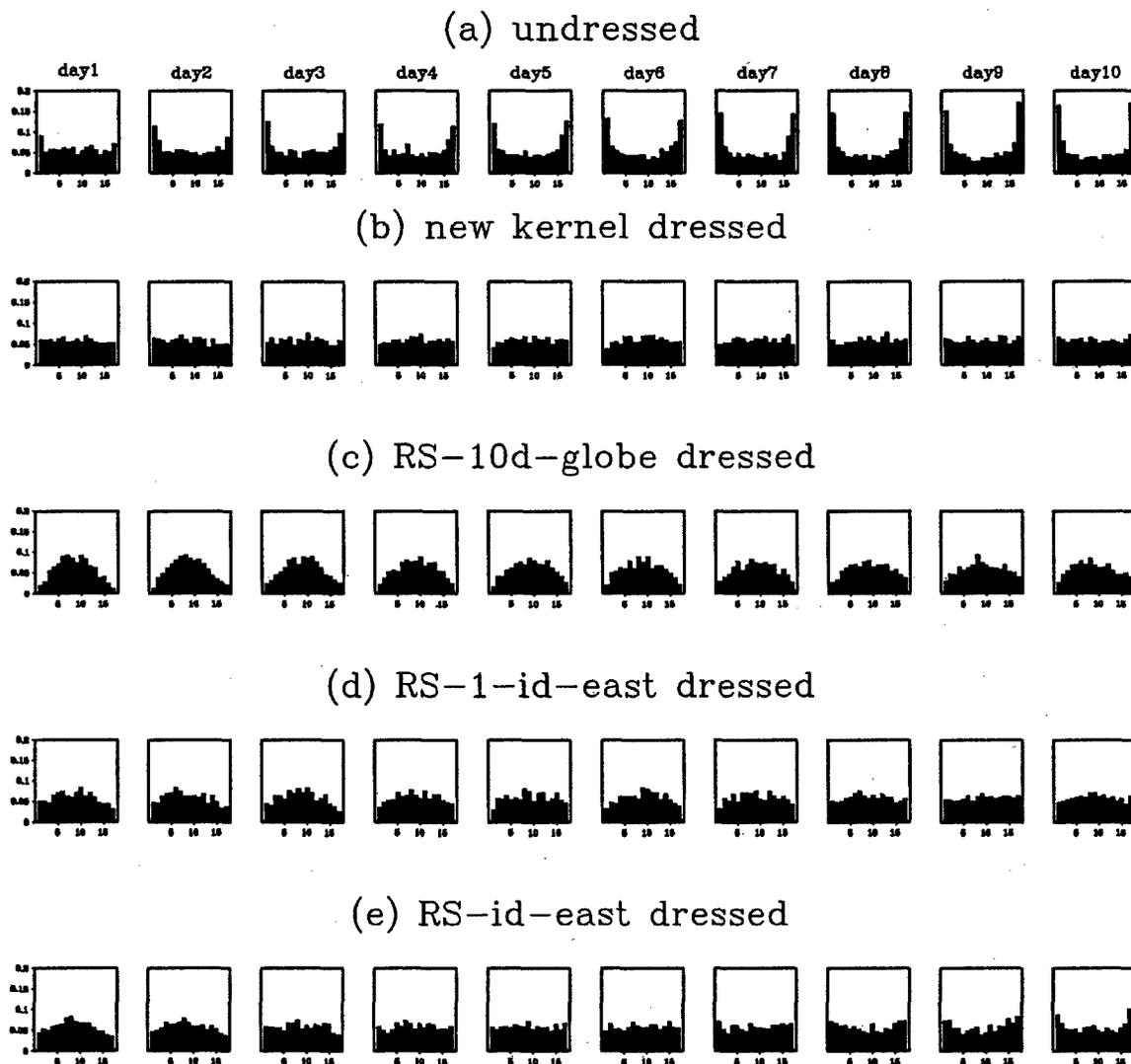


Figure 4.4: Rank histograms for (a) undressed, (b) new kernel dressed, (c) RS-10d-globe dressed, (d) RS-1-id-east dressed and (e) RS-id-east dressed CCM3 ETKF 500hPa U ensembles over the 14 sites from 1-day to 10-day lead times.

In Figure 4.5 we show the Brier score (BS; Brier 1950; Murphy 1973; Wilks 1995) measurement results. Four climatologically equally likely bins are defined by using 1999 summer 500hPa U verifications over the 14 verification sites. For each lead time, the BS is first calculated for each of the 14 sites for each of 92 forecasts of the 2001 summer and then averaged over all 14 sites throughout all of the season's forecasts. The number of samples of BS for each lead time is thus $14 \times 92 = 1288$. The BS corresponding to using the sample climatology, i.e., the uncertainty term when decomposing the BS, is also shown in Figure 4.5. To estimate the significance of the differences between curves, a bootstrap resampling technique (Efron and Tibshirani 1986; Wilks 1995; Hamill 1999; Mullen and Buizza 2001; Roulston and Smith 2003) is used to estimate the $\pm \sigma$ bounds (i.e., standard error) for each curve.

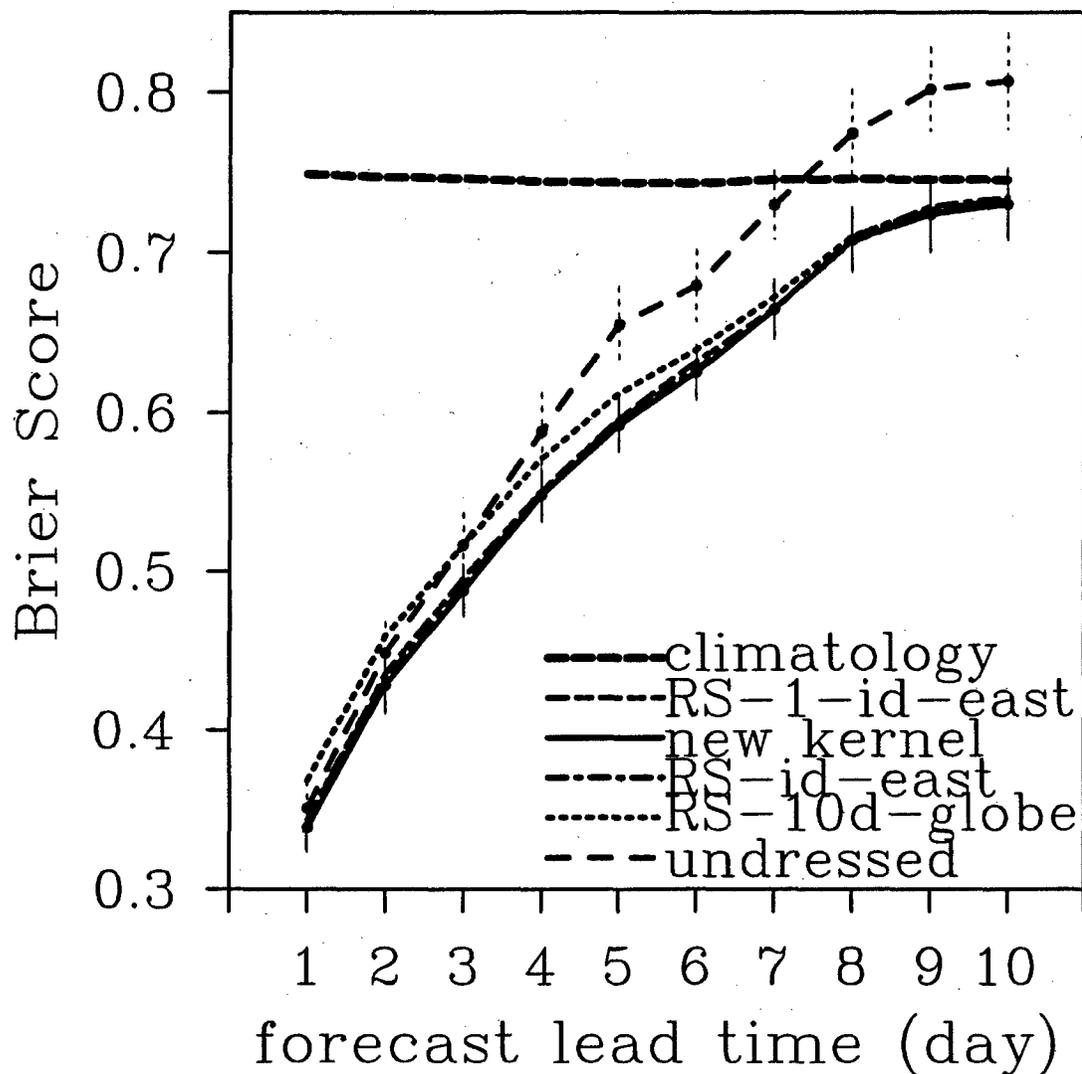


Figure 4.5: Brier scores for the undressed, new kernel dressed, RS-10d-globe dressed, RS-1-id-east dressed and RS-id-east dressed CCM3 ETKF 500hPa U ensembles from 1-day to 10-day lead times. Brier score from the sample climatology is also shown. The vertical solid and dashed lines are the standard errors of Brier score calculation with given samples for the new kernel dressed and undressed ensembles respectively.

