Global Ensemble Forecast System (GEFS) and Northern American Ensemble Forecast System (NAEFS)

Version 1.0 User's Guide

(Draft)

Ensemble and Post-process Team
Environmental Modeling Center
National Centers for Environmental Prediction, NOAA

November 12 2009

Foreword

This User's Guide describes the Global Ensemble Forecast System (GEFS) / Northern American Ensemble Forecast System (NAEFS) Version 1.0, released in January 2010. As the GEFS/NAEFS is developing further, this document will be continuously enhanced and updated to match the released version.

For the latest version of this document, please visit the GEFS/NAEFS User's Website at http://wwwt.emc.ncep.noaa.gov/gmb/ens/index.html .

Please send feedback to Yuejian.Zhu@noaa.gov .

Contributors to this guide:
Yuejian Zhu, Richard Wobus, Mozheng Wei, Dingchen Hou, Bo Cui and Qingfu Liu

Table of Contents

Chapter 1: Overview

1.1 Historic Review:

NCEP's Global Ensemble Forecast System (GEFS) has been in operation since 1993, using the NCEP Global Forecast System (GFS) model for integration and breeding technique to generate perturbations in the initial conditions. After the Aug. 25, 2005 implementation, GEFS runs four times per day (0000, 0600, 1200 and 1800 GMT). At each time, 10 (5 pairs) perturbed members are initialized using breeding method with a breeding cycle of 6 hours. A control forecast, initialized with unperturbed initial condition, is also run at 00Z. The control and all perturbed forecasts are integrated at T126 resolution and 28 vertical levels (L28) up to 384 hours. A "relocation" technique is applied in the initial condition of each run to adjust the initial central location of tropical storms to the actual location.

1.2 Recent Changes for GEFS (May 2006, March 2007, Dec. 2009)

(Refer to: http://www.emc.ncep.noaa.gov/gmb/ens/ens_imp_news.html)

1.1.1 Horizontal Resolution

In the coming implementation to be finished in late 2009, the horizontal resolution of GEFS runs will be increased to T190.

1.1.2 Membership

The number of perturbed members was increased to 14 in May 2006 and 20 in March 2007. Since the May 2006 implementation, ensemble control forecast has been included for all four forecast cycles.

1.1.3 Generation of the Initial Perturbations

Breeding Method (BM) is modified by applying Ensemble Transformation (ET) to the ensemble perturbations in short-range forecasts. The resulted initial perturbations are then rescaled, leading to ET with Rescaling (ETR) method. ETR was introduced to the breeding method in May 2006.

1.1.4 Representation of Model Related Uncertainty

In the coming implementation in late 2009, a Stochastic Total Tendency Perturbation Scheme (STTP) will be included to represent uncertainties associated with the NWP model used for the integration. STTP is based on the hypothesis that tendencies of the ensemble perturbations provide a representative sample of the random total model errors.

1.3 NAEFS (North American Ensemble Forecast System)

1.3.1 General Description.

The Canadian (Meteorological Service of Canada, MSC), the Mexican (National Meteorological Service of Mexico, NMSM), and the US (National Weather Service, NWS) NMS established the North American Ensemble Forecast System (NAEFS). The NAEFS was inaugurated in November 2004, and the first operational implementation of NAEFS products occurred in May 2006. In December 2007, down-scaling products for Continental United States (CONUS) have been implemented in NWS/US operation. Within the NAEFS, ensemble producing centers (currently MSC and NWS) (1) exchange their raw forecast data (operational since September 2004); (2) statistically post-process (include down-scaling) all ensemble members; and (3) jointly with other members (currently NMSM) develop and produce end products based on the combined ensemble of forecasts.

1.3.2 Basic products.

Statistical post-processing involves (a) the correction of all ensemble members for biases (first and higher moments), (b) the establishment of weights for the combination of all members, and (c) the expression of each bias-corrected forecast member in terms of percentile values within a long-term climatological distribution of the NCEP-NCAR reanalysis. The participating centers collaborate in the development of post-processing algorithms and software and share a common procedure to generate the basic products of bias-corrected forecasts, the corresponding weights and climatological percentile values. The products for probabilistic forecast (10%, 90%, 50%, mean, mode and spread) have been generated after statistical bias correction for all ensemble members. The free ftp distributions of these basic products were operationally implemented in May 2006 and December 2007.

1.3.3 End products.

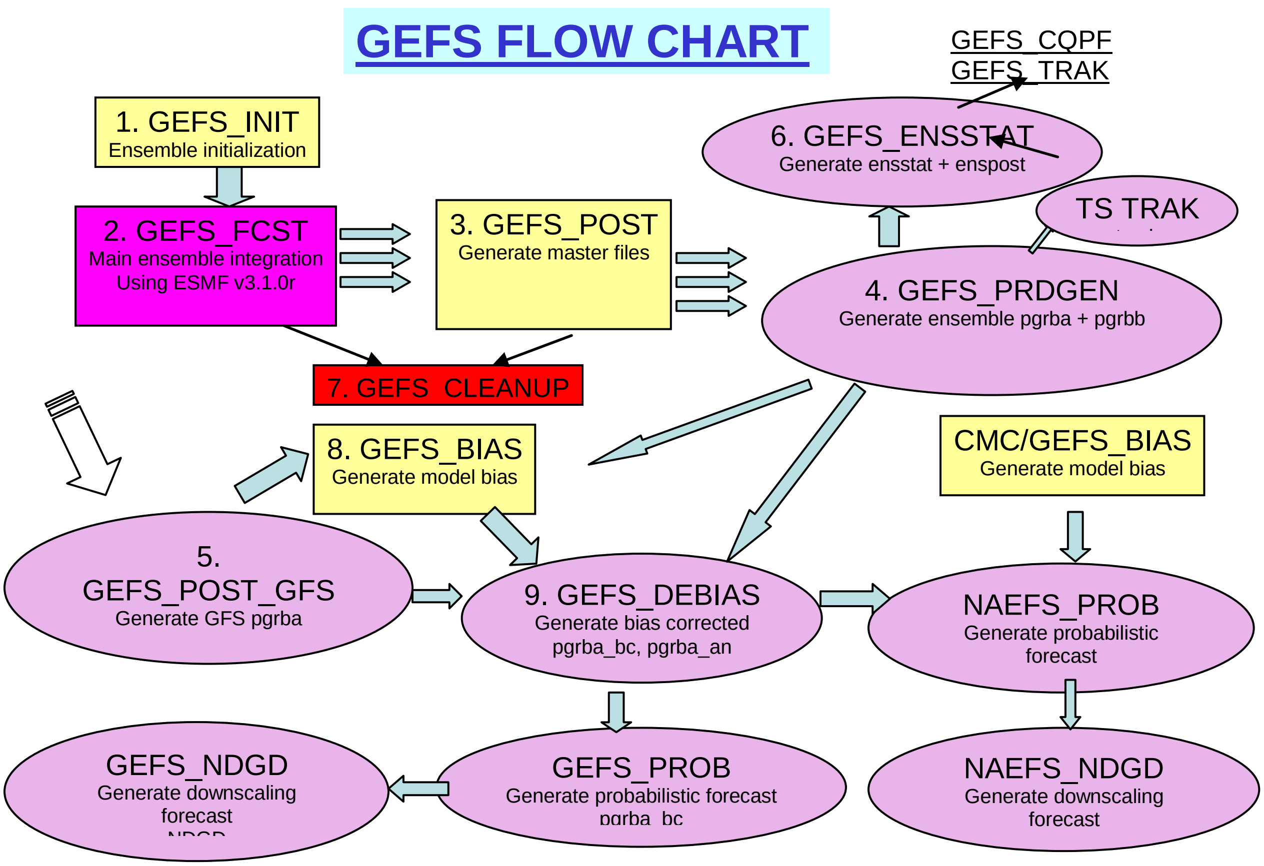
The final goal of the NAEFS is the generation of end products for the use of the participating and other NMS, including those used for severe weather warnings. Down-scaling probabilistic products for CONUS are generated in NDGD grid by using Real Time Meso-scale Analysis (RTMA) as proxy truth. Some of the end products are developed jointly (such as the North American week-2 temperature and precipitation anomaly forecast), while others will be provided by individual participating centers. In all cases, end-products will be based on the common set of basic products described above, ensuring the consistency of all NAEFS end products. NAEFS participants actively seek input from potential users from developing regions (such as the Caribbean, South America and Africa) regarding desired end products for these areas.

1.3.4 Expansion of NAEFS.

The current NAEFS can be considered as a prototype for a multi-center ensemble forecast system, envisaged by the THORPEX research program. The US Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC) will be next one to plan to joint NAEFS, while the US Air Force Weather Agency (AFWA) as a user. The Japan Meteorological Agency (JMA) expressed an interest in joining the NAEFS as producing centers. The UK Met Office also considers its participation, pending the results of a multi-year testing and evaluation phase. These possible expansions will broaden the scope of the NAEFS and may lead to the development of a Global Ensemble Forecast

System (GEFS), as the ensemble forecast component of the Global Interactive Forecast System (GIFS), foreseen by the THORPEX program. The NAEFS, and a possible future GEFS will well represent the spirit of the enhanced international collaboration sought by the THORPEX research program. In particular, the NAEFS/GEFS can provide a framework of operational requirements and constraints within which new research initiatives must be conceived on one had, and will offer a receiving end for any new methods developed based on the THORPEX Interactive Grand Global Ensemble (TIGGE) data archive, or related to other THORPEX initiatives.

