U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION

OFFICE NOTE 416

THE USE OF BRED VECTORS IN THE NCEP GLOBAL 3-D VARIATIONAL ANALYSIS SYSTEM

ZHAO-XIA PU
UCAR VISITING SCIENTIST PROGRAM

EUGENIA KALNAY, DAVID PARRISH
ENVIRONMENTAL MODELING CENTER

WAN-SHU WU, ZOLTAN TOTH
GENERAL SCIENCES CORPORATION

JUNE 1996

THIS IS AN UNREVIEWED MANUSCRIPT, PRIMARILY INTENDED FOR INFORMAL EXCHANGE OF INFORMATION AMONG NCEP STAFF MEMBERS

* To be submitted to *Weather and Forecasting*

ABSTRACT

The errors in the first guess (forecast field) of an analysis system vary from day to day, but, as is the case in all operational data assimilation systems, forecast error covariances are assumed to be constant in time in the NCEP 3-dimensional variational analysis system (SSI). This study focuses on the investigation of the impact of modifying the error statistics by including effects of the "errors of the day" on the analysis system. An estimate of forecast uncertainty, as defined from the bred growing vectors of the NCEP operational global ensemble forecast, is applied in the NCEP operational SSI analysis. The growing vectors are used to estimate the spatially and temporally varying degree of uncertainty in the first guess forecasts used in the analysis. The measure of uncertainty is defined by a ratio of the local amplitude of the growing vectors, relative to a background amplitude measure over a large area. This ratio is used in the SSI system for adjusting the observational error term. Preliminary experiments show positive impact of this virtually cost-free method on the quality of the analysis and medium range weather forecasts, encouraging us to test it in the operational practice. The results of a 45-day parallel run, and a discussion of other methods to take advantage of the knowledge of the day-to-day variation in forecast uncertainties provided by the NCEP ensemble forecast system, are also presented in the paper.

1. Introduction

A 3-dimensional variational data assimilation system (3-D VAR), known as the SSI (spectral statistical interpolation) analysis system was implemented into the operational global medium range weather forecast system at NCEP in 1991(Parrish and Derber 1992; Derber et al. 1991). This advanced data assimilation scheme has played a vital role in recent data assimilation and research at NCEP and is still under development (Derber and Wu 1996; Parrish et al. 1996; Derber et al. 1994). The analysis is performed every 6 hours and the 6-hour forecast field is used as a first guess in the system. As in all general variational data assimilation systems, the objective function to minimize is defined as:

$$2J=(x-x_b)^T B^{-1}(x-x_b)+(K(x)-y)^T O^{-1}(K(x)-y))+J_c \qquad (1)$$

where X is the analysis variable, X_b is the first guess and y is the observational vector. B and O denote the 6 hour forecast and observational error covariance matrices, respectively. K is a transformation operator which transforms analysis variables into simulated observations. J_c denotes a dynamical constraint penalty which enforces a global balance of the analysis increments. This system is used to find an analysis field which best fits both the first guess and observations. It assumes that the forecast error and observation error are not correlated. The error covariance matrices B and O are currently held constant in time (Parrish et al. 1992). However, the uncertainties in the first guess will change from day to day. It is necessary to modify such covariance weights (Wahba et al. 1995). It is our purpose to investigate the impact of modifying the error statistics by introducing the effect of "errors of the day". Because the large spread in the short-range ensemble forecasts usually can be used to point out areas of large uncertainty, an estimate of forecast uncertainty, as defined from the bred growing vectors of the NCEP operational global ensemble forecast (Toth and Kalnay 1993 & 1996), is introduced in the NCEP SSI analysis system. Since currently the background error covariance in the SSI is cast in spectral and not physical space, the impact of large forecast error is done by modifying the observational error covariance.

2. Methodology

The spatially and temporally varying degree of uncertainty in the first guess forecast is estimated by 24-hour operational global ensemble forecasts. It is defined as a ratio of the local amplitude of the growing vectors, relative to a reference amplitude measured over a large area:

$$\sigma^2 = \frac{\sum_{i=1}^{n} T_S(F_i - F_c)^2}{\sum_{i=1}^{n} T_L(F_i - F_c)^2} = \frac{Spread\ (small\ scale)}{Spread\ (large\ scale)} \tag{2}$$