Bootstrap is a computer-based method to assess the accuracy of the estimation of an unknown parameter θ by n actual samples. A bootstrap sample is a random sample of size n drawn *with replacement* from the actual n samples. A random number generator is first used to draw a large number (m) of bootstrap samples. For each of the m bootstrap samples, the unknown parameter θ is estimated and the estimated value is denoted as $\hat{\theta}$. The sample standard deviation of the m $\hat{\theta}$ then gives the standard error of the parameter estimated by the actual n samples. In our experiment, the parameter of interest is the mean and we are interested in assessing the accuracy of the sample mean from 1288 BS samples. Note that in bootstrap resampling the n actual samples are required to be independent. Since the 1288 BS samples could be spatially and temporally correlated, before resampling we first estimate the number of independent samples within the 1288 BS samples. Following Roulston and Smith (2003), we divide the total 1288 samples into independent blocks and take the BS's averaged over each block as n actual independent samples. We first divide the 14 sites into groups to ensure that the BS time series averaged over each group are uncorrelated among different groups. We end up having 3 independent groups. Then for each group, we work out the length of the temporal block in a way to ensure that the autocorrelation of the BS time series given by averaging the BS's over each temporal block is nearly zero. After we get the independent samples, 100 bootstrap samples were generated by resampling the independent samples with replacement as recommended by Efron and Tibshirani (1986). These 100 bootstrap samples were used to estimate the $\pm \sigma$ bounds, i.e., the standard error of each curve in Figure 4.5.

In Figure 4.5, $\pm \sigma$ bounds for the curves of the undressed ensemble and the new dressing kernel are shown. From Figure 4.5, the dressed ensemble with new kernel performs better than the undressed ensemble for 1-10 day forecast lead times but only the improvements for 4-10 day forecasts are statistically significant. It is also better than the best member dressed ensemble RS-10d-globe for 1-10 day lead times with significance for 1-2 day lead times. The RS-10d-globe ensemble is worse than the undressed ensemble for 1-2 day lead times. The RS-10d-globe ensemble is significantly better than the undressed ensemble for 5-10day lead times. The scores for the best member dressed ensembles, RS-id-east and RS-1-id-east, are statistically indistinguishable from the new kernel dressed ensemble. Note that RS-10d-globe has worse BS than both RS-id-east and RS-1-id-east, which is inconsistent with the argument from Roulston and Smith (2003) that full space or high dimensional space should be used to identify the best member. To explain why the RS-10d-globe ensemble is worse than the RS-id-east and RS-1-id-east ensembles, we first notice that the error variance of the best member defined in RS-10d-globe is only 10% smaller than the worst member. In other words, all members can be regarded as "the worst" or "the best" if identified in such high dimensional space.

We also tried (not shown) the continuous ranked probability score (CRPS; Hersbach 2000) and the ignorance score (IGN; Roulston and Smith 2002). The comparison results from CRPS and IGN are qualitatively the same as that from BS. Note that in computing these probability scores, the ensemble size for the dressed ensemble is 512, which is much larger than the undressed ensemble size 16. The improvement of the dressed ensemble scores relative to the undressed ensemble scores thus may partly come from the increase of the ensemble size (Richardson 2001; Roulston and Smith 2003).

This is confirmed when we randomly select 16 out of 512 members to calculate the BS for the dressed ensemble. The result (not shown) is qualitatively the same as Figure 4.5, but the improvement of the dressed ensemble relative to the undressed ensemble is smaller than the improvement shown in Figure 4.5.

The rank histogram and Brier score tests only measure the skill of forecasts of individual variables. As discussed in section 4.3, the new dressing kernel is not only able to produce reliable spread for an individual variable but also able to generate a reliable estimate of the error covariance among variables of interest. As discussed in section 4.5, distributions of weather indices that depend on more than one variable are not only sensitive to the forecast error for an individual variable but also sensitive to the covariance of the forecast errors among these variables. As the new kernel is designed to consider the error covariance of variables of interest, it is expected to provide reliable ensemble forecasts for such weather indices. In this section, we first use a simple measure to show that the new kernel can provide a reliable estimate of forecast error covariance. Then in section 4.5, we further demonstrate this property of the new dressing kernel by applying it to the accumulative cooling degree days forecasts, a weather index useful for weather derivative users.

To check the reliability of the dressed covariance estimates, for each forecast at a particular lead time, we first calculate the 500hPa U ensemble covariance between any two of the 14 verification sites. There are 91 pairs among the 14 verification sites. We then average the ensemble covariances collected from the 91 pairs and from all forecasts of the 2001 summer. Then we calculate the product of the ensemble mean errors of any two of the 14 sites and average these products collected from the 91 pairs and from all

forecasts of the 2001 summer. For ensembles that provide a reliable estimate of the forecast error covariance, the averaged ensemble covariance and the averaged ensemble mean error covariance calculated above are equal to each other. In Figure 4.6, we plot the averaged ensemble covariance and the averaged ensemble mean error covariance for 1-10 day lead times. The undressed ensemble covariance underpredicts the ensemble mean error covariance from 4 to 10 day lead times. After dressing with the new kernel, the ensemble covariance matches with the ensemble mean error covariance for 1 to 10 day lead times except that it may overpredict the ensemble mean error covariance at the 6-day lead time. The exception at the 6-day lead time may be due to the limited training data sample size. For example, it might be that there was an extreme event in the training data set that made our kernel too wide. We expect this problem would go away if we had a longer training data set. The RS-10d-globe dressed ensemble covariance overpredicts the ensemble mean error covariance for all lead times. The RS-1-id-east dressed ensemble appears to overpredict the ensemble mean error covariance before 6 day lead time and underpredict the ensemble mean error covariance at 8-9 day lead times. The RS-id-east dressed ensemble covariance underpredicts the ensemble mean error covariance at 7-10 day lead times. The new kernel dressed ensemble covariance is the most reliable among these dressed ensembles.