Chapter 2: Software Installation



Chapter 3: Running GEFS/NAEFS

3.1 Full Cycling:

3.2 Forecast Only (no Cycling)

3.3 Downstream Dependences (For Implementation Use only)

3.3.1 Sigma (hybrid) files

SREF – use sigma/hybrid files as initial/boundary conditions

3.3.2 Pressure grib files

Wave ensembles, tracking, precipitation verification, MDL-GMOS

Chapter 4: GEFS/NAEFS Products

GEFS products are stored at CCS computer which is in the main directory: **/com/gens/prod**

For each cycle [hh:00,06,12,18] of year [yyyy]. Month [mm], day [dd], there is a subdirectory: **gefs.[yyyy][mm][dd]/[hh]/.**

4.1 Basic products:

Subdirectories:

pgrba:

pgrbb:

ensstat:

pgrba_bc:

pgrba_an:

pgrba-wt:

4.2 Derived products:

Subdirectories:

ndgd:

4.3 Public Access (ftp access):

NCEP anonymous ftp: ftp://ftpprd.ncep.noaa.gov/pub/data/nccf/com/gens/prod/

NWS Gateway anonymous ftp: ftp://tgftp.nws.noaa.gov/SL.us008001/ST.opnl/

NCEP NOMADS (Server 5): http://nomad5.ncep.noaa.gov/ncep_data/

4.4 Web Display:

NCEP operational images: http://www.nco.ncep.noaa.gov/pmb/nwprod/analysis/

CPC experimental images:

http://www.cpc.ncep.noaa.gov/products/predictions/short_range/NAEFS/Outlook_D264.00.php

EMC experimental images: http://wwwt.emc.ncep.noaa.gov/gmb/ens/NAEFS/NAEFS-prods-NCEP.html

MSC experimental images: http://www.meteo.gc.ca/ensemble/naefs/index_e.html

Chapter 5: Ensemble Based Probabilistic Forecast Verification

(Yuejian.Zhu@noaa.gov)

5.1 Introduction:

The NCEP ensemble verification system was developed to evaluate ensemble based probabilistic forecast in the 90s (Zhu et al., 1996). This system mainly focuses on two attributes: the ***reliability*** and ***resolution*** (Toth et al., 2003, 2006) of the NCEP ensemble based probabilistic forecast, in addition to the traditional verification measures such as Pattern Anomaly Correlation (PAC) and Root Mean Square (RMS) error for the ensemble mean, rank histogram, and outliers (Zhu, 2004; Toth et al., 2003), and Perturbation versus Error Correlation Analysis (PECA) (Wei and Toth, 2003), etc. For precipitation verification, Equitable Threat Score (ETS), True Skill Statistics (TSS) and Bias (BI) have been used to measure the ensemble mean (Zhu, 2007). In this ensemble based probabilistic verification system, the definitions of events are based on 1) user defined thresholds, 2) climatological percentiles, and 3) the ensemble members. In practice at NCEP, the climatological percentiles (10 climatologically-equally-likely bins) have been used for NCEP/GEFS (Global Ensemble Forecast System) daily verification. Therefore, the probabilistic skill scores for current NCEP/GEFS forecasts are based on the NCEP/NCAR 40-year reanalysis climatology (references). On a routine basis, this system generates a Brier Score (BS), Brier Skill Score (BSS) with its decomposition of reliability and resolution, Ranked Probability Skill Score (RPSS), Continuous Ranked Probability Skill Score (CRPSS), Relative Operational Characteristics (ROC) area score, Relative Economic Value (REV) score for selected loss/cost ratios to apply to upper atmospheric variables such as 500hPa geopotential height, and 850hPa temperature and near surface variables such as 1000hPa geopotential height, 2-meter temperature, and 10-meter wind (u and v). In terms of the ensemble mean, as in a deterministic forecast, ensemble spread and RMS error have been introduced, histogram (or Talagrand) distributions and outliers have been generated to measure the ensemble's reliability and consistency. This system was recently upgraded and applied to the Northern American Ensemble Forecast System (NAEFS), which combines the NCEP and CMC ensemble forecasts. This article mainly summarizes this verification system.

5.2 Methodology of Verification:

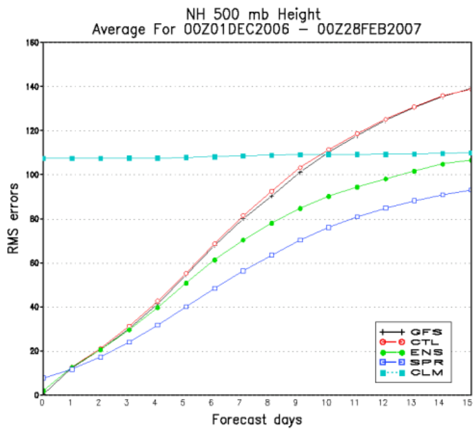
5.2.1 RMS error and SPRD (ensemble spread)

RMS errors of the ensemble mean measure the distance between forecasts and analyses (or observations). SPRD (ensemble spread) is calculated by measuring the deviation of ensemble forecasts from their mean (Zhu, 2005). Figure 1 is an example of a display of RMS errors and ensemble spread (SPRD) for a 15-day lead-time forecast. Usually, SPRD is defined as:

$$SPRD = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(\overline{f} - f(n))^2}$$

Where $\overline{f} = \frac{1}{N}\sum_{n=1}^{N} f(n)$ is for the ensemble mean and f is for the ensemble forecast.

In general, an ideal ensemble forecast will be expected to have the same size of ensemble spread as their RMS error at the same lead time in order to represent full forecast uncertainty (Zhu, 2005, Buizza et al., 2005). But most of the ensemble systems are under-dispersed (less spread) for longer lead times due to an imperfect model system (or physical parameterizations) and other things. Therefore, a stochastic process will be introduced to increase ensemble spread for longer lead-time forecasts (Hou et al., 2008). On the other hand, the ensemble mean consistently performs better than the high resolution deterministic forecast GFS (T382L64) after a 2-day lead time, while the high resolution GFS uses similar (or more) resources than the global ensembles (20 members at T126L28 resolution).



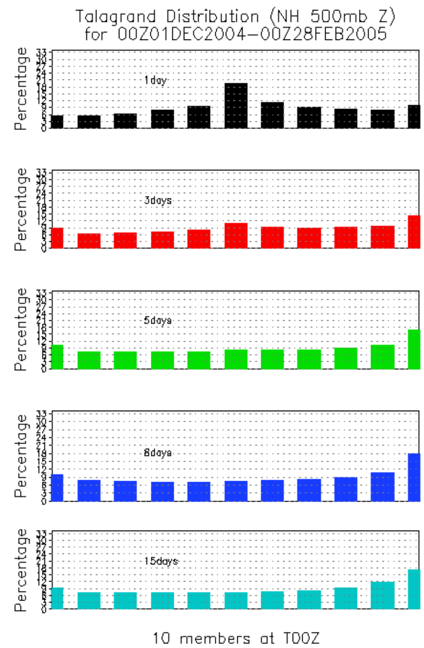


Fig. 1. RMS error for ensemble mean (blue) and ensemble spread (green) for NH ex-tropical 500hPa geopotential height of the 2006-2007 winter season, compared to the GFS (black) and ensemble control (CTL, red) RMS errors. The top curve (cyan) is for RMS error of climatology.

Figure 2 shows an example of a NCEP GEFS (10-member) forecast for the period December 1st 2004 – February 28th 2005, for 1-, 3-, 5-, 8-, and 10-day NH 500hPa geopotential height. The histogram distribution for the raw forecast is over-dispersed at the short lead-time at that time period. There is a little cold bias for longer lead-times as you see the high bars move right as lead time increases.

5.2.2 Histogram Distribution:

A Histogram (or Talagrand) distribution is a simple measurement used to verify an ensemble system and its forecast distribution. The calculation formula of the Histogram Distribution (*HD*) for one grid point, at time *t*, analysis or observation (*a*), *N* ensemble forecasts *f(1,2,…N)* after re-ordering from low to high according to their values could be written as:

$$HD(n) = \begin{cases} \frac{1}{N}, n = 1, a \le f(n) \\ \frac{1}{N}, n = 2,..., N, f(n-1) < a \le f(n) \\ \frac{1}{N}, n = N+1, a > f(N) \end{cases}$$

There are a few resulting common shapes such as a U-shape, L-shape and A-shape. The U-shape represents an over-dispersed ensemble (more spread), A-shape means the ensemble is under-dispersed (less spread), and the L-shape represents a typically biased forecast. The best ensemble system will be expected to have a constant (or flat line) HD.

Fig. 2. NCEP global ensemble (10 member) histogram (Talagrand) distribution for NH ex-tropical 500hPa geopotential height for 1-, 3-, 5-, 8-, 15-day forecasts of the 2004-2005 winter.

5.2.3 CRPS and RPS

Continuous Ranked Probability Skill Score (CRPSS) and Ranked Probability Skill Score (RPSS) measure the reliability and resolution. The formulas can be written as follows:

$$CRPS = \int_{-\infty}^{+\infty} \left[F(x) - H(x - x_0) \right]^2 dx$$

Where the Heaviside Function H is $H(x - x_0) = \begin{cases} 0, x \le x_0 \\ 1, x > x_0 \end{cases}$ and

$$CRPSS = \frac{CRPS_r - CRPS_f}{CRPS_r}$$

Where *r* is for a reference and *f* is for a forecast.

$$RPS = 1 - \frac{1}{k-1} \left[\sum_{i=1}^{k} (\sum_{n=1}^{i} p_n - \sum_{n=1}^{i} o_n)^2 \right]$$

And $RPSS = \frac{RPS_f - RPS_r}{1 - RPS_r}$

Where *P* is a forecast probability, and *O* is for an observation or analysis.

For statistics over a long period, CRPS is very similar to RPSS. Therefore, we consider it possible to use either one of these two measures, whichever is more convenient. There is a very good example of this in Figures 5 and 6 for NH extra-tropical 850hPa temperature.