where σ^2 is the ratio and it is defined in grid space. T denotes a smoothing convolution which is performed using spectral transforms (Purser et al. 1994), F represents the 24h forecast at the same verification time, subscript i and c denote the ensemble member and control forecast (operational forecast), respectively. L and S mean the length of large scale (reference) and small scale (local) used in the smoothing operator. The initial perturbation amplitudes in the NCEP ensemble forecast are spatially varying, smaller in data-rich continents, and larger over oceans, in order to be proportional to the average analysis error. The ratio between the small scale and the large scale ensemble spread is introduced in (2) in order to capture only the day-to-day variation in forecast uncertainty, rather than the time averaged distribution of analysis errors. When the ratio is large, there is more uncertainty in the first guess, and we assume that the guess field (6-hour forecast) is likely to have larger error in these areas. Since it is difficult to introduce local adjustments to the background error B because it is defined spectrally, instead, more weight is given to observational data in areas of large uncertainty by directly dividing the observation error variance by $1/\sigma^2$, whenever this factor is less than 1. Hence the analysis is driven closer to the data in these areas more than in the other areas. If the ratio is smaller than 1 (the forecast uncertainty is small), the observation error is maintained without change at its nominal value.

3. A Numerical Experiment

The method is tested in the current operational SSI analysis and medium range forecast system. The model used in experiments is the same as the NCEP operational global spectral model, but with lower horizontal resolution, with T62 with 28 sigma vertical levels (Pan et al. 1995). In the data assimilation cycle, an analysis is performed every 6 hours. The NCEP ensemble forecast system is also performed in a T62 L28 version, and it generates 5 pairs of ensemble forecast members (Toth and Kalnay 1996) at 0000UTC and 2 pairs of ensemble members at 1200UTC.

As described in (2), the value of the ratio is an RMS average over the member of ensemble pairs and it will depend on the number of the ensemble members which are used in the computational procedure. In our experiment, we only calculate the ratio at every 0000UTC by using all 5 pairs of 24-hour forecast ensemble members, then the ratio at 0600 UTC, 1200UTC and 1800UTC is computed by linear interpolation between two adjacent ratios at 0000UTC. The small and large scales used in (2) were chosen as in Experiment 1 (see below), with values that were assumed to be representative scales of local growth and smooth variation of the analysis errors.

a. Characteristics of the uncertainty ratio, and impact on the analysis field and first guess

The goal of this work is to point out areas where large uncertainties are present in the first guess, and to make appropriate adjustments to the error covariance only in such areas. In formula (2), if σ^2 greater than 1, it means that the local forecast spread is large. Based on this assumption, we set a lower bound of 1 on σ^2 (i.e., we set $1/\sigma^2 \leq 1$ before multiplying the observational errors). The distribution of the ratio varies not only in time, but also with the component of the field (temperature, wind, moisture), and vertical level. Fig.1 shows an example of distribution of $1/\sigma^2$ at different sigma levels for wind field at 0000 UTC 18 June 1996. It shows that there are distinct values of the forecast spread at different vertical levels. The impact of changes in moisture is larger in the tropics (figure not shown).

Then we introduce the effect of large uncertainties into the NCEP 3-dimensional SSI analysis system by multiplying the observational error covariance by $1/\sigma^2$ (with an upper bound of 1). The results are compared with a control run which keeps the error covariances constant in time. To verify the quality of the analysis, the RMS fit of both temperature (in K) and vector wind (in m/sec) against rawinsonde data are presented in Fig.2 for the analysis field itself and in Fig.3 for the next first guess which is the 6-hour forecast started from the analysis field. This particular case is verified at the analysis time of 0000UTC 18 June 1996. The figure shows that the method drives the analysis fields closer to the observations when compared to the control analysis field (as could be expected from the reduction of the observational errors in areas of large forecast uncertainty). However, this improved fit is preserved in the next first guess field, indicating that the analysis has been improved by this procedure. We have obtained similar results in most cases.

b. Impact of variable uncertainty on medium range weather forecasts

We first tested the method using the data period 0000UTC 1 August 1995 to 0000UTC

14 August 1995. According to the scale of smoothing operators, two experiments were performed: In Experiment 1 the large scale smoothing is taken as L_H=1500 km for horizontal smoothing and L_V=4km for the vertical. For small scale smoothing, the horizontal scale is S_H=300km and the vertical scale S_V=1 km. In Experiment 2, we used the same scales except for the large scale horizontal smoothing, which is taken as L_H=2000km. We use the obtained ratio to adjust the observation error covariance in the SSI system, then 5 day forecasts from every 0000UTC analysis field were compared with the corresponding control (operational T62) forecast. Table 1 shows the comparison of the 1-5-day forecast average anomaly correlation scores for 500mb geopotential height. It shows a positive impact of the experiments with respect to the control (which did not account for the forecast time verying uncertainty) in the short and medium range weather forecast. Experiment 2 is slightly better than Experiment 1 in the Southern Hemisphere, and it was used for the rest of the experiments presented in the next section.