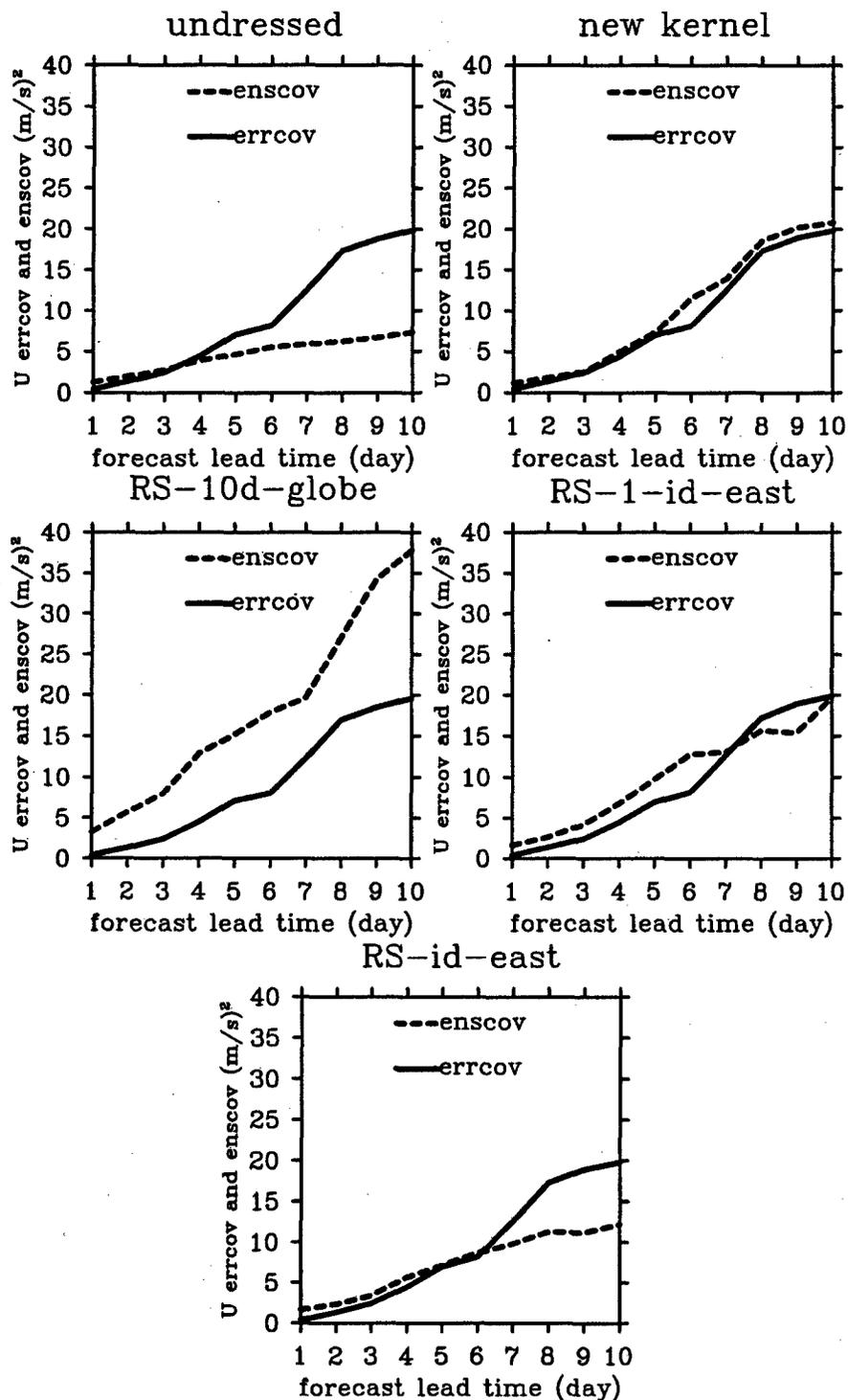


Figure 4.6: seasonal mean 500hPa U ensemble covariance averaged over 91 pairs of sites among the 14 verification sites (dashed) and seasonal mean 500hPa U ensemble mean error covariance averaged over 91 pairs of sites among the 14 verification sites (solid) as a function of forecast lead times for undressed ensemble, new kernel dressed ensemble, RS-10d-globe best member dressed ensemble, RS-1-id-east best member dressed ensemble and RS-id-east best member dressed ensemble.

In summary, our tests with the CCM3 ETKF ensembles show that the performance of the best member dressed ensemble is highly dependent on the choice of subspace used to define the best member and that the new dressing kernel can provide a more reliable estimate of the second moment of the forecast errors than the best member dressed ensembles.

4.5 Application on Cooling degree days forecasts at Boston

In this section we apply and further test the new dressing kernel for forecasting the probability distribution of the accumulated cooling degree days (CDD), a frequently used weather index for weather derivatives. Weather indices such as CDD, depend nonlinearly on multiple meteorological variables, in which case the distribution of CDD is sensitive to both the error variance of individual variables and the error covariance among these variables. Therefore, CDD provides an appropriate test bed to test the new dressing kernel that is designed to provide reliable estimates of both error variance and error covariance among variables of interest. Another purpose of this section is to show how ensemble forecasts can be fed in a quantitative user application model and how the resulting output can be used to form probabilistic forecasts of the user relevant economic variable (Palmer 2001).

4.5.1 CDD definition

To manage the risks associated with abnormally warm or cool summers, a frequently used weather index is "accumulated cooling degree days" or CDDs for short. (See the web site <http://www.cme.com/prd/wec/abtwthder2766.html> of the Chicago Mercantile Exchange for more information). The accumulated CDD is defined as

$$CDD = \sum_{i=1}^{N_d} \max(0, T_i - 65^\circ F), \quad (4.13)$$

where N_d is the number of days over which the CDD is accumulated (i.e., the contract period) and T_i is the arithmetic average of the daily maximum and minimum 2m temperatures in Fahrenheit on the i th day of the period. Denotations follow Zeng (2000). Note that knowing the distribution of temperature forecast errors on each of the N_d days defining the CDD is *not* sufficient to determine the pdf of CDDs. One must also know how the temperature errors are correlated through time because if a temperature error in the day 2 forecast is positively correlated to temperature errors in the day 1 and day 3 forecasts then the distribution of CDDs will be broader than it would be if there were no such correlation.

4.5.2 Application of dressing

In the following experiment, we only consider samples over one single site, Boston, for one season. In order to increase the number of independent samples, we consider CDD accumulated over only 3 days. (The Chicago Mercantile Exchange's CDD contracts pertain to CDDs accumulated over a month or a season.). There are two ways

to augment the CDD ensemble derived from the 16-member CCM3 ETKF ensemble forecasts for daily 2m temperature. One is to dress CDD ensemble forecasts directly and the other is to dress T_i and substitute the dressed T_i in (4.13). However, if we were to dress CDDs directly we would have to modify our dressing algorithm to account for the fact that CDDs are positive definite. Because of this and because we want to demonstrate how the new dressing technique can account for correlations of temperature errors through time, we choose to dress T_i . Specifically, to obtain a dressed ensemble forecast of the 3-day CDDs, we first dress 1-3 day T_i output from the CCM3 ETKF ensemble and then substitute each of the dressed 1-3 day T_i forecasts for Boston into (4.13). This also demonstrates how to feed ensemble forecasts to user application functions (Palmer 2002).

The CCM3 ETKF T_i outputs are interpolated to the single verification site at Boston. The verifications for CDD and T_i for summer 1999 and 2001 are obtained from the Chicago Mercantile Exchange at <http://www.cme.com/dta/hist>. The training and dressing procedures are similar as described in section 4.4.1 except (a) the bias for T_i is computed from the previous 2 weeks' forecasts; (b) to account for the correlation of errors, the second moment constraint dressing kernel is built by simply placing 1-3 day T_i forecasts for Boston and the corresponding verifications in sample vectors with size of 3 elements when constructing the terms in Eq. (4.9); (c) the subspace to identify the best member is over Boston from 1 to 3 day lead times and thus the best member error samples for T_i , $i=1,2,3$ is stored in 3-element vectors for archiving the best member errors; and (d) to overcome the limitation that the number of best-member dressing

perturbations drawn is limited by the length of time period during which the best member error archive is built, the best member dressing perturbations are drawn from a zero-mean multi-dimensional (3-dimensional in this case) normal distribution whose covariance is consistent with the covariance of the archived best member errors.