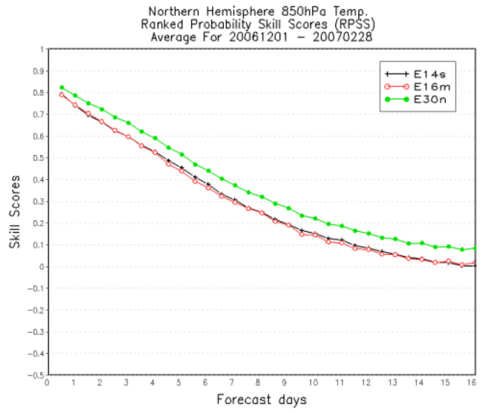


Fig. 3. CRPSS for the NCEP 14 global ensemble raw forecast (black) compared to the CMC 16 global raw forecast (red) and the combined NCEP and CMC ensembles (green) for NH extra-tropical 500hPa geopotential height for the winter 2006-2007.

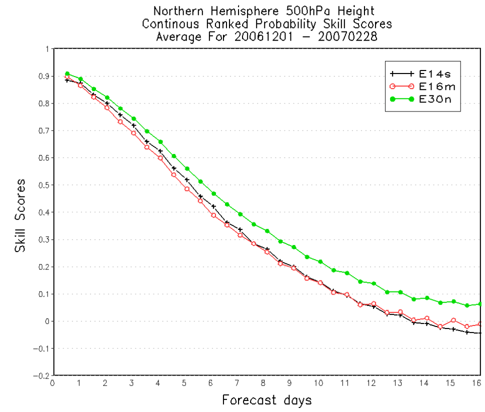


Fig. 4. RPSS for the NCEP 14 global ensemble raw forecast (black) compared to the CMC 16 global raw forecast (red) and the combined NCEP and CMC ensemble (green) for NH extra-tropical 850hPa temperature for the winter of 2006-2007.

5.2.4 Brier Score and Decomposition

There are many classical references that discuss the Brier score (BS), and its decomposition for reliability and resolution. (Wilks, 1995; Toth et al., 2003; 2006). In general, BS can be expressed as the summation of reliability, resolution and uncertainty (Wilks, 1995). CRP or RPS can be considered as a total integration of all probabilities. Users can review all the references to understand BS, reliability, resolution and uncertainty. And here is the final formula for decomposition:

BS = Reliability – Resolution + Uncertainty

5.2.5 Hitting Rate, False Alarm Rate and Economic Value

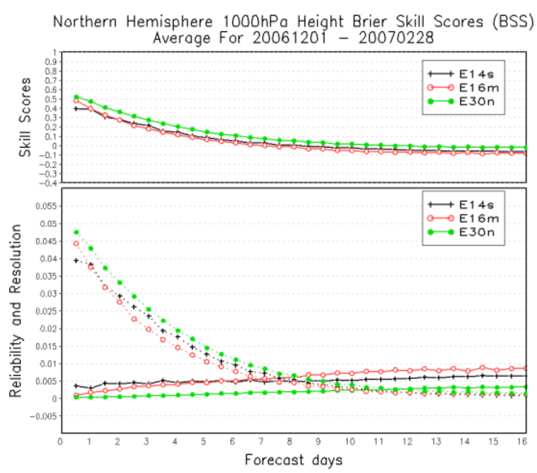
There is a traditional consideration for the hitting rate and false alarm rate. The typical application for this is the Relative Operational Characteristics (ROC) curve (Toth et al., 2003), or sometimes called the ROC area. Another application is the Relative Economic Value (REV), used when evaluating the loss and cost (Zhu and etc. 2002) which is very useful for decision makers.

5.3 Verification Statistics and Applications:

The NCEP/GEFS and NAEFS unified verification system will focus on probabilistic forecast verification for mainly short- and medium-range ensemble forecasts. Currently, it is available for the global ensemble forecast only, but it will be soon applied to the short-range ensemble forecast system as well. The discussion in Section 2 (Methodology of ensemble verification) describes the main characteristics of a probabilistic forecast which are more completely measured in terms of reliability and resolution. The NCEP/GEFS and NAEFS product verification statistics have been generated for the seasonal average and the skill scores have been posted at: http://wwwt.emc.ncep.noaa.gov/gmb/yzhu/html/opr/naefs.html since June 2006.

The Figures 4-9 are all skill scores for the 2006-2007 winter season which compare NCEP/GEFS 14-member raw forecasts (black, E14s), CMC/GEFS 16-member raw forecasts (red, E16m) and the combined NAEFS 30-member (NCEP(14) + CMC(16)) raw forecasts (green, E30n). Figure 4 shows CRPSS for NH extra-tropical 500hPa heights. NAEFS (green) raw forecasts are significantly improved in skill for all lead times, especially for longer lead times. In Section 2, CRPSS and RPSS (Figures 5 and 6) were discussed for NH extra-tropical 850hPa temperature. They are very similar to each other which suggest that either score could be used to verify the ranked probability. .

BSS and its decomposition (reliability and resolution) are shown in Figure 7 for NH extra-tropical 1000hPa height. The results are very similar to 500hPa height and 850hPa temperature. According to the formula in Section 2.d, BSS is equal to zero when resolution (going down with time from high to low) equals reliability (which goes up with time). The Reliability diagram (Figure 8) is more popular with many users. The diagonal line is the perfect line for a reliable forecast and the further you get from this line the worse the forecast. Apparently a bias corrected forecast has more reliability (Figure 8, red line).



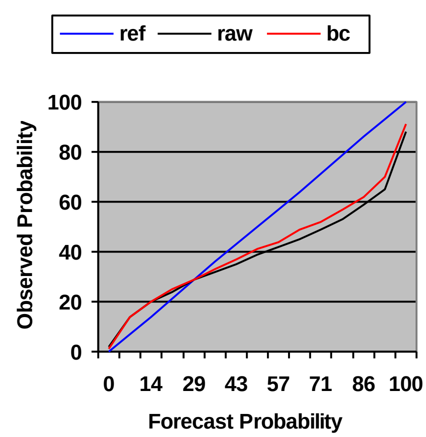


Fig. 5. BSS (top plot), Reliability (bottom plot, solid) and Resolution (bottom, dotted) for the NCEP 14 global ensemble raw forecast (black), compared to the CMC 16 global raw forecast (red) and the combined NCEP and CMC ensemble (green) for NH extra-tropical 1000hPa geopotential height for the winter of 2006-2007.

Fig. 6. Reliability diagram of the NCEP 14 global ensemble raw forecast (black)

compared to the bias corrected forecast (red) for a 48 hour forecast of NH extra-tropical 1000hPa height for the winter of 2006-2007.

There are some differences between NCEP/GEFS and CMC/GEFS raw forecasts at the near surface (1000hPa geopotential height) when considering the ROC area (Figure 9). There is a similar result for 2-meter temperature for the 2006 CMC/GEFS model. There is much improvement after the CMC/GEFS implementation in July 2007 (personal communication). However, NAEFS still shows a significant improvement at all lead times.

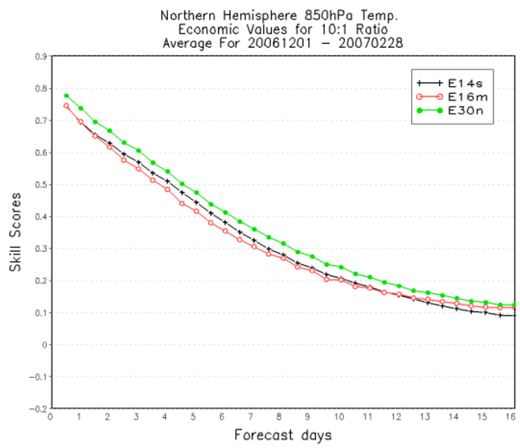


Fig. 7. ROC areas (from 0 to 1) for the NCEP 14 global ensemble raw forecast (black) compared to the CMC 16 global raw forecast (red) and the combined Fig. Hurricane track forecast from GEFS individual members for Hurricane Frances (2004) without tropical storm relocation.

Fig. 8. REV (10:1 loss/cost ratio) for the NCEP 14 global ensemble raw forecast (black) compared to the CMC 16 global raw forecast (red) and the combined NCEP and CMC ensemble (green) for NH extra-tropical 850hPa temperature for the winter of 2006-2007.

Economic values (Figure 10) really depend on the loss/cost ratio. These values vary when the loss/cost ratio changes (Zhu et al., 2002). In most cases, a 10:1 loss/cost ratio shows (near) maximum economic value.

Chapter 6: GEFS/NAEFS Developments

6.1 ETR Initialization and Cycling:

(Mozheng.Wei@noaa.gov)

6.1.1 Introduction

The initial perturbations for the GEFS are generated by the ET (Ensemble Transform) with rescaling or ETR. The ETR method was adopted and implemented successfully at NCEP on May 30, 2006 for operational forecasts. Due to the limitations on computing resources at the time of the implementation, the NCEP global GEFS ran only 56 ETR-generated members for the four daily cycles at 00Z, 06Z, 12Z and 18Z. At each cycle, only 14 members were integrated for the 16 day forecasts. The NCEP operational configuration has been switched to that described in Fig. 1 of this document since March 27, 2007. At every cycle, there are 20 long forecasts out of 80 ensemble members. More details about the ETR methodology can be found in Wei *et al. (*2005, 2006a, 2006b, 2008).