Table 1. Comparison of 1-5 day forecast anomaly correlation average score verified against control analysis for 500mb geopotential heights (n=14 cases, 1-20 waves)

	N. Hem.			S. Hem.		
Day	Ctrl.	Exp.1	Exp.2	Ctrl.	Exp.1	Exp.2
1	.978	.978	.978	.977	.977	.978
2	.937	.938	.938	.929	.929	.930
3	.872	.873	.873	.862	.863	.865
4	.800	.803	.803	.787	.790	.794
5	.709	.713	.710	.712	.715	.722

4. The results of parallel tests

The method has been tested in parallel within the current NCEP global medium range weather forecast system starting from 0000UTC 23 April 1996, comparing as a control with the lower resolution version (T62/L28) of the NCEP global operational forecast model. The parallel test is designed following Exp. 2 in section 3b. As of 0000UTC 25 June, there were 45 cases available for comparison. Fig. 4 shows the 5-day forecast anomaly correlation score for 1000mb and 500mb geopotential height verified against the control analysis. The results indicate that the method improves the medium range weather forecast for most cases. Table 2 shows the comparison of the 1-5-day forecast average anomaly correlation scores for geopotential height, demonstrating that the method gives significant benefit to the medium range forecast skill in both the Northern and the Southern Hemispheres, the improvement being larger in the Southern Hemisphere.

Table 2. Comparison of 1-5-day forecast anomaly correlation average score verified against the control analysis for 500mb and 1000mb geopotential heights (n= 45 cases, 1-20 waves)

	Northern. Hemisphere				Southern Hemisphere			
	1000mb		500mb		1000mb		500mb	
Day	Ctrl.	Test	Ctrl.	Test	Ctrl.	Test	Ctrl.	Test
1	.964	.962	.983	.983	.961	.961	.976	.977
2	.911	.910	.950	.950	.899	.900	.976	.977
3	.853	.854	.898	.899	.809	.816	.850	.861
4	.764	.769	.822	.824	.709	.727	.751	.765
5	.636	.644	.711	.717	.604	.624	.639	.652

5. Summary and discussion

This study shows that the use of the bred vectors of the ensemble forecast in the NCEP SSI analysis system has improved the quality of short and medium range weather forecasts, and therefore that the analysis is also better. The method drives the analysis field closer to the observational data in the areas where the ensemble identifies large forecast uncertainty. It also improves the next guess, as well as the medium range weather forecast. The method only requires calculating the ratio of the ensemble spread in small and large scales, and the inverse of this ratio (bounded by 1) is used to adjust the observational errors. The computational cost of this method is negligible, since the ensemble forecasts are already available. The positive impact of the experiments encourages further exploration of the use of the bred vectors in improving the analysis system by taking into account the forecast "errors of the day" rather than assuming that the forecast error covariance is constant in time, as currently done in all operational systems.

In this study we took a very simple approach, by reducing the observational errors in areas identified by the ensemble as areas of large forecast uncertainty. The results suggest that the NCEP bred vectors provide a good representation of forecast errors even at the shortest ranges. Other (more advanced) methods to take advantage of this knowledge of the day-to-day variability in the forecast errors are also possible and will be explored in the future. Two such methods are an improvement of the first guess by minimizing the distance between the first guess and the observations, but moving only along the direction of the bred growing vectors (Kalnay and Toth 1994; Purser et al. 1994), and the inclusion of an error covariance based on the bred vector perturbations into the forecast error covariance.

Acknowledgments

We are most grateful to Drs. Mark Iredell, Peter Caplan, Hua-Lu Pan and to Mr. Yuejian Zhu for their help in creating the executables for the parallel tests. The first author is supported by UCAR/NCEP Visiting Scientist Program.

REFERENCES

Derber J. and W.-S. Wu, 1996: The use of cloud-cleared radiances in the NCEP's SSI analysis system. Preprints, *11th Conf. On Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc..

Derber, J., D. Parrish, W.-S. Wu, Z. Pu and S. Rizvi, 1994: Improvements to the operational SSI global analysis system. Preprints, *10th Conf. On Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 149-150.

Derber, J., D. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Weather & Forecasting*, 6, 538-547.

Kalnay, E., and Toth, Z., 1994: Removing growing errors in the analysis. *10th conference on numerical weather prediction*, July 18-22, 1994, Portland, AMS, 212-215.

Pan, H.-L., J. Derber, D. Parrish, W. Gemmill, S.-Y. Hong, and P. Caplan, 1995: Changes to the 1995 NCEP operational MRF model analysis/forecast system. *Technical Procedures Bulletin*, No.428. National Weather Service. 30pp.