4.5.3 Results on the reliability of the dressed CDD ensemble spread

Figure 4.7 shows the reliability of the spread of the accumulated CDD ensembles measured by the rank histograms. The figure shows that the undressed CDD ensemble underpredicts the CDD forecast uncertainty. After dressing with the best member method, it is still underdispersive. In comparison, the new dressing kernel can provide reliable spread for the 3-day accumulated CDD forecasts. Note that the number of realizations of verifications for one season's forecasts over a single site is limited for constructing the rank histogram if using all ensemble members as ranks. To overcome this problem, as in section 4.4.2 we randomly choose a relatively small number of ensemble members out of all ensemble members to define the ranks for the rank histogram. The result shown in Figure 4.7 corresponds to the case where we randomly choose 3 members out of 4096 dressed ensemble members to build 4 ranks for each ensemble forecast. Also note for situations where the verification exactly equals some of the ensemble members, such as CDD forecasts of zero and a verification of zero, the number of members (m) equal to the verification was first counted. Then we assigned uniform random numbers between 0 and 1 to the m members and the verification. The m members are ordered according to the assigned random numbers. The rank of the

verification is then determined by the rank of the random number assigned to the verification among the m random numbers assigned to the m tied ensemble members. This is similar to the method for constructing rank histogram for precipitation discussed in Hamill and Colucci (1997). The χ^2 test for the uniformity of the rank histogram confirms the flatness of rank histogram of the new kernel dressed CDD ensembles (P value as large as 0.74) and the unflatness of those of the undressed (P value as small as 0.0001) and the best member dressed CDD ensembles (P value as small as 0.02). The underdispersion of the best member dressed ensemble indicates that the best member dressing kernel is either failing to provide reliable error variance estimates for individual T_i and/or it cannot reliably represent the temporal correlation of forecast errors. We also measure the skills of the CDD ensembles with the ignorance score. Four climatologically equally likely categories are built from 2001 summer CDD verifications on Boston. The results of the ignorance scores for the CDD ensembles are shown in Figure 4.8. The smaller the score, the less ignorant of the CDD probabilistic forecast. Statistical t test (Ott 1993) shows that the ignorance score for the new kernel CDD ensemble is significantly better than those of the best member CDD ensemble and the undressed CDD ensemble. So the probabilistic CDD forecast generated from the CDD ensemble augmented by the new dressing kernel is more skillful than the undressed CDD ensemble and the best member dressed CDD ensemble. The Brier score measurement (not shown) provides qualitatively the same result.

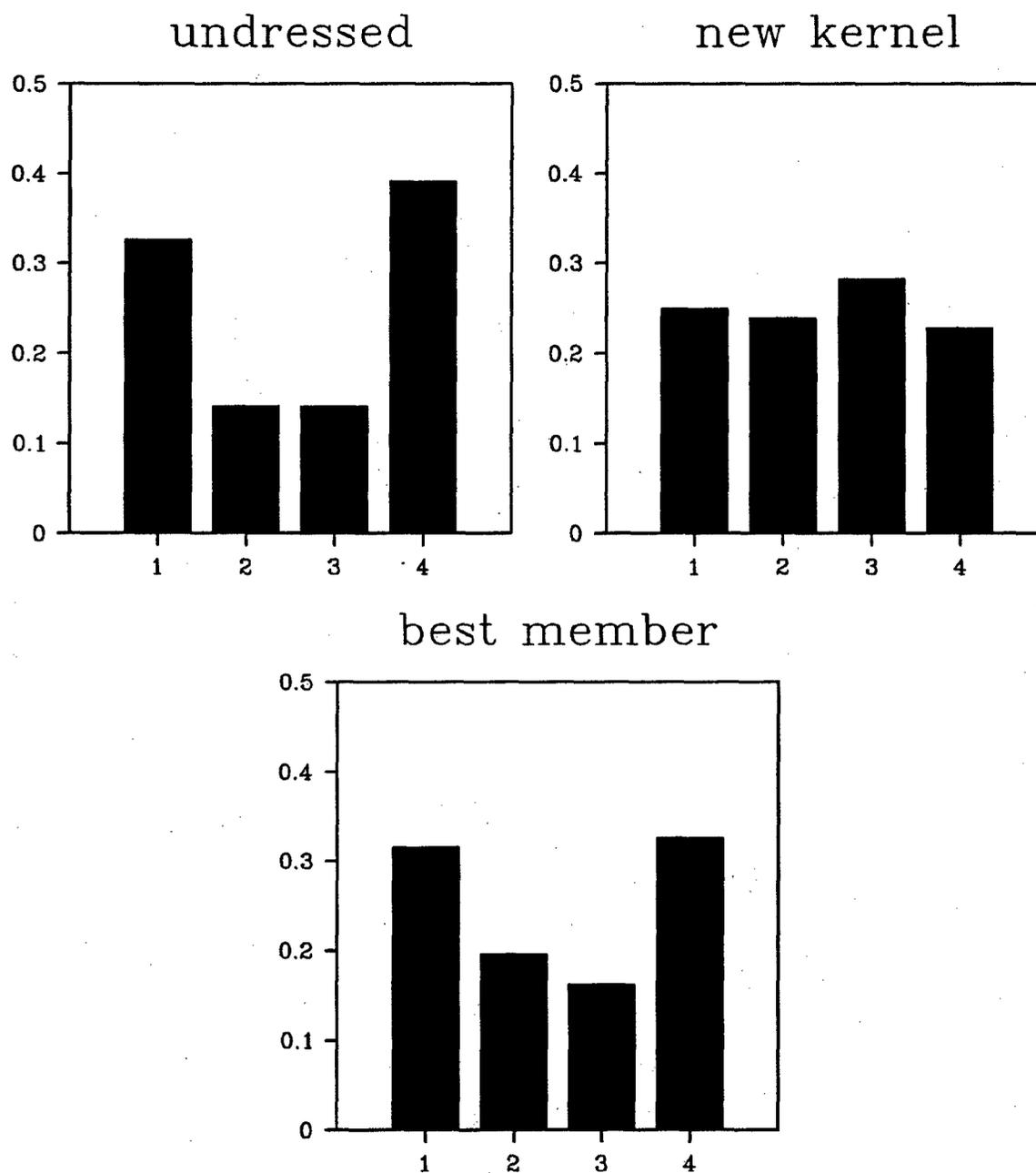


Figure 4.7: Rank histograms for undressed, new kernel dressed, and the best member dressed 3-day accumulated CDD ensembles over Boston during 2001 summer.

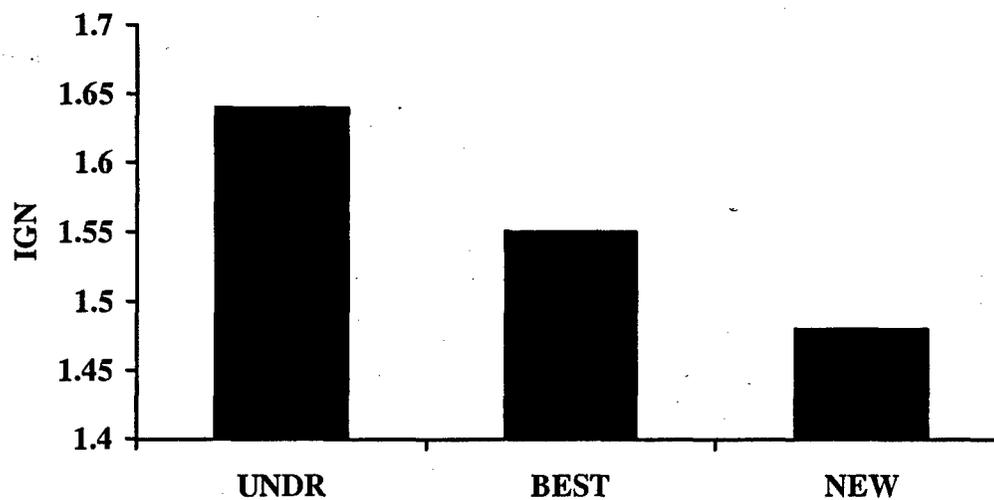


Figure 4.8: Ignorance scores for the undressed (UNDR), the best member dressed (BEST) and the new kernel (NEW) dressed CDD ensembles.