6.1.2 Basic formulation

Let

$$\mathbf{Z}^f = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1{}^f, \mathbf{z}_2{}^f, ..., \mathbf{z}_k{}^f], \qquad \mathbf{Z}^a = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1{}^a, \mathbf{z}_2{}^a, ..., \mathbf{z}_k{}^a] \qquad (1)$$

where the n dimensional state vectors $\mathbf{z}_i{}^f = \mathbf{x}_i{}^f - \mathbf{x}^f$ and $\mathbf{z}_i{}^a = \mathbf{x}_i{}^a - \mathbf{x}^a$ $(i = 1, 2, ...k)$ are k ensemble forecast and analysis perturbations respectively. In our experiments, $\mathbf{x}^f$ is the mean of k ensemble forecasts and $\mathbf{x}^a$ is the analysis from the independent NCEP operational DA system. Unless stated otherwise, the lower and upper case bold letters will indicate vectors and matrices, respectively. In the ensemble representation, the $n \times n$ forecast and analysis covariance matrices are approximated, respectively, as

$$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT} \quad \text{and} \quad \mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}, \qquad (2)$$

Where superscript T indicates the matrix transpose. For a given set of forecast perturbations $\mathbf{Z}^f$ at time t, the analysis perturbations $\mathbf{Z}^a$ are obtained through an ensemble transformation $\mathbf{T}$ such that

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \qquad (3)$$

In the ETR method, we want to use analysis error variances from the best possible DA system to restrain the initial perturbations for our GEFS. At NCEP, the best analysis error variances can be derived from the NCEP operational DA system which uses all kinds of real-time observations.

Suppose $\mathbf{P}_{op}{}^a$ is the diagonal matrix with the diagonal values being the analysis error variances obtained from the operational DA system, the ET transformation matrix $\mathbf{T}$ can be constructed as follows. For an ensemble forecast system, the forecast perturbations $\mathbf{Z}^f$ can be generated by equation (1). One can solve the following eigenvalue problem.

$$\mathbf{Z}^{f^T}\mathbf{P}_{op}^{a^{-1}}\mathbf{Z}^f = \mathbf{C}\,\mathbf{\Gamma}\,\mathbf{C}^{-1} \qquad (4)$$

Where $\mathbf{C}$ contains column orthonormal eigenvectors ($\mathbf{c}_i$) of $\mathbf{Z}^{f^T}\mathbf{P}_{op}^{a^{-1}}\mathbf{Z}^f$ (also the singular vectors of $\mathbf{P}_{op}^{a^{-1/2}}\mathbf{Z}^f$), and $\mathbf{\Gamma}$ is a diagonal matrix containing the associated eigenvalues (λ_i) with magnitude in decreasing order, that is, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k]$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}$ and $\mathbf{\Gamma} = diag(\lambda_1, \lambda_2, ..., \lambda_k)$.

The ET analysis perturbations can be constructed through transformation

$\mathbf{T}_p = \mathbf{C}\mathbf{G}^{-1/2}$, that is

$$\mathbf{Z}_p{}^a = \mathbf{Z}^f\mathbf{T}_p = \mathbf{Z}^f\mathbf{C}\mathbf{G}^{-1/2} \qquad (5)$$

From equation (5), it can be shown that the analysis perturbations are orthogonal with respect to an inverse analysis error variance norm. The analysis error covariance matrix can be approximated through analysis perturbations such as Eq. (2) if the number of ensemble members is large, i.e. when $k \to n$.

The sum of the analysis perturbations is

$$\sum_{q=1}^{k}(\mathbf{Z}_p{}^a)_{mq} = \sum_{q=1}^{k}\sum_{i=1}^{k}\mathbf{Z}_{mi}{}^f\mathbf{C}_{iq}g_q{}^{-1/2} = \sum_{i=1}^{k}\sum_{q=1}^{k-1}\mathbf{Z}_{mi}{}^f\mathbf{C}_{iq}\lambda_q{}^{-1/2} \neq 0 \qquad (6)$$

Equation (6) shows that the sum of all perturbations defined by equation (5) is not zero, although the forecast perturbations are centered. It is desirable that all initial perturbations are centered around the best possible analysis field in order to get better ensemble mean performance. Thus, $\mathbf{Z}_p{}^a$, which are the transformed perturbations by $\mathbf{T}_p$, are not centered. A transformation that will transform these perturbations into k centered perturbations and preserve ensemble analysis covariance $\mathbf{P}^a$ is the simplex transformation (ST).

In the practical implementation, we use the transformation $\mathbf{C}^T$, which is a by-product from the eigenvalue solution in equation (4), to act on all k perturbations $\mathbf{Z}_p{}^a$ to produce k centered perturbations with simplex structure. Therefore, the final ET solution with ST is

$$\mathbf{Z}^a = \mathbf{Z}_p{}^a\mathbf{C}^T = \mathbf{Z}^f\mathbf{C}\mathbf{G}^{-1/2}\mathbf{C}^T \qquad (7)$$

The sum of the final perturbations is:

$$\sum_{q=1}^{k}\mathbf{Z}_{mq}{}^{a}=\sum_{i=1}^{k}\mathbf{Z}_{mi}{}^{f}\sum_{q=1}^{k}\sum_{l=1}^{k-1}\mathbf{C}_{il}\mathbf{C}_{ql}\lambda_{l}{}^{-1/2}=\sum_{i=1}^{k}\mathbf{Z}_{mi}{}^{f}\sum_{l=1}^{k-1}\mathbf{C}_{il}\lambda_{l}{}^{-1/2}\sum_{q=1}^{k}\mathbf{C}_{ql}=0 \qquad (8)$$

This shows that all perturbations after ET and ST transformations are centered.

Since all perturbations are centered, they are not strictly orthogonal as they were before ST. The ideal initial perturbations in an ensemble system must be centered and span a subspace that has maximum number of degrees of freedom. Let's now look at the orthognality of the perturbations defined in equation (7) in the following.

$$\mathbf{J}=(\mathbf{P}^{a^{-1/2}}\mathbf{Z}^{a})^{T}(\mathbf{P}^{a^{-1/2}}\mathbf{Z}^{a})=\mathbf{Z}^{a^{T}}\mathbf{P}^{a^{-1}}\mathbf{Z}^{a}=\mathbf{C}\mathbf{G}^{-1/2}\mathbf{\Gamma}\mathbf{G}^{-1/2}\mathbf{C}^{T} \qquad (9)$$

Consequently equation (9) results in

$$\mathbf{J}=\mathbf{I}-\mathbf{c}_{k}\mathbf{c}_{k}{}^{T} \qquad (10)$$

This equation shows that $\mathbf{J}_{ii}=1-1/k$, and when $i\neq j$, we have

$$\mathbf{J}_{ij}=-\frac{1}{k}, \qquad \lim_{k\to\infty}(\mathbf{J}_{ij})=0 \qquad (11)$$

Equation (11) shows that for a finite number of ensemble members, the analysis perturbations after ET and ST transformations are not exactly orthogonal. The perturbations are uniformly centered and distributed in different directions. The larger the number of ensemble members, the more orthogonal the perturbations will become. If the number of ensemble members approaches infinity, then the transformed perturbations will be orthogonal under this norm.

The properties of the initial perturbations generated from equation (7) can be summarized as follows. **(a)** The initial perturbations will be centered around the analysis field to improve the score of ensemble mean. **(b)** They have simplex, not paired, structure. The ST, which preserves the analysis covariance, ensures that the initial perturbations will have the maximum number of effective degrees of freedom. The variance will be maintained in as many directions as possible within the ensemble subspace. **(c)** The perturbations are uniformly centered and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will become. **(d)** The initial perturbations have time and flow-dependent spatial structure if the analysis error variance is derived from operational DA system at every cycle. **(e)** The covariance constructed from the initial perturbations is approximately consistent with the analysis covariance from the DA if the number of ensemble members is large.

To make the initial spread distribution more similar to the analysis error variance, we impose a regional rescaling process on the initial perturbations generated by the above process, i.e., each initial perturbation after ET and ST from eq. (7) will be rescaled by the analysis error variance using

$$\mathbf{y}_{m}{}^{a}(i,j,l)=\alpha(i,j,l)\mathbf{z}_{m}{}^{a}(\mathrm{i,\ j,l}) , \qquad (12)$$

where α is the rescaling factor derived from analysis error variance; i, j, l are indices for the horizontal and vertical directions in grid point space; and $m = 1, 2, \ldots k$ is the index for the ensemble member.

6.1.3 Experimental results

Our original experiments run from 31 Dec 2002 to 17 Feb 2003, however, our study will focus on the 32-day period from 15 Jan 2003 to 15 Feb 2003. There are 10 ensemble members in both the ETKF and breeding-based systems. The observations used for ETKF are from the conventional data set in the NCEP global DA system. This conventional data set contains mostly rawinsonde and various aircraft data, and wind data from satellites. The ETKF results displayed in most figures are mainly for comparison with various ET experiments. We also ran 10-member ET experiments with (ETR) and without rescaling to compare with our previous experiments with breeding and ETKF.

In addition, we test ETR experiments with more members. In particular, we run an 80-member ETR at every cycle. However, due to the computing resource limit only 20 members will be integrated for long forecasts. The other 60 members are used only for cycling (integrated to 6 hours). At every cycle, both ET and ST are imposed on all 80 members, followed by ST on the 20 members used for the long forecasts. At different cycles, a different 20 members will be used for long forecasts. A schematic of this configuration is depicted in Fig. 1. All the ensembles are cycled every 6 hours in accordance with the NCEP DA system, in which new observations are assimilated in consecutive 6-hour time windows centered at 00, 06, 12 and 18 UTC.

One good measure of ensemble forecast performance is a direct comparison of the ensemble perturbations to the forecast errors. We have computed PECA values as described in Wei and Toth (2003) for all the ensemble systems mentioned in the previous paragraph. The PECA values for 500mb geopotential height for a 10-member ETR (solid), ET (dotted), breeding (dashed) and ETKF (dash-dotted) are shown in Figs. 2a, b, c and d for the globe, Northern and Southern Hemispheres, and the tropics. In each panel, the PECA for the optimally combined perturbations and the PECA averaged from individual perturbations are displayed in thick and thin lines, respectively.