Parrish, D. F., J. Derber, J. Purser, W. Wu, and Z. Pu, 1996: The NMC Global Analysis system: Recent improvements and future plans. *J. Meteor. Soc. Japan*, in press.

Parrish, D.F., J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon.Wea. Rev.*, 120, 1747-1763.

Purser, R. J., D. F. Parrish, Z. Toth, and E. Kalnay, 1994: Numerical filtering applied to the enhancement of pre-analysis corrections of background errors expressible as rapidly amplifying modes. Preprints, *10th Conf. On Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc..

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of initial perturbation. *Bulletin of the Amer. Meteor. Soc.*, 74, 2317-2330.

Toth, Z., and E. Kalnay, 1996: Ensemble forecasting at NCEP and the breeding method. Submitted to *Mon. Wea. Rev.*

Wahba,G., D R Johnson, F Gao, J Gong, 1995: Adaptive tuning of numerical weather prediction model: Randomized GCV in three-and four-dimensional data assimilation, *Mon.Wea.Rev.*, 123. 3358-3369.

Figure Captions:

Fig.1 an example of distribution of the ratio at different sigma levels. This contour illustrated the inverse of the ratio ($1/\sigma^2$) at sigma level 7 (sigma=0.846), level 13 (sigma=0.501) and level 18 (sigma=0.210) for wind field at 0000 UTC 18 June 1996.

Fig. 2 Root-Mean-Square fit of temperature and vector wind against the rawinsonde data. for the analysis field at 0000 UTC 18 June 1996. The vertical axis denotes the pressure level(unit: hpa) and the horizontal axis denotes the RMS error of temperature (unit: K) or vector wind(unit: m/s). The dashed line for experiment and solid line for control analysis.

Fig. 3 Same as Fig.2. except the RMS fit for the 6-hour forecast.

Fig. 4 Scatter diagrams of the 5-day forecast anomaly correlation (AC) scores for geopotential height field in experiment and control forecast.
a). Northern Hemisphere 500 mb. b). Southern Hemisphere 500mb
c). Northern Hemisphere 1000mb. d). Southern Hemisphere 1000mb