4.6 Conclusion

A new multi-variate dressing method is designed to make the distributions from which dressed ensemble members are drawn indistinguishable from the distribution from which verifying observations are drawn under a seasonally averaged second moment measure. Ensemble bias is removed first before building training statistics for the

dressing kernel and before dressing the current ensembles. The CCM3 ETKF ensemble dressed with the second moment constraint method is more skillful than the corresponding undressed ETKF ensemble. With both a random number generator experiment and the CCM3 ETKF ensemble framework, Roulston and Smith's (2003) original best member dressing method was compared with the second moment constraint dressing method. It was found that the spread of the best member dressed ensemble can be over-dispersive or under-dispersive depending on such factors as the undressed ensemble size, how under-dispersive the undressed ensemble is and the subspace from which the best member is identified. In contrast, the ensembles dressed with the second moment constraint dressing kernel always gave about the right amount of dispersion.

The utility of the second moment constraint dressing relative to the best member dressing and the importance of accurately accounting for the temporal correlation of forecast errors was demonstrated by comparing predictions of accumulative cooling degree days from undressed, second moment dressed and best member dressed ensembles. It was found that the new second moment constraint dressing kernel provided a 3-day accumulated CDD ensemble with more reliable spread and better skill than the CDD ensemble augmented with the best member dressing kernel.

In sections 4.3 and 4.4 of this paper, the dressing perturbations for the new kernel were drawn from a multi-variate normal distribution. As in the best member method, the dressing perturbations for the new kernel can also be based on an archive of past errors rather than a prescribed distribution. This is achieved by first grouping the historical errors of all ensemble members and then transforming these errors by premultiplying a matrix so as to make the covariance of the transformed errors to be equal to the \mathbf{Q} matrix

in Eq. (4.10)–(4.12). In our experiment, dressing with the archive and the prescribed distribution produce similar results for 500hPa U and 2m T. So we only show the results corresponding to the prescribed distribution. Also note that the assumption of a Gaussian dressing kernel is likely poor for positive-definite quantities such as precipitation and 10 m wind speed. To extend the usage of the new dressing kernel for such quantities, a possible option is to transform such quantities to make the transformed quantities have more Gaussian type of distributions (Wilks 2002).

In the new dressing method, no dressing is performed for directions where the undressed ensemble is already overdispersive (Eq. (4.10)–(4.12)). For future work, to correct the directions where the undressed ensemble is overdispersive, we can try to dress each ensemble member differently. A possible solution would be to dress the central members with more dressing perturbations than the outside members so that the pdf of the dressed ensemble is narrower than the undressed ensemble.

Given large enough data sets, it would be of interest to condition the dressing kernel on flow regimes known to have profound impacts on model error. For example, different dressing kernels might be used on convectively stable and unstable days and they may also be constructed to be regionally dependent. The dressing technique and the model output statistics (MOS) technique can also be integrated. For example, we can first apply MOS for each member of the dynamic ensemble and then perform dressing on the ensemble processed by MOS.

Chapter 5

Concluding remarks and remaining challenges

In the thesis, a new initial perturbation generation method, the ensemble transform Kalman filter (ETKF), has been developed and compared with the breeding method. With a little more computational expense, the ETKF provides a significantly more skillful ensemble generation scheme to sample initial condition uncertainty than the breeding method. A new initial perturbation centering scheme called spherical simplex centering was also introduced and found to provide a more useful ensemble than that obtained from the traditional positive/negative paired centering. A second moment constraint “dressing” method to postprocess ensemble output is explored and found to improve the reliability of ensemble spread better than the previously proposed best member dressing method. The statistically augmented spherical simplex ETKF ensemble was then applied for the probabilistic forecasts of a user specific quantity, the cooling degree days (CDD). It was found that the ETKF ensemble dressed with the second moment constraint method provides reliable spread for CDD forecasts and has better skill than the CDD ensemble augmented by the best member method.

The low-rank estimates of the error covariance by ETKF because of the limited ensemble members may cause the magnitude of the ETKF initial perturbations to be too small. Besides ameliorating such deficiency by using an inflation factor as in chapter 2, a bias ameliorated form of ETKF formulation is being developed and tested.

Future research will address the question of how to optimally combine statistically postprocessed ensemble forecast systems from different operational centers to provide probabilistic forecasts. Bayesian averaging may provide a theoretical foundation for such problems. Furthermore, as the ensemble output is often at a grid resolution incapable of resolving some key aspects of severe weather, there is also a need to develop postprocessing methods to downscale ensemble output from the grid resolution to the weather impacted sites of interest. A feasible option would be to apply the Model Output Statistics (MOS) technique to each ensemble member. Another option would be to use an analog downscaling technique on each ensemble member to find the possible subgrid scale states associated with the grid-scale state given by the ensemble member.

Developing methods to effectively evaluate the ensemble is also important. Besides the standard skill scores, evaluating the economic value of the ensemble generated probabilistic forecasts for weather-sensitive users has become a standard way to assess the quality of the ensemble forecasts. We are currently exploring and testing a new economic evaluation method relevant to weather derivative users.

Research on methods to generate initial conditions continues to be important. To improve sampling of the probability density function (pdf) of initial conditions, ensemble forecasts and data assimilation may need to be coupled. Ensemble based data assimilation provides a natural framework for integrating the two. Although early studies with simulated observations and perfect models (see Hamill 2004 and references therein) have shown promising results of ensemble-based data assimilation, it has not been

demonstrated that these methods can produce superior forecasts to those generated by current operational methods in an operational setting.

Various approaches to account for model errors have been tried such as constructing the ensembles with multi-models and single model with perturbed physics parameters or with different parameterization schemes. However, these methods are more empirical than theoretically justified. Developing stochastic-dynamic ensembles based on stochastic parameterizations of sub-grid scale processes may provide a first principle basis for accounting for model error. Such effort is just beginning (e.g., Buizza et al. 1999; Palmer 2001).

How many forecasts should be in the ensemble? Given the constraints on time and computing resources, how should ensemble size and model resolution be balanced? Besides assessing the performances of different configurations of ensemble size and model resolution with the standard forecast skill measures, such question may be answered more appropriately with considerations of customers' needs. Challenges also remain for human forecasters who will need to interpret a larger amount of information from ensemble forecasts than the conventional single deterministic forecasts and who may also need to help customers to analyze the risks associated with different weather conditions rather than just to report what the weather is going to be like.

Although this thesis focused on ensemble weather forecasting, ensemble forecast also provides tools in other areas such as targeted observations, data assimilation, climate forecasting and it also has applications in oceanography and other geosciences. As such, ensemble forecasting is likely to remain an active area of research and development for the next few years.

Bibliography

- Anderson, J.L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518-1530.
- Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**, 1941-1953.
- , 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: consequences for calibration. *Mon. Wea. Rev.*, **131**, 1509-1523.
- Bishop, C. H., B. J. Etherton and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.
- Buizza, R., 2001: Accuracy and potential economic value of categorical and probabilistic forecasts of discrete events. *Mon. Wea. Rev.*, **129**, 2329-2345.
- , M. Miller and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **125**, 2887-2908.
- Courtier, P., J.N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-var, using an incremental approach. *Quart. J. Royal Meteor. Soc.*, **120**, 1367-1387.

- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- , 1997: Atmospheric data assimilation. *J. Meteor. Soc. Japan*, **75**, 319-329.
- Du, J., S. L. Mullen and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427-2459.
- , —, and —, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev.*, **128**, 3347-3351.
- Eckel, F.A. and M.K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting*, **13**, 1132-1147.
- Efron, B. and R. Tibshirani, 1986: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, **1**, 54-77.
- Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703-713.
- , 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Mon. Wea. Rev.*, **122**, 714-728.
- Epstein, E. S., 1969 : Stochastic dynamic prediction. *Tellus*, **21**, 739-759.
- 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., 1992: Using the extended Kalman filter with a multiplayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905-17924.
- , 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

- Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model consensus. *Wea. Forecasting*, **15**, 571-582.
- 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.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155-167.
- , 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550-560.
- , 2003: Ensemble based data assimilation: a review. Submitted to *Mon. Wea. Rev.*
- , and 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.
- , J.S. Whitaker and X. Wei, 2004: Ensemble re-forecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, in press.
- , J. A. Hansen, S. L. Mullen, and C. Snyder, 2004: Workshop on ensemble forecasting in the short to medium range. *Bull. Amer. Meteor. Soc.*, in press.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559-570.
- Hou, D., E. Kalnay, and K.K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73-91.