In each of these regions, ETR (solid) has the highest average PECA values (thin lines) for short lead times, with breeding (dashed) next. The gap between ETR and breeding is even larger for the optimally combined perturbations (thick). This is due to the structural difference between the two methods. The perturbations in ET and ETR are simplex structures, while in breeding the positive/negative paired strategy is used. In a paired strategy, the effective number of degrees of freedom (EDF) of ensemble subspace is reduced by half by construction, while a simplex structure has a maximum EDF. It is interesting to see that the PECA values for both optimally combined and individual averages are similar for ET and ETKF.

It is noteworthy that the rescaling imposed on the ETR perturbations improves PECA values in almost all the domains we have chosen, particularly for the lead times up to a few days. In order to see the improvement in PECA from the increase of members, we compare a 10-member ET and a 20-of-80-member ETR (see Fig. 1 for the configuration). In Fig. 3, we show PECA values for the 10-member ETR (solid) and without (dotted) rescaling, the 20-of-80-member ET with (dashed) and without (dash-dotted) rescaling for Northern Hemisphere, North America, Europe and the globe. Again, the average PECA from the individual members and that from the optimally combined perturbations are indicated by thin and thick lines, respectively. It is clear that rescaling can increase the PECA value for a 20-member ensemble as well (see thick dashed and dash-dotted lines) as for a 10-member ET. Another message from this figure is that increasing the number of ensemble members will significantly increase the PECA value for optimally combined perturbations in all domains (thick solid vs. dashed line; dotted vs. dash-dotted).

Also plotted in Fig. 3 are the PECA values from the optimally combined perturbations for 80-member ET ensembles with (diamond) and without (square) rescaling at a 6-hour lead time. Since we have integrated only 20 members for the long forecasts due to computing resource limits, the remaining 60 members are integrated for only 6 hours, for cycling. Again, rescaling increases the PECA values for the ETR ensembles, especially for large domains like the Northern Hemisphere and the globe. The difference between ET and ETR ensembles is smaller for smaller domains, such as North America and Europe. The PECA value for ETR is about 0.9 and 0.95 for North America and Europe, respectively. This means that the 80-member ETR perturbations with rescaling can explain about 80% to 90% of forecast errors at 6-hour lead time. In all domains, the PECA values at a 6-hour lead time from the 80-member ETR are much larger than those from 20 members. This implies that the forecast error covariance at 6-hour lead times constructed from the 80-member ETR forecast perturbations will be a very good approximation to the real background covariance matrix, which can be used to improve DA quality.

Fig. 4 shows the Brier Skill Score (BSS) for 500mb geopotential height over the Northern Hemisphere, which is calculated by averaging the BSS for 10 climatologically equally likely events using climatology as a reference forecast. For shorter forecast lead times at least up to day 7, and for ensembles with 10 members ETR is best, while ETKF is the worst and breeding is in the middle. If we use 20 members out of the 80-member ETR as described in Fig. 1, its BSS value is higher than all the other experiments at all forecast lead times.

Shown in Fig. 5 is the ROCA (relative operating characteristic area) for the same experiments over the Northern Hemisphere. ROCA is a measure of discrimination. The results show that a 10-member ETR is better than 10-member breeding, while a 10-member ETKF has the lowest value of ROCA. Again, when the ensemble membership is increased to 20 members out of 80-member ETR, the ROCA is significantly higher than for all the other three experiments with only 10 members. We have also computed the EV for all these ensemble systems, which is shown in Fig. 6. In terms of EV, the 10-member ETR is similar to the 10-member breeding, and both are better than the 10-member ETKF. Again, the 20 out of 80 member ETR is better than all the other ensembles.

6.1.3 Future work

Future work is focusing on generating the time-dependent 3D initial analysis error variance $\mathbf{P}_{op}{}^{a}$. Efforts have been made in different directions, such as using the multi-center analysis data (Wei *et al.* 2009a), and estimating the analysis error variance directly from the NCEP operational data assimilation system (GSI) as described in Wei *et al.* *(*2009b).

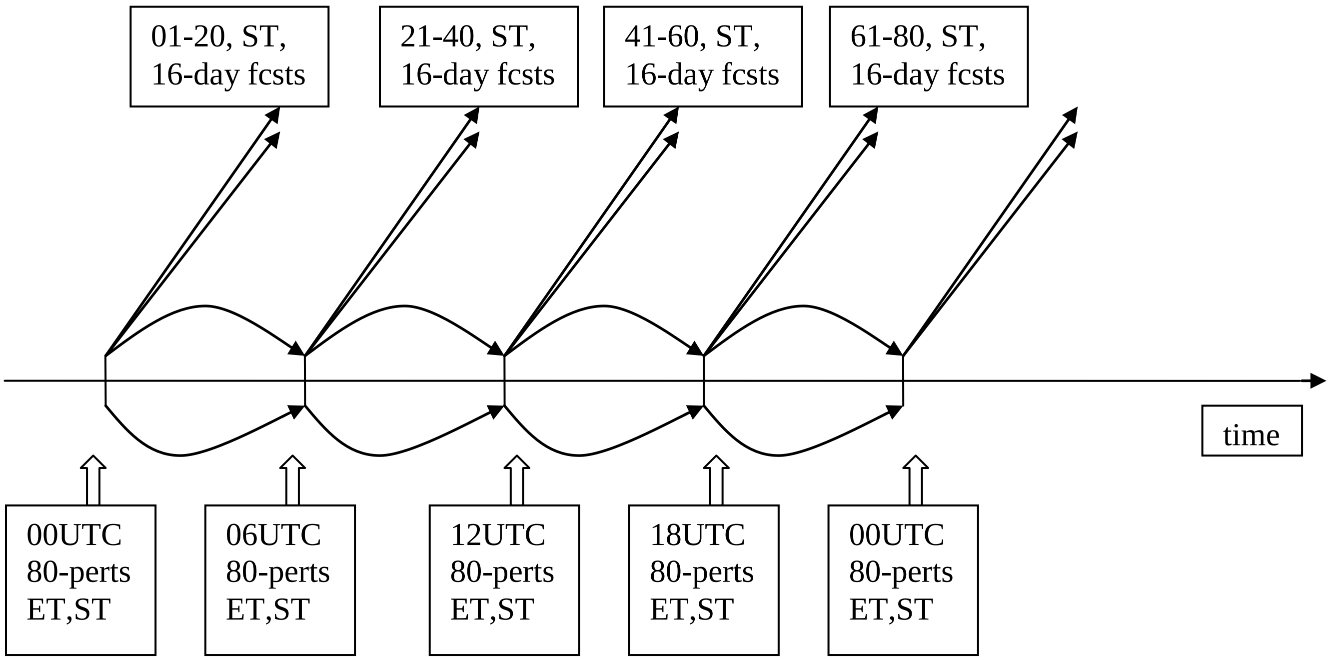


Fig.1. Schematic of the configuration of the 80-member ETR-based ensemble experiment. At each cycle ET transformation is carried out in all 80 perturbations, followed by the ST transformation. ST is also imposed on the 20 perturbations that will be used for long-range forecasts.

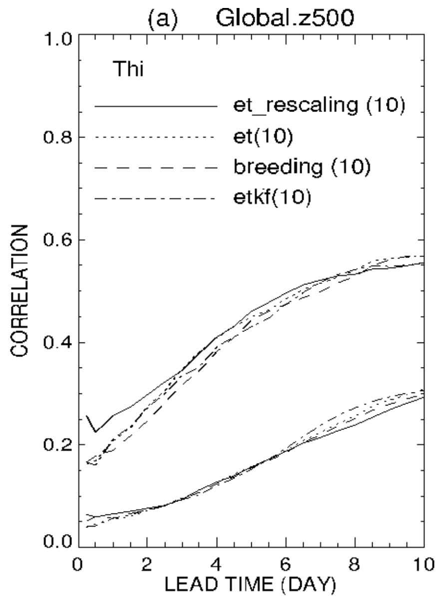


Fig. 2. PECA values for ETR (solid), ET without rescaling (dotted), breeding (dashed) and ETKF (dash-dotted) ensembles with 10 members for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from the individual perturbations, respectively.

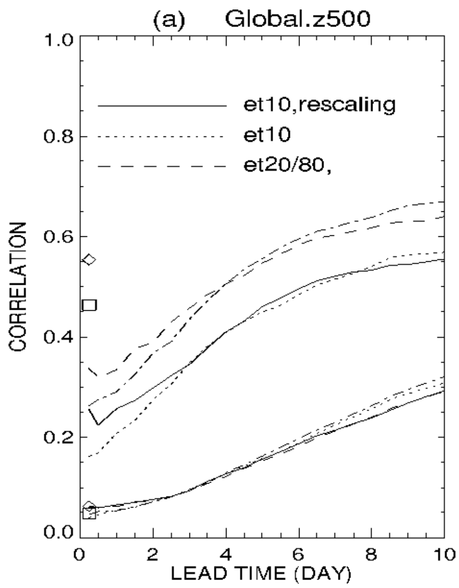


Fig. 3. PECA values for a 10-member ETR (solid), 10-member ET without rescaling (dotted), 20 of 80 member ET with rescaling (dashed) and 20 of 80 member ET without rescaling (dash-dotted) ensembles for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from individual perturbations, respectively.