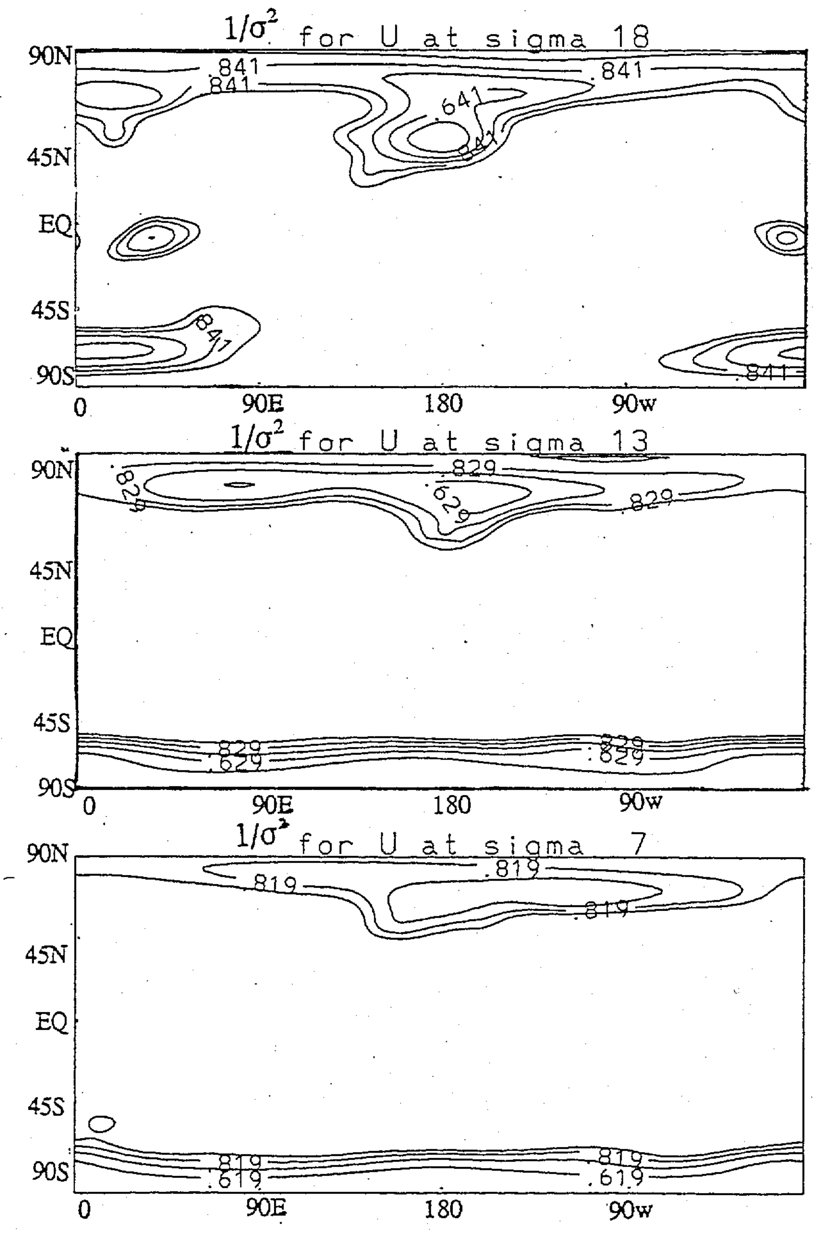

$1/\sigma^2$ for U at sigma 18
90N
45N
EQ
45S
90S
0
90E
180
90w
.841
.641
$1/\sigma^2$ for U at sigma 13
.829
.629
$1/\sigma^2$ for U at sigma 7
.819
.619

Fig. 1

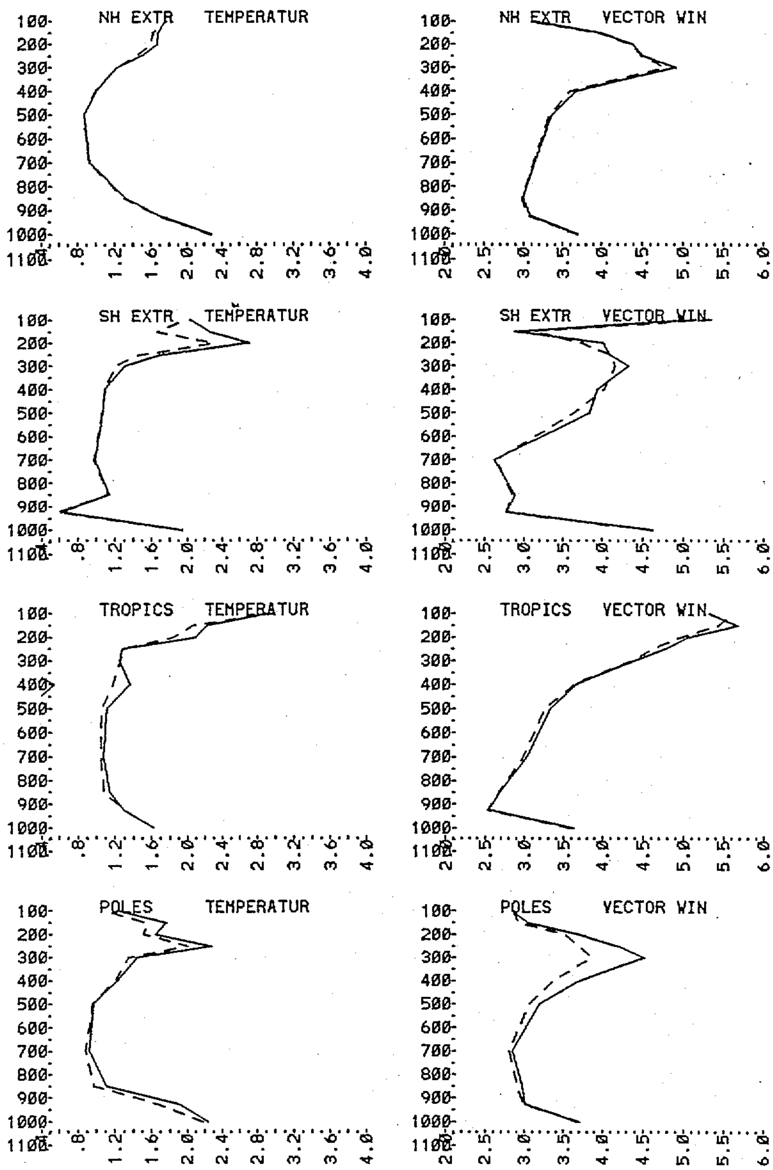

NH EXTR TEMPERATUR
NH EXTR VECTOR WIN
SH EXTR TEMPERATUR
SH EXTR VECTOR WIN
TROPICS TEMPERATUR
TROPICS VECTOR WIN
POLES TEMPERATUR
POLES VECTOR WIN
100- 200- 300- 400- 500- 600- 700- 800- 900- 1000- 1100-
.4 .8 1.2 1.6 2.0 2.4 2.8 3.2 3.6 4.0
2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0