- Houtekamer, P.L., L. Lefaiivre and J. Derome, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.
- Jeffery, T.K., J.H. Hack, B.B. Gordon, B.A. Boville, B. P. Briegleb, D.L. Williamson, and P.J. Rasch, 1996: Description of the NCAR community climate model (CCM3). NCAR Tech. Note NCAR/TN-420+STR, 152pp.
- Kalman, R., 1960: A new approach to linear filtering and predicted problems. *J. Basic Eng.*, **82**, 35-45.
- , and R. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95-105.
- Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.
- Kass, R.E. and A.E. Raftery, 1995: Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.
- Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi and C. E. Williford, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196-4216.
- Krzysztofowicz, R. and A.A. Sigrest, 1999: Calibration of probabilistic quantitative precipitation forecasts. *Wea. Forecasting*, **14**, 427-442.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.
- Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130-141.
- , 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.

- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaxis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.
- Mullen, S.L. and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **129**, 638-663.
- Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595-600.
- Mylne, K. R., 2002: Decision-making from probability forecasts based on forecast value. *Meteorol. Appl.*, **9**, 307-315.
- , R. E. Evans, and R. T. Clark, 2002: Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. *Quart. J. Roy. Meteor. Soc.*, **128**, 361-384.
- Orrell, D., L. A. Smith, J. Barkmeijer, and T. N. Palmer, 2001: Model error in weather forecasting. *Nonlin. Proc. Geophys.*, **8**, 357-371
- Ott, R. L., 1993: *An Introduction to statistical methods and data analysis*. 4th ed. Duxbury Press, 1051pp.
- Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parameterization in weather and climate prediction model. *Quart. J. Royal Meteor. Soc.*, **127**, 279-304.
- , 2002: The economic value of ensemble forecasts as a tool for assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747-774.
- Parish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical -interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

- Peixoto, J.P. and A. H. Oort, 1992: *Physics of Climate*. Springer-Verlag New York, Inc. 520pp.
- Raftery, A.E., F. Balabdaoui, T. Gneiting and M. Ploakowski, 2003: Using Bayesian model averaging to calibrate forecast ensembles. Technical report no. 440. Department of Statistics, University of Washington.
- Richardson, D.S., 2000a: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649-667.
- , 2000b: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847-1864.
- , 2001: Measures of skill and values of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473-2489.
- Ross, Sheldon, 1998: *A first course in Probability*. Prentice Hall, 514pp
- Roulston, M. S. and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653-1660.
- , and —, 2003: Combining dynamical and statistical ensembles. *Tellus*, **55A**, 16-30.
- , D.T. Kaplan, J. Hardenberg, and L.A. Smith, 2003: Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy*, **28**, 585-602.
- Smith, L. A., 2001: *Nonlinear dynamics and statistics (chapter 2)*, Alistair I. Mees (ed.), Birkhauser.

- Stensrud, D. J. and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England. *Mon. Wea. Rev.*, **131**, 2510-2524.
- , D. J., J. Bao, T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective system. *Mon. Wea. Rev.*, **128**, 2077-2107.
- , 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.
- , Y. Zhu, and T. Marchok, 2001: The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 463-477.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729-747.
- 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 S. J. Julier, 2004: Which is better, an ensemble of positive/negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590-1605.
- Whitaker, J.S. and A.F. Loughe, 1998: The relationship between ensemble spread and the ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292-3302.

- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- , 2001: A skill score based on economic value for probability forecasts. *Meteor. Applic.*, **8**, 209-219.
- , 2002: Smoothing forecast ensembles with fitted probability distributions. *Quart. J. Roy. Meteor. Soc.*, **128**, 2821-2836.
- Zeng, L., 2000: Weather derivative and weather insurance: concept, application, and analysis. *Bull. Amer. Meteor. Soc.*, **81**, 2075-2082.
- Zhu, Y., G. Yeengar, Z. Toth, S. M. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th Conf. On Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J79-J82.
- Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73-83.

Appendix A

Derivation on the new dressing kernel

A1 Derivation on equation (4.6)

To derive Eq. (4.6) first note that using Eq. (4.4)

$$\begin{aligned} (\Psi_{li} - \Psi_{lj}) &= (\bar{\mathbf{x}}_l + \mathbf{x}_{li} + \boldsymbol{\varepsilon}_{li} - \bar{\mathbf{x}}_l - \mathbf{x}_{lj} - \boldsymbol{\varepsilon}_{lj}) \\ &= ((\mathbf{x}_{li} - \mathbf{x}_{lj}) + (\boldsymbol{\varepsilon}_{li} - \boldsymbol{\varepsilon}_{lj})) \end{aligned} \quad (\text{A.1})$$

Using (A.1) on the left side of Eq. (4.5) gives

$$\begin{aligned} & \left\langle \left\langle (\Psi_{li} - \Psi_{lj})(\Psi_{li} - \Psi_{lj})^T \right\rangle_{i \neq j} \right\rangle_l \\ &= \left\langle \left\langle ((\mathbf{x}_{li} - \mathbf{x}_{lj}) + (\boldsymbol{\varepsilon}_{li} - \boldsymbol{\varepsilon}_{lj}))((\mathbf{x}_{li} - \mathbf{x}_{lj}) + (\boldsymbol{\varepsilon}_{li} - \boldsymbol{\varepsilon}_{lj}))^T \right\rangle_{i \neq j} \right\rangle_l \quad (\text{A.2}) \\ &= \left\langle \left\langle (\mathbf{x}_{li} - \mathbf{x}_{lj})(\mathbf{x}_{li} - \mathbf{x}_{lj})^T \right\rangle_{i \neq j} \right\rangle_l + \left\langle \left\langle (\boldsymbol{\varepsilon}_{li} - \boldsymbol{\varepsilon}_{lj})(\boldsymbol{\varepsilon}_{li} - \boldsymbol{\varepsilon}_{lj})^T \right\rangle_{i \neq j} \right\rangle_l \\ &= 2 \langle \boldsymbol{\Sigma}_l^2 \rangle_l + 2 \langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle \end{aligned}$$

Note that the covariance of the dressing perturbations $\langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle$ is the same for all ensemble members for all cases. So we put no subscript on this term. Also note that from Eq. (4.4)

$$(\Psi_{li} - \mathbf{y}_l) = (\bar{\mathbf{x}}_l + \mathbf{x}_{li} + \boldsymbol{\varepsilon}_{li} - \mathbf{y}_l). \quad (\text{A.3})$$

Hence the right side of Eq. (4.5) is

$$\begin{aligned}
& \left\langle \left\langle (\boldsymbol{\psi}_i - \mathbf{y}_i)(\boldsymbol{\psi}_i - \mathbf{y}_i)^T \right\rangle \right\rangle_i \\
&= \left\langle \left\langle (\bar{\mathbf{x}}_i + \mathbf{x}_{ii} + \boldsymbol{\varepsilon}_{ii} - \mathbf{y}_i)(\bar{\mathbf{x}}_i + \mathbf{x}_{ii} + \boldsymbol{\varepsilon}_{ii} - \mathbf{y}_i)^T \right\rangle \right\rangle_i \\
&= \left\langle \left\langle ((\mathbf{x}_{ii} + \boldsymbol{\varepsilon}_{ii}) - (\mathbf{y}_i - \bar{\mathbf{x}}_i))((\mathbf{x}_{ii} + \boldsymbol{\varepsilon}_{ii}) - (\mathbf{y}_i - \bar{\mathbf{x}}_i))^T \right\rangle \right\rangle_i \\
&= \langle \boldsymbol{\Sigma}_i^2 \rangle_i + \langle \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \rangle + \langle (\mathbf{y}_i - \bar{\mathbf{x}}_i)(\mathbf{y}_i - \bar{\mathbf{x}}_i)^T \rangle_i
\end{aligned} \tag{A.4}$$

Substituting Eq. (A.1)-(A.4) into Eq. (4.5) gives Eq. (4.6).

A2 Derivation on equation (4.9a)

To derive equation (4.9a), first we start with the first term on the right side of Eq.