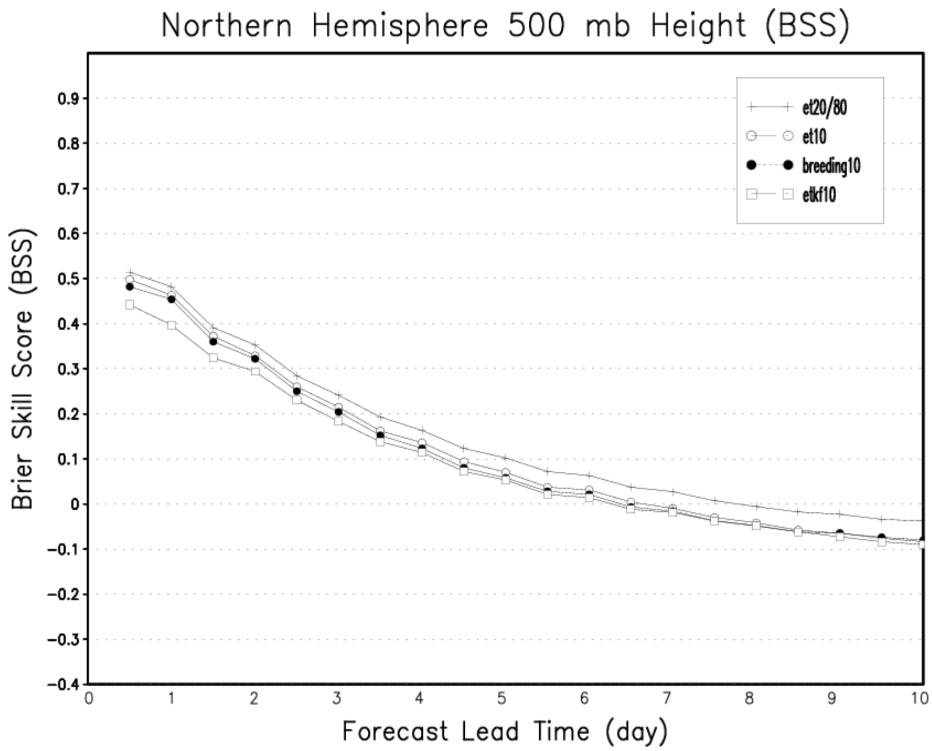


Fig. 4. Averaged Brier Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ETR (cross), 10-member ETR (open circle), 10-member breeding (full circle) and 10-member ETKF (open square) ensembles.

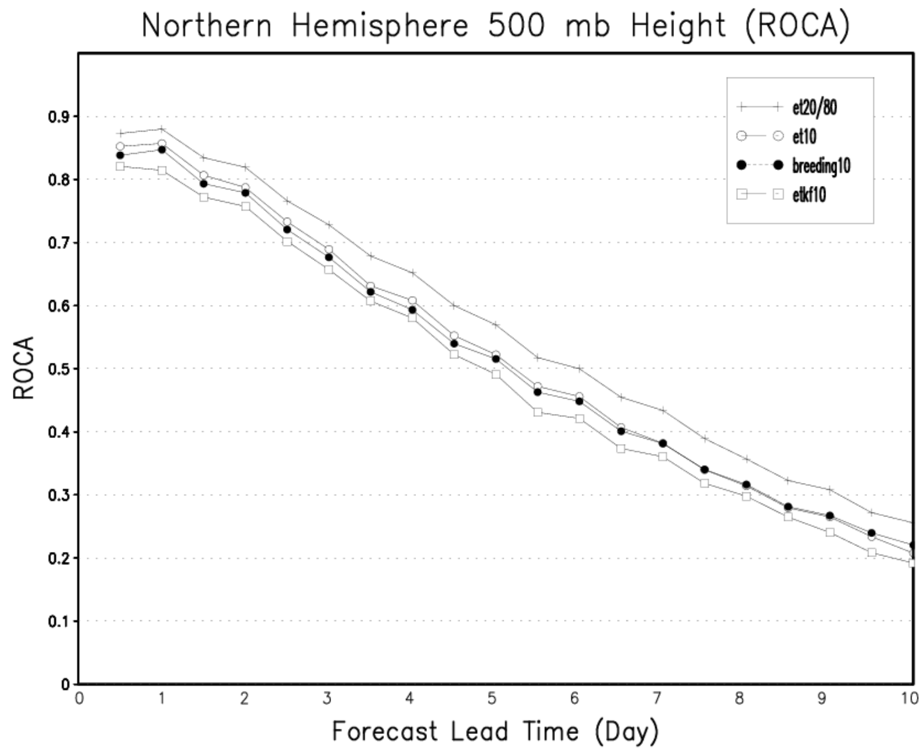


Fig.5. The same as Fig. 4, but for the relative operating characteristic area.

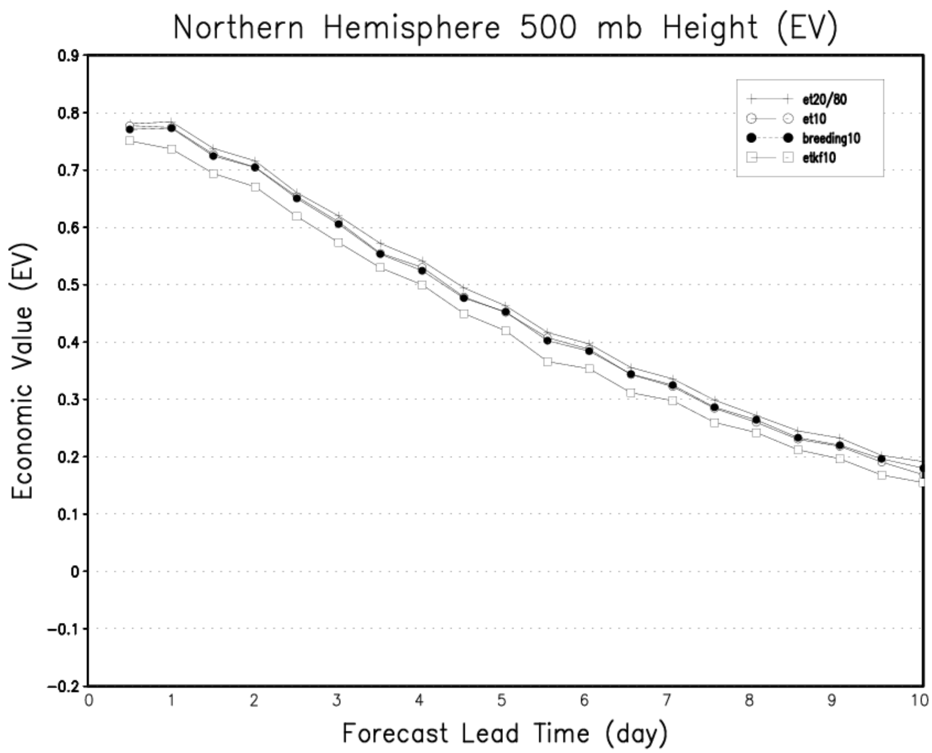


Fig. 6. The same as Fig. 4, but for the economic value.

6.2 Tropical Storm Relocation:

(Qingfu.Liu@noaa.gov)

6.2.1 Introduction

A hurricane relocation system was implemented in Global Forecast System (GFS) at the National Centers for Environmental Prediction (NCEP) in 2000 (Liu, et al., 2000). The hurricane relocation system moves the hurricane vortex in the model guess to the observed location before the 3D-VAR updates the analysis. It contains the following major steps: a) locate the hurricane vortex center in the guess field, b) separate the hurricane model's vortex from its environment field (Kurihara et al., 1995), c) move the hurricane vortex to the observed position, and d) if the vortex is too weak in the guess field, add a bogus vortex to the 3D-VAR data analysis (Lord, 1991). After the successful implementation of the hurricane relocation system in the GFS model, we tried to use it in the Global Ensemble Forecast System (GEFS). However, due to large differences in the hurricane structure among individual ensemble members, the relocation system does not have much effect on the statistics of the forecast tracks. In 2004 we modified the hurricane relocation system and added it to the GEFS. Test results show that the track spread from individual members was significantly reduced. The modified system was implemented in GEFS in July 2005 (Zhu et al., 2005). This paper explains the hurricane relocation system in the GEFS and shows some of our test results with and without hurricane relocation.

6.2.2 Hurricane Relocation in GEFS

The hurricane relocation system from the GFS model was modified to be used in GEFS. The new system can be summarized as follows: 1) Split the forecast fields from ensemble members (including the control) into environmental fields and hurricane components; 2) Compute global ensemble perturbations without the hurricane component (breeding cycle); 3) Compute hurricane perturbations (after relocating the hurricane to the observed location) for individual ensemble member P1 (hurricane perturbation for P1) = C*(XP1- Xn1)*|| Xc0||/|| XP1- Xn1|| ; 4) Add the hurricane perturbation and global ensemble perturbation to the analysis fields to create the model initialization. C is the scaling factor and XP1 (XN1) represents the 3D (or 2D) variables such as wind, temperature, mixing ratio and surface pressure for ensemble member P1 (N1). Xc0 represents the same variable for the model control. ||X|| is the square root of the summation of X2 over the whole hurricane area. The hurricane perturbation is scaled to be 5% of the magnitude of the control (C=0.05).

6.2.3 Test Results

After the modified hurricane relocation system was added to the GEFS model, we ran a series of experiments from 20040824 to 20040930. The test results are summarized in Figs 1, 2 and 3. Fig. 1 shows the differences in hurricane track forecast with and without the hurricane relocation for hurricane Frances (2004082800). You can see that there is a significant reduction in the initial track spread. Fig. 2 compares the statistical results of the track error and the track spread for all the forecasts with and without hurricane relocation. The initial reduction of the track spread is maintained throughout the 5 day forecast period. Fig. 3 shows the improvement in track forecasts compared to the operational models in the Atlantic and East Pacific Basins. The average track errors are smaller compared to the operational ensemble track forecasts throughout the forecast period, and the forecast tracks from the GEFS model are also better than those from the GFS model (higher resolution).