Fig. 2

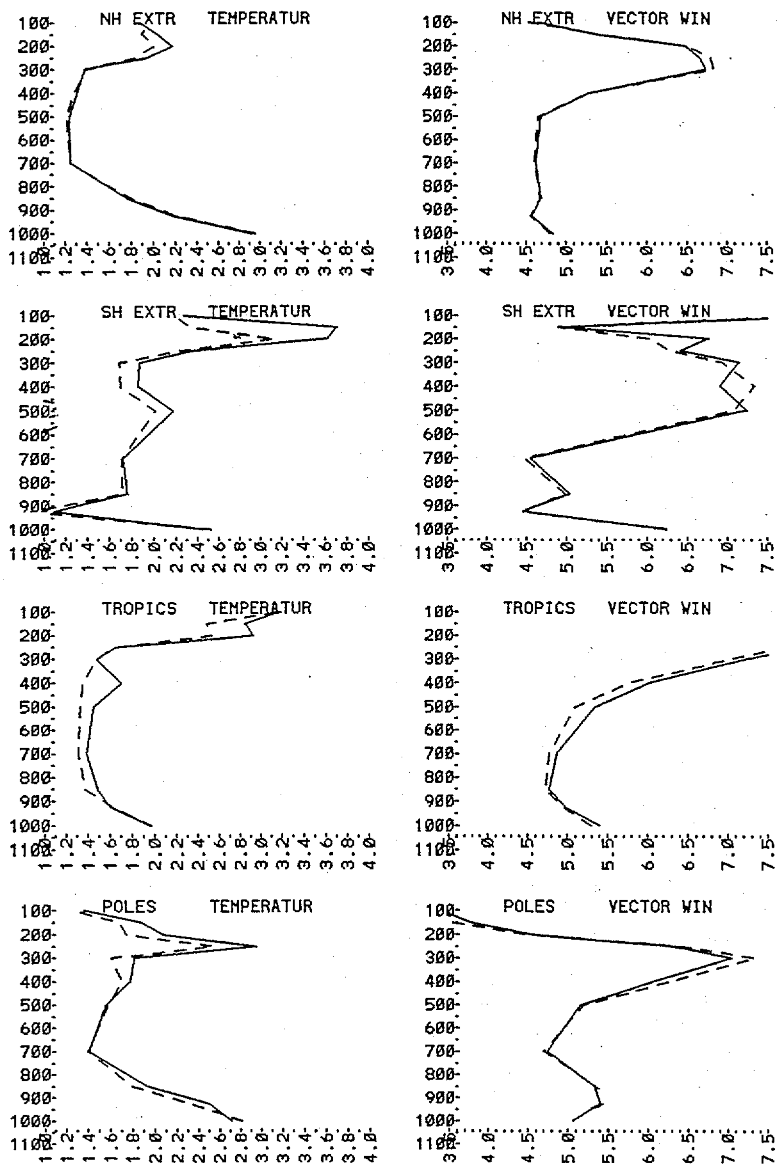

NH EXTR TEMPERATUR
NH EXTR VECTOR WIN
SH EXTR TEMPERATUR
SH EXTR VECTOR WIN
TROPICS TEMPERATUR
TROPICS VECTOR WIN
POLES TEMPERATUR
POLES VECTOR WIN