(4.6). First note that

$$\begin{aligned}
& \left\langle (\bar{\mathbf{x}}_i^s - \mathbf{y}_i)(\bar{\mathbf{x}}_i^s - \mathbf{y}_i)^T \right\rangle_i \\
&= \left\langle \left((\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i) + (\bar{\mathbf{x}}_i - \mathbf{y}_i) \right) \left((\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i) + (\bar{\mathbf{x}}_i - \mathbf{y}_i) \right)^T \right\rangle_i \\
&= \left\langle (\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i)^T \right\rangle_i + \left\langle (\bar{\mathbf{x}}_i - \mathbf{y}_i)(\bar{\mathbf{x}}_i - \mathbf{y}_i)^T \right\rangle_i
\end{aligned} \tag{A.5}$$

Note in deriving the last step in Eq. (A.5), we use the assumption

$\left\langle (\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i)(\mathbf{y}_i - \bar{\mathbf{x}}_i)^T \right\rangle_i = \mathbf{0}$, which means the difference between the sample ensemble

mean and the underlying ensemble mean does not co-vary with the difference between

the verifications (e.g., observations) and the underlying ensemble mean over seasons'

forecasts. Also recall that $\left\langle (\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}}_i^s - \bar{\mathbf{x}}_i)^T \right\rangle_i = \boldsymbol{\Sigma}_i^2 / K$. Then from Eq. (A.5), the

first term on the right side of Eq. (4.6) can be approximated as

$$\left\langle (\bar{\mathbf{x}}_i - \mathbf{y}_i)(\bar{\mathbf{x}}_i - \mathbf{y}_i)^T \right\rangle_i = \left\langle (\bar{\mathbf{x}}_i^s - \mathbf{y}_i)(\bar{\mathbf{x}}_i^s - \mathbf{y}_i)^T \right\rangle_i - \frac{1}{K} \left\langle \boldsymbol{\Sigma}_i^2 \right\rangle_i. \quad (\text{A.6})$$

Approximate $\boldsymbol{\Sigma}_i^2$ in the last term of Eq. (A.6) and the second term on the right side of

Eq. (4.6) with $\boldsymbol{\Sigma}_i^{s^2}$. Then we get Eq. (4.9a).

Appendix B

A list of basic concepts in data assimilation and ensemble forecasting

1. Atmospheric data assimilation

Atmospheric data assimilation is an objective analysis process that involves a linear combination of observations with a background (or “first guess”) forecast, which is usually a short-term forecast. The purpose of atmospheric data assimilation is to produce a regular, physically consistent four-dimensional representation of the state of the atmosphere from a heterogeneous array of in-situ and remote instruments that sample imperfectly and irregularly in space and time (Daley 1997). Mathematically, the data assimilation process is expressed as

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{W}(\mathbf{y} - H(\mathbf{x}^f)), \quad (\text{B1})$$

where \mathbf{x}^f and \mathbf{x}^a are column vectors containing n elements of forecast and analyzed values on regular model grids, \mathbf{y} is a column vector containing p elements of observed values on observation sites and H is an operator mapping the forecast/analyzed variables on the analysis grids to the observed variables at the observation locations. The vector $\mathbf{y} - H(\mathbf{x}^f)$ is called the innovation vector or observation increment. Matrix \mathbf{W} contains the weights for linearly combining \mathbf{x}^f and \mathbf{y} . In (B1) and the discussion below for simplicity we assume the observations in \mathbf{y} are collected at the validation time of the

short-term forecast \mathbf{x}^f . In operations, observations collected in a certain time window are assimilated. The vector \mathbf{x}^a in (B1) is called the analysis.

2. Optimal data assimilation

In (B1), \mathbf{W} is a $n \times p$ weight matrix for applying the observation increment (i.e., the innovation vector) to correct the background forecast in order to obtain an analysis. For the i th element of \mathbf{x}^a , the i th row of \mathbf{W} contains p weights for linearly combining the i th element of \mathbf{x}^f and the p elements in the innovation vector. The goal in data assimilation research is to find weight \mathbf{W} so that the error variance for each element of \mathbf{x}^a is *minimized*. These weights are called optimal weights and the corresponding data assimilation is called *optimal data assimilation*.

In optimal data assimilation, the optimal weight matrix \mathbf{W} is derived as

$$\mathbf{W} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (\text{B2})$$

and the corresponding error covariance of \mathbf{x}^a , denoted as \mathbf{P}^a , is

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}^f. \quad (\text{B3})$$

For detailed derivations on (B2) and (B3) please refer to (Daley 1991). Here the author gives explanations on terms in (B2) and (B3). First, \mathbf{H} is the tangent linear of H , i.e., $\mathbf{H} = \partial H / \partial \mathbf{x}$. We call \mathbf{H} the linear observation operator. To further understand the terms on the right sides of (B2) and (B3), we first define forecast error vector, $\mathbf{e}^f = \mathbf{x}^f - \mathbf{x}_m^f$, observation error vector $\mathbf{e}^o = \mathbf{y} - \mathbf{x}_o^f$ and observation mapping operator error vector

$\mathbf{e}^h = H(\mathbf{x}_m^f) - \mathbf{x}_o^t$. In these error vector definitions, \mathbf{x}_m^f is the true atmospheric state expressed with forecast variables on the regular model grids and \mathbf{x}_o^t is the true atmospheric state expressed with observed variables on the observation sites. These errors \mathbf{e}^f , \mathbf{e}^o and \mathbf{e}^h are all multi-dimensional random variables. We assume that these errors are unbiased, that is, $\langle \mathbf{e}^f \rangle = \mathbf{0}$, $\langle \mathbf{e}^h \rangle = \mathbf{0}$, $\langle \mathbf{e}^o \rangle = \mathbf{0}$. Hereafter $\langle \rangle$ is a symbol (commonly used in statistics) which means expectation or average over an infinite sample. The elements in \mathbf{e}^f are errors for different variables at the same model grid or same variable on different grids. Correlations among the elements over the same grid and over adjacent grids are not zero. Similarly, correlations among the elements of \mathbf{e}^o and among the elements of \mathbf{e}^h over the same site and over adjacent sites are not zero. The covariance of \mathbf{e}^f , denoted by matrix \mathbf{P}^f , is called the forecast error covariance. Mathematically $\mathbf{P}^f = \langle \mathbf{e}^f (\mathbf{e}^f)^T \rangle$. The matrix \mathbf{R} is called observation error covariance, which is defined as $\mathbf{R} = \langle \mathbf{e}^o (\mathbf{e}^o)^T \rangle + \langle \mathbf{e}^h (\mathbf{e}^h)^T \rangle$. Note in data assimilation the matrix \mathbf{R} not only includes the part associated with the measurement error \mathbf{e}^o but also includes the part associated with the errors in the forward interpolation of the mapping operator \mathbf{e}^h . In data assimilation, it is often assumed that there is no correlation among the three types of errors, that is, $\langle \mathbf{e}^o (\mathbf{e}^f)^T \rangle = \mathbf{0}$, $\langle \mathbf{e}^h (\mathbf{e}^f)^T \rangle = \mathbf{0}$ and $\langle \mathbf{e}^o (\mathbf{e}^h)^T \rangle = \mathbf{0}$. On the right side of (B3) is the analysis error covariance, defined as $\mathbf{P}^a = \langle \mathbf{e}^a (\mathbf{e}^a)^T \rangle$ where

$\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}_m^f$. The diagonal element of \mathbf{P}^a in (B3) gives the minimum error variance of the corresponding analyzed variable associated with the optimal weights (B2). The diagonal elements of the second part on the right side of (B3) are the reduction (or “shrinking”) of forecast error variance due to the optimal assimilation of observations. If we precisely know \mathbf{P}^f and \mathbf{R} (called true forecast error covariance and true observation error covariance), then we can find the optimal weights by (B2) and know exactly what the analysis error covariance associated with the optimal data assimilation by (B3) (called true analysis error covariance associated with the optimal data assimilation scheme) is. We do not, however, know what the true atmospheric state \mathbf{x}^f is and thus we do not know \mathbf{P}^f and \mathbf{R} . A major theme of data assimilation research is to find ways to estimate and approximate \mathbf{P}^f and \mathbf{R} . Data assimilation schemes associated with approximated \mathbf{P}^f and \mathbf{R} are called sub-optimal schemes. This thesis focuses on ensemble forecasting not data assimilation. For readers who are interested in the details of different types of data assimilation schemes please start with Daley (1991, 1997), Evensen (1992,1994), Courtier et al. (1994) and Parrish and Derber (1992).