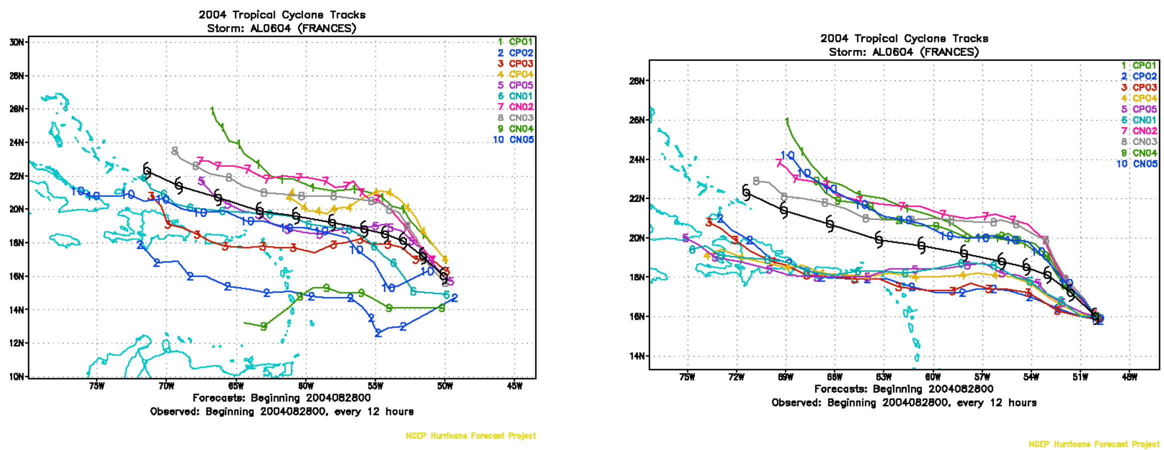


Fig. Hurricane track forecast from GEFS individual members for Hurricane Frances (2004) without tropical storm relocation (left) and with tropical storm relocation (right).

6.2.4 Summary and Discussion

The modified hurricane relocation system significantly reduces the spread of the track forecast in the GEFS model. However, it only slightly reduces the mean track error. The values of the global scaling factor and constant C used here are empirical factors, and need lots of ensemble experiments to determine their values.

6.3 A Stochastic Total Tendency Perturbation (STTP) Scheme with GEFS

(Dingchen.Hou@noaa.gov)

In early 2010, a Stochastic Total Tendency Perturbation (STTP) Scheme will be implemented with GEFS to represent uncertainties associated with the NWP model used for the integration. For further details of the scheme see Hou et al. (2006, 2008).

6.3.1 Basic assumption and strategy

Stochastic Tendency Perturbation (STTP) Scheme is based on the hypothesis that differences in the tendencies among the ensemble perturbations provide a representative sample of the random total model errors associated with the formulation of the dynamic and physical processed, truncation and parameterizations. The ensemble perturbation tendencies (represented by P) are first randomly combined following certain rules to form Stochastic (Total Tendency) Perturbations, which are then scaled to appropriate size and used as stochastic forcing terms in the model equations. Mathematically, in an ensemble system with N members, the stochastic perturbation for ensemble member i at time t is expressed as

$$S_i(t) \propto \sum_j^N w_{i,j}(t) P_j(t) \qquad \text{for i=1,2, … N} \qquad (1)$$

6.3.2 Simplifications in the current version

The current implementation of STTP is based on a simplified version which estimates the perturbation tendencies using finite differences with a time interval of a few hours. It is implemented by periodically stopping the concurrently running ensemble member integrations and perturbing the model states by adding the rescaled SPs. With a specified time interval of application, designated as Δt, STTP can be implemented by integrating the original model equation but modifying the model state variables (X) by using

$$X_i^{'} = X_i + \gamma(t)\sum_{j=1}^{N} w_{i,j}(t)\left\{\left[\left(X_j\right)_t - \left(X_j\right)_{t-\Delta t}\right] - \left[\left(X_0\right)_t - \left(X_0\right)_{t-\Delta t}\right]\right\} \qquad (2)$$

for i=1,2, …, N at t=Δt, t=2Δt, … . γ(t) is rescaling coefficient varying with time but uniform across all ensemble members. Its values are dependent on the choice of time interval Δt and its temporal variation is related to the temporal evolution of the ensemble perturbations.

Δt=6 hours is used in the current implementation while the generation of the random combination coefficients (w) and the specification of rescaling coefficient (γ) are described in the following paragraphs.

6.3.3 Generation of stochastic combination coefficients

Using matrix notation and omitting the time t, the relation (1) can be rewritten as

$$S_{NM} \sim W_{NN} P_{NM} \qquad (3)$$

where the subscripts indicate the dimensions of the matrix and M is the number of grid points. The problem of generating the combination coefficients is to specify a random, but orthonormal matrix W as a function of time. The temporal variation of the W matrix is represented by random rotations from one application to the next, or mathematically

$$W_{NN}(t) = W_{NN}(t-1) R_{NN}(t) \qquad (4)$$

where R is a random matrix only slightly different from the identity matrix I, representing a random and slight rotation of the N w vectors in an N-dimensional space. The rotation at a particular time, R(t), can be viewed as the combination of a steady rotation, which is represented by a random but temporally invariant Matrix R^0, and a random rotation R^1, which changes at every application of the scheme, i.e.,

$$R_{NN}(t) = R^0{}_{NN} R^1{}_{NN}(t-1) \qquad (5)$$

James Purser (personal communication) developed the methodology and software to generate a random orthonormal matrix and a random rotation matrix. Both procedures start with filling an NxN matrix A with independent random numbers from a Gaussian distribution and the orthonormalization is realized by applying the Gram-Schimidt procedure (e.g. Golub and Van Loan, 1996) to A. The rotation matrices R^0 and R^1 are generated by applying the same procedure to $(I+\alpha(A-A^T))$ where I is the identity matrix, A^T the transpose of A, and α the "amount" of rotation. These routines are used to generate the temporally varying weighting matrix W via the following procedures: (1) Initializing W by generating a random orthonormal matrix W(t=0); (2) specify the fractional numbers, α_0 to prescribe the "degree" of rotation in the steady rotation and generate R^0; (3) specify another fractional number α_1 for the degree of random rotation to find R^1; (4) at each time the stochastic perturbation scheme is applied, generate a random slight rotation matrix using the same α_0 and α_1 but different seed, and use (5) and (4) to update the W matrix.

The temporal evolution of the weighting matrix W can be viewed as N vectors in the N-dimensional space, changing their directions slightly with random vibrations (R^1)

imposed on a steady rotation (R^0). Similarly, the evolution of each weighting factor $w_{i,j}$ is a random walk (corresponding to R^1) superimposed on a periodic function (corresponding to R^0) with the level of noise (due to the random walk) and the period controlled by α_0 and α_1, respectively. α_0 and α_1 are the only two parameters required to specify W(t). While higher value of α_1 defines noisier curves, larger α_0 is corresponding to shorter period. Fig. 1 depicts some examples of these curves in a 10 member (N=10) ensemble system, showing the curves for i=10 and j=1, 2, .., 10, i.e., the temporal variation of the weight factors to determine the stochastic forcing for ensemble member 10. In this particular case (same as the current implementation) with $\alpha_0=\alpha_1=0.05$, the period is about 6 days and the curves look fairly noisy. For reference, $\alpha_1=0.005$ defines smooth curves while $\alpha_0=0.005$ is corresponding to a much longer period (>10 days).

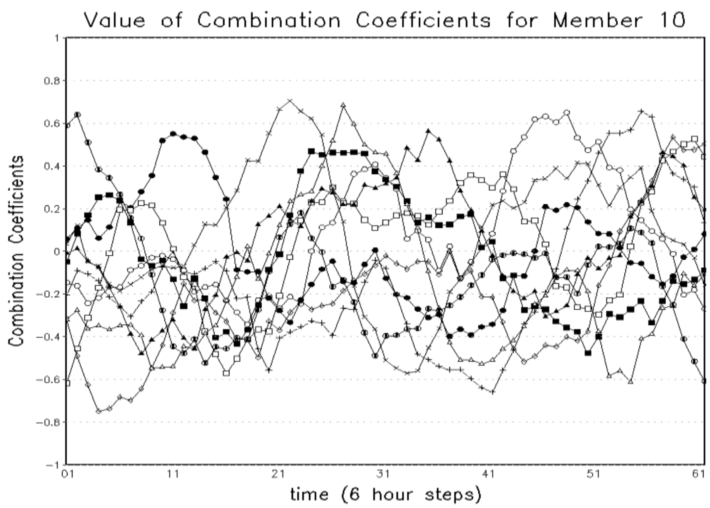


Fig.1 Examples of combination coefficients $w_{i,j}$ as a function of forecast lead time, in a 10 member (N=10) ensemble system. Shown are 10 curves for j=1,2,…,10 with fixed i=10 and $\alpha_0=\alpha_1=0.05$.

6.3.4 Specification of rescaling factors

The rescaling coefficient g is factorized as a global rescaling factor γ_0 and a regional rescaling factor γ_1, i.e,

$$\gamma = \gamma_1(\varphi, d)\gamma_0(t) \tag{6}$$

The global rescaling factor is a function of forecast lead time only and expressed as

$$\gamma_0(t) = \pm[p_2 + (p_1 - p_2)\left\{1.0 - \frac{1.0}{1.0 + e^{-p_3(t-p_4)}}\right\}] \tag{7}$$

For the Feb 2010 implementation, with T190L28 resolution throughout the integration from 0h to 384h, the parameters in (7) are specified as

$p_1=0.100$, $p_2=0.01$, $p_3=0.11$ and $p_4=252$ hours. (8)

For the Jan 2012 Implementation, with T254L42 resolution from 0h to 192h and then reduced to T190L42, the parameters used are

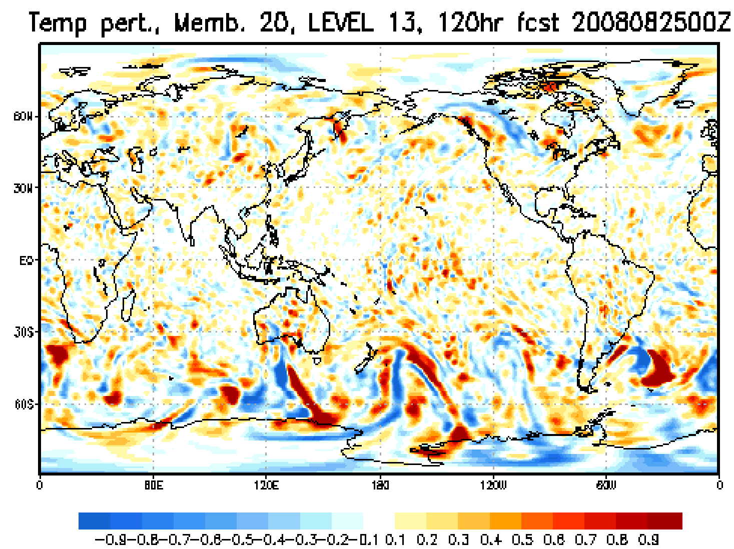
$p_1=0.105$, $p_2=0.03$, $p_3=0.12$ and $p_4=252$ hours. (9)

Experiments with horizontal resolution of T126, T190 and T254 suggest that (8) and (9) are generally suitable for uniform horizontal resolution and variable resolution with truncation at 180-192h, respectively, although fine tuning and optimization may ne necessary.