Fig. 3

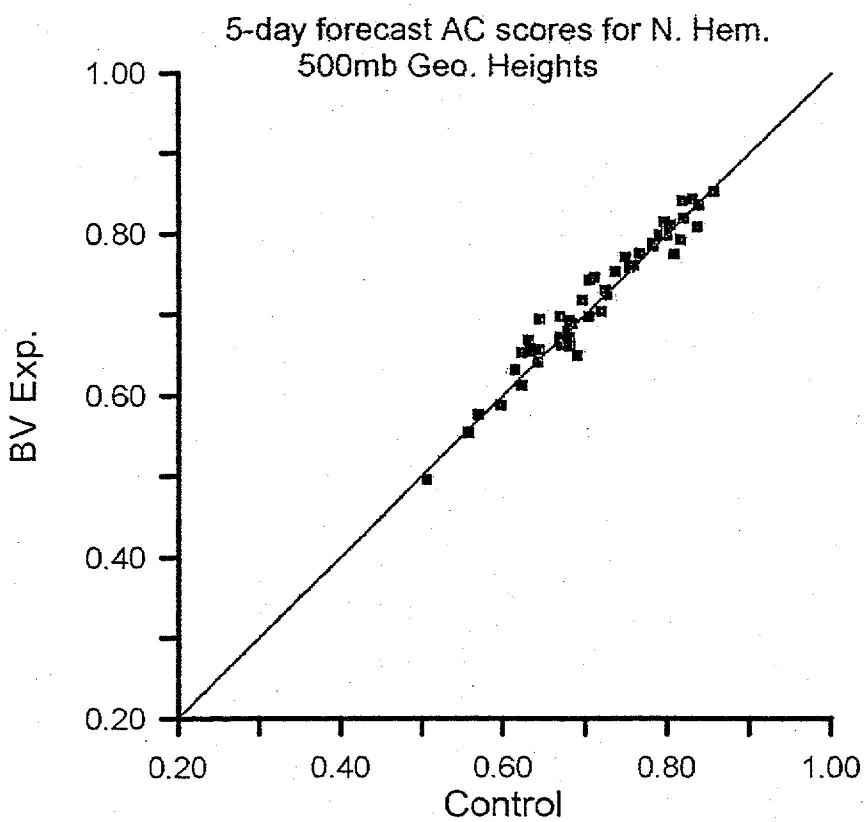
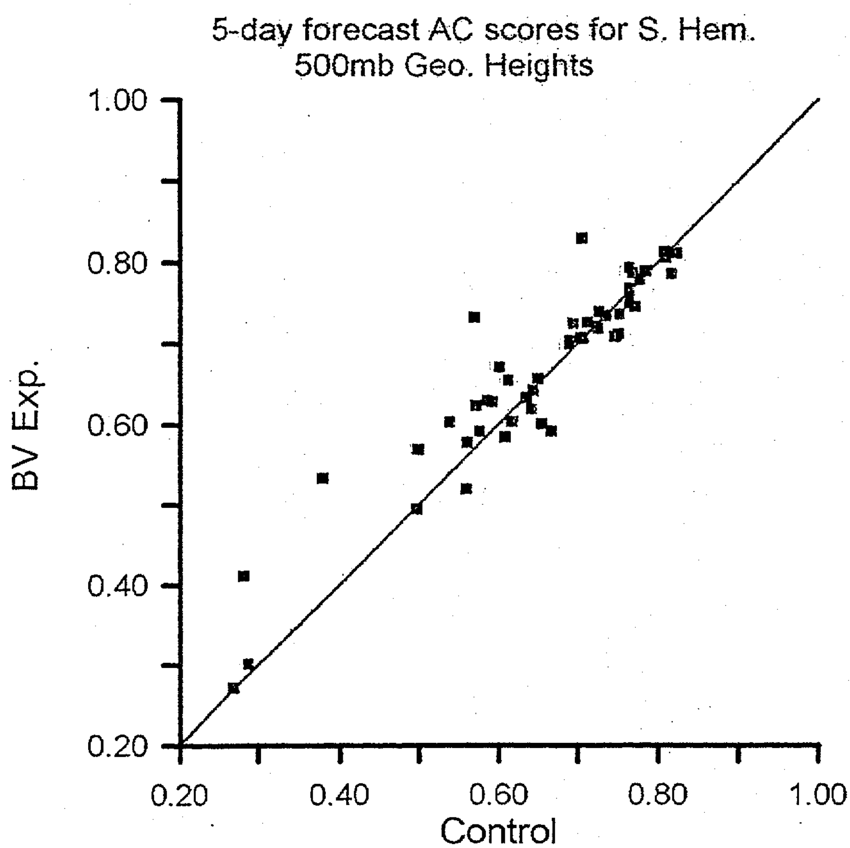
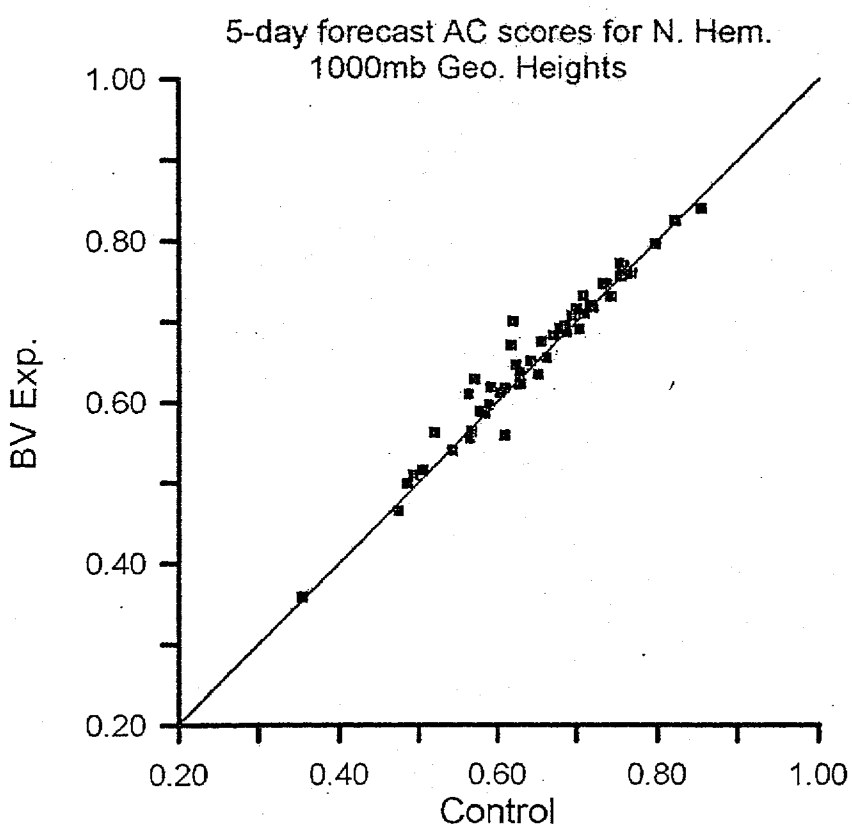