3. Data assimilation cycle

Figure B.1 illustrates a typical 6-hour data assimilation cycle, that is, the analysis is generated four times a day at synoptic data collection time. At 00Z, a model starts from initial conditions given by a previously completed atmospheric analysis and is integrated for a short (6 hour) forecast. The 6-hour forecast and the observations collected at 06Z are linearly combined by a data assimilation scheme to generate the analysis at 06Z.

Then the model is integrated from the analysis at 06Z. The same procedures are followed at 12Z and so on. At ECMWF and NCEP, one analysis is generated each time and the forecast started with the analysis is called the control forecast. The control forecast initialized at 00Z, 06Z, 12Z and 18Z can be run up to 10-day or longer for the purpose of medium range forecast.

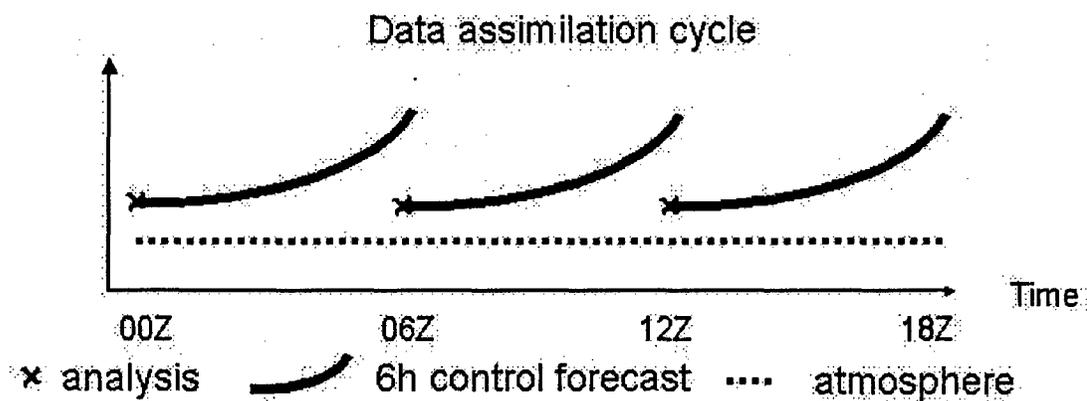


Figure B.1: Cartoon for a typical operational data assimilation cycle.

4. Ensemble forecast cycle

Figure B.2 shows a typical operational ensemble forecast cycle. Given an analysis generated from a data assimilation scheme, there are errors associated with the

analysis. See definitions of "error" in section 2 of this appendix. In ensemble forecasting, an ensemble of initial conditions is generated to sample the possible true atmospheric states around the analysis. The ensemble of initial conditions is called perturbed initial conditions. The difference of a perturbed initial condition and the analysis is called an initial perturbation. One active research topic in ensemble forecasting is how to generate initial perturbations to realistically and effectively sample the possible true atmospheric state around the analysis. Please refer to the introduction for details. The forecast starting from the perturbed initial condition is called the perturbed forecast. The control forecast and the perturbed forecasts together are called the ensemble forecast. For simplicity, Figure B.2 only shows one perturbed forecast. In operations, more than one perturbed initial conditions and thus more than one perturbed forecasts are generated. At operational centers, ensemble forecasts can be generated four times a day with the forecasts initialized at 00Z, 06Z, 12Z and 18Z. The ensemble forecast are then run up to the forecast lead time depending on the purpose of the forecast.

The perturbed forecasts cycle may or may not interact with the data assimilation cycle, depending on the types of data assimilation scheme and initial perturbation generation scheme. Currently at NCEP and ECMWF, the process of generating initial perturbations does not interact with the process of data assimilation. For the ETKF initial perturbation generation method described in chapter 2 and 3, these two processes do not interact either. At the Canadian Meteorological Center, these two processes do interact. Please see discussions in the introduction and references therein for details.

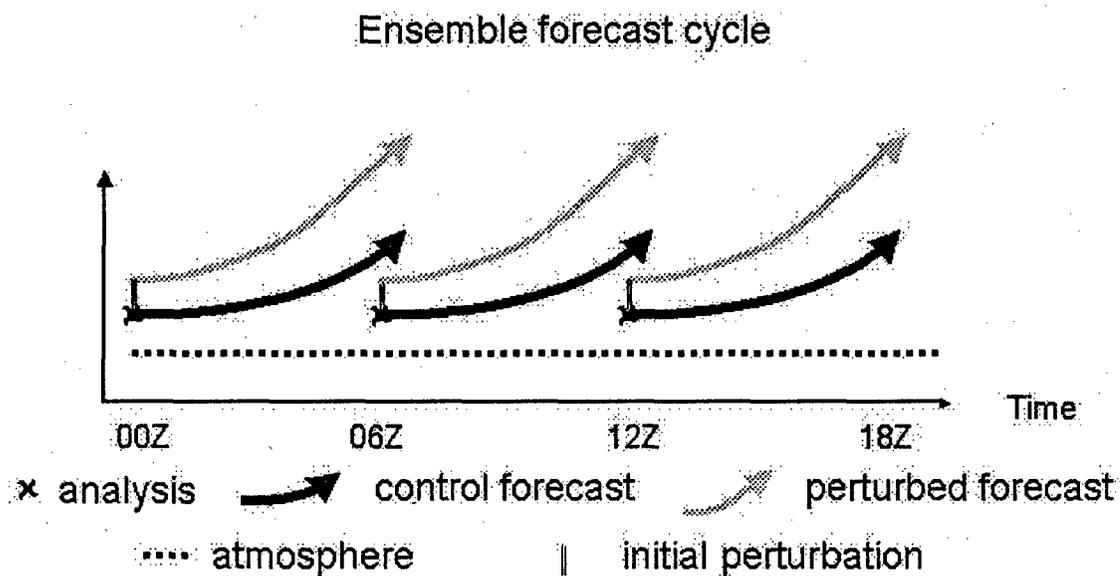


Figure B.2: Cartoon for a typical operational ensemble forecast cycle.

4. Eigenvector of a covariance matrix

A covariance matrix of a K -dimensional variable \mathbf{x} is written as $\langle \mathbf{x}\mathbf{x}^T \rangle$. This covariance matrix can be decomposed into the following format,

$$\langle \mathbf{x}\mathbf{x}^T \rangle = \mathbf{E}\mathbf{\Gamma}\mathbf{E}^T, \quad (\text{B4})$$

where columns of \mathbf{E} contain orthonormal vectors, called eigenvectors of $\langle \mathbf{x}\mathbf{x}^T \rangle$. The matrix $\mathbf{\Gamma}$ is a diagonal matrix whose i th diagonal element is the eigenvalue for the i th eigenvector (i.e., the i th column of \mathbf{E}). The diagonal elements of $\mathbf{\Gamma}$ are in general

ordered in decreasing value. The leading eigenvector, i.e., the eigenvector corresponding to the largest eigenvalue, explains the most variance of \mathbf{x} (Peixoto and Oort 1992). The eigenvectors in \mathbf{E} form an orthonormal basis of vectors for any K -dimensional vector. In other words, The K -dimensional variable \mathbf{x} can be expressed in the form of linear combination of the eigenvectors.

VITA

Xuguang Wang

Education

2004, Ph.D. in Meteorology, The Pennsylvania State University

1998, B.S. in Atmospheric Science, Beijing University, China

Experience

1999-2004, Graduate research assistant, Department of Meteorology, the Pennsylvania State University

Refereed Journal Publications

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.

Wang, X., C. H. Bishop and S.J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590-1605.