5-day forecast AC scores for N. Hem.
500mb Geo. Heights
1.00
0.80
0.60
0.40
0.20
BV Exp.
0.20
0.40
0.60
0.80
1.00
Control

Fig. 4a



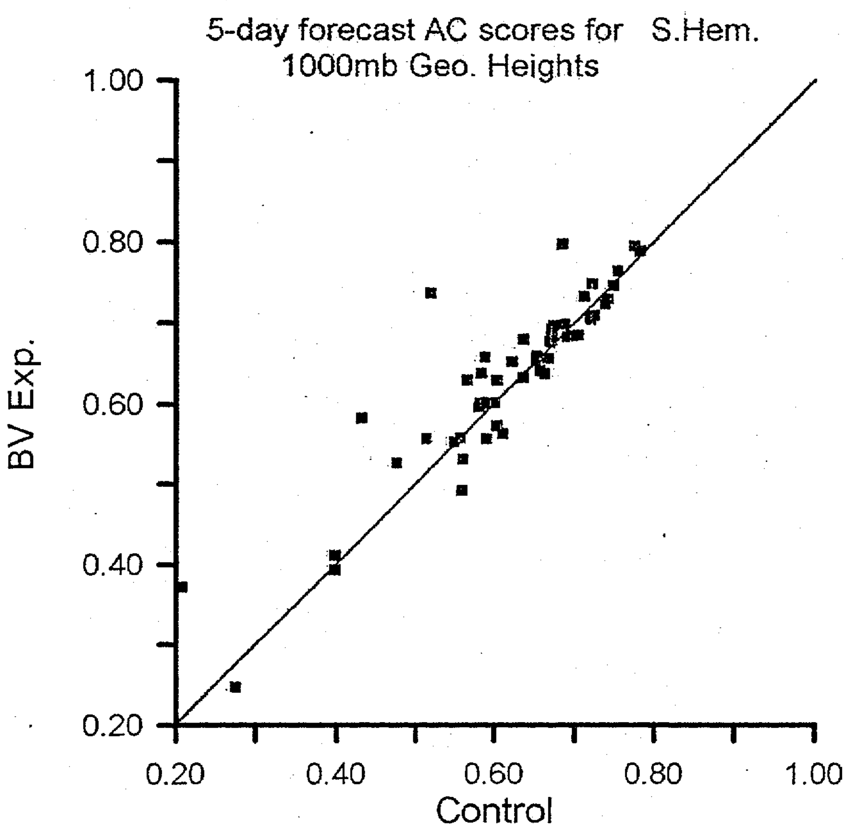
5-day forecast AC scores for S. Hem.
500mb Geo. Heights
1.00
0.80
0.60
0.40
0.20
BV Exp.
0.20
0.40
0.60
0.80
1.00
Control

Fig. 4b



5-day forecast AC scores for N. Hem.
1000mb Geo. Heights
1.00
0.80
0.60
0.40
0.20
BV Exp.
0.20
0.40
0.60
0.80
1.00
Control

Fig. 4c



5-day forecast AC scores for S.Hem.
1000mb Geo. Heights
1.00
0.80
0.60
0.40
0.20
BV Exp.
0.20
0.40
0.60
0.80
1.00
Control

Fig. 4d