The regional rescaling factor, in its current form, changes with latitude and the day of the year (season):

$$\gamma_1(\varphi, d) = 1.0 + 0.2\sin(\varphi)\cos\frac{2\pi d}{364} \tag{10}$$

Equation (10) indicates that the perturbation size in the winter hemisphere is larger than that in the summer hemisphere. As shown as an example in fig.2, the stochastic forcing vectors have structures of random noise and its size, represented by a vector norm similar to total energy, shows a flow-dependent global distribution with largest amplitudes associated with the mid-latitude jets in both hemispheres.



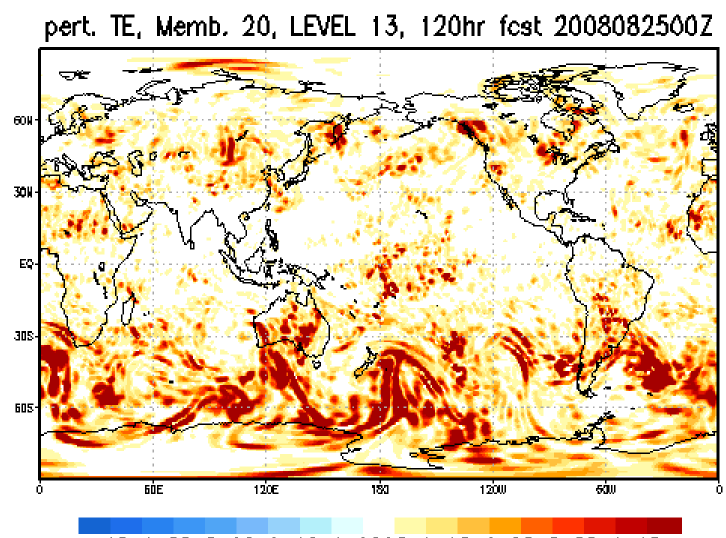


Fig. 2 An example of stochastic perturbations added to modify the model states of the 20-member NCEP Global Ensemble Forecasting System. Shown are (upper panel) the temperature perturbation (unit: K) associated with ensemble number 20, and (lower panel) the corresponding perturbation size defined as the square root of "total energy" norm of the perturbation vector (unit: ms^{-1}), at 120h integration time starting from 00Z, Aug. 25, 2008.

6.3.5 Performance:

The inclusion of STTP significantly increases the spread of the ensemble, reduces outliers, reduces systematic errors of the ensemble mean forecast, and improves ensemble-based probabilistic forecasts as well as the ensemble forecast distribution. Forecast improvement is more consistent in the tropics than in the extratropics, and more prominent in the cool season than in the warm season. In the tropics, the scheme improves the forecast by increasing both the statistic reliability and the resolution, while the resolution is hardly affected in the extratropics. In addition, the impact of the scheme is complementary to model improvements in formulation and increase in spatial resolution. Fig.3 compares the Continuous Ranked Probability Skill Score (CRPSS) of H500 and T850 forecast from ensembles with and without STTP.

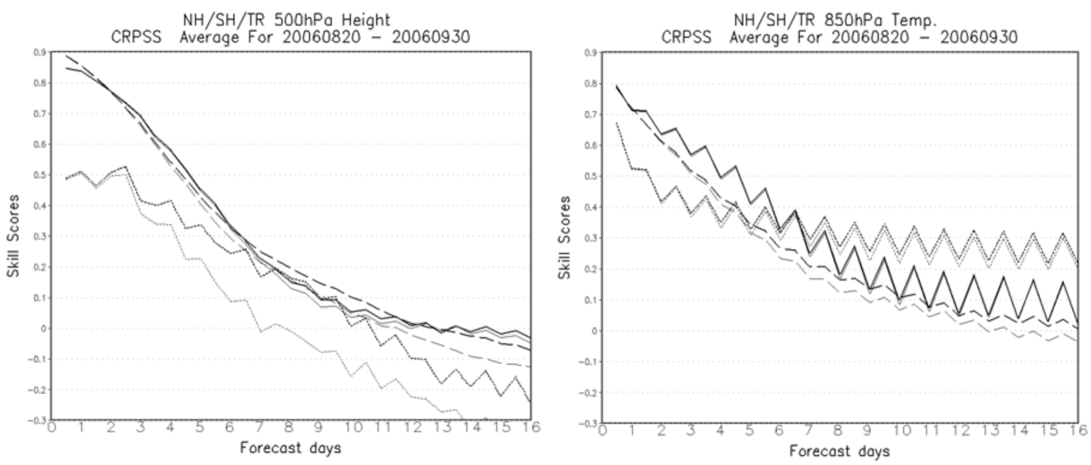


Fig. 3 CRPSS of 500 hPa geopotential height (left panel) and 850 hPa temperature, averaged over the Northern (solid) Southern (long dash) hemisphere Extratropics and the Tropics (short dash) for the ensembles without (light lines) and with STTP (dark lines), for the Northern summer period of Aug. 20 to Sep. 30, 2008.

6.3.6 Software Engineering

Currently, the implementation of STTP requires concurrently running N perturbed members and a control run, all at the same resolution. The concurrency and synchronism of all integrations are facilitated by the Earth System Modeling Framework (ESMF) software system. A user can modify the parameter list to change the time interval Δt, the parameters α_0 and α_1 discussed in 6.3.3, and p_1, p_2, p_3 and p_4 in equation (7).

6.4: Summary of GEFS/GFS bias correction

(Bo.Cui@noaa.gov and Yuejian.Zhu@noaa.gov)

EMC has experimentally summary for GFS bias correction since 2006, which is using the same algorithm as global ensemble forecast system (GEFS) bias correction. The description of the method is as following:

6.4.1 Bias Estimation:

The bias (b) for each lead-time (t) (6-hour interval up to 180 hours), each grid point (i,j) is defined as the different of analysis (a) and forecast (f) at the same valid time (t0) which is up on latest available analysis.

$$b_{i,j}(t) = f_{i,j}(t) - a_{i,j}(t_0) \qquad (1)$$

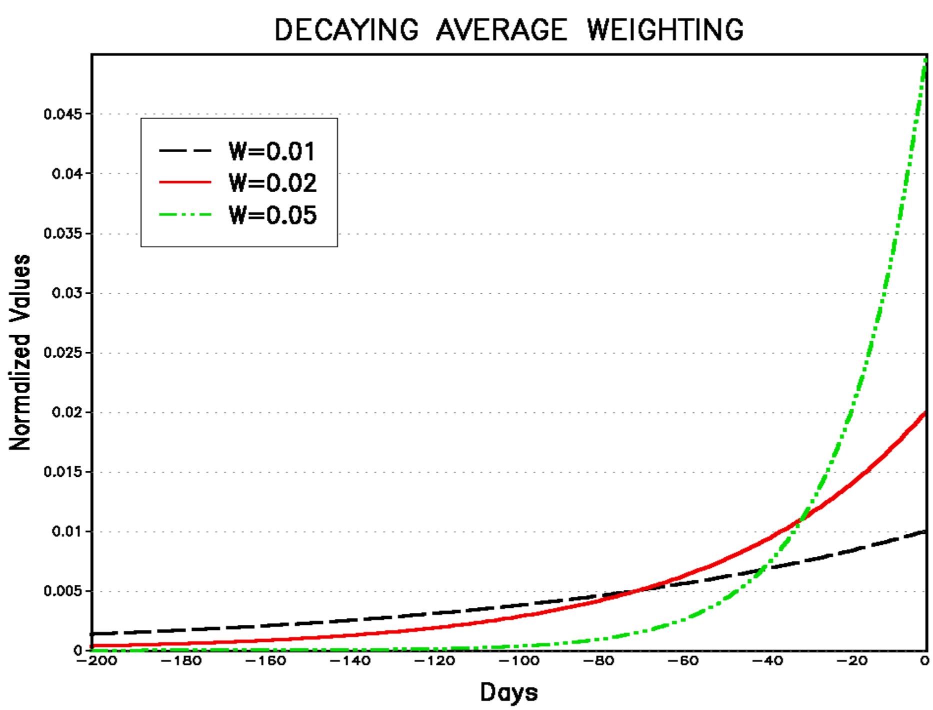
6.4.2 Decaying Average:

Average bias will be updated by considering prior period bias and current bias by using decaying average with weight coefficient (w).

$$B_{i,j}(t) = (1-w)B_{i,j}(t-1) + wb_{i,j}(t) \qquad (2)$$

6.4.3 Weight:

By previous experiments for different weights (0.01, 0.02, 0.05, 0.1 and etc…), w equals to 0.02 has been used for GEFS bias correction which is mainly using past 50-60 days information (see figure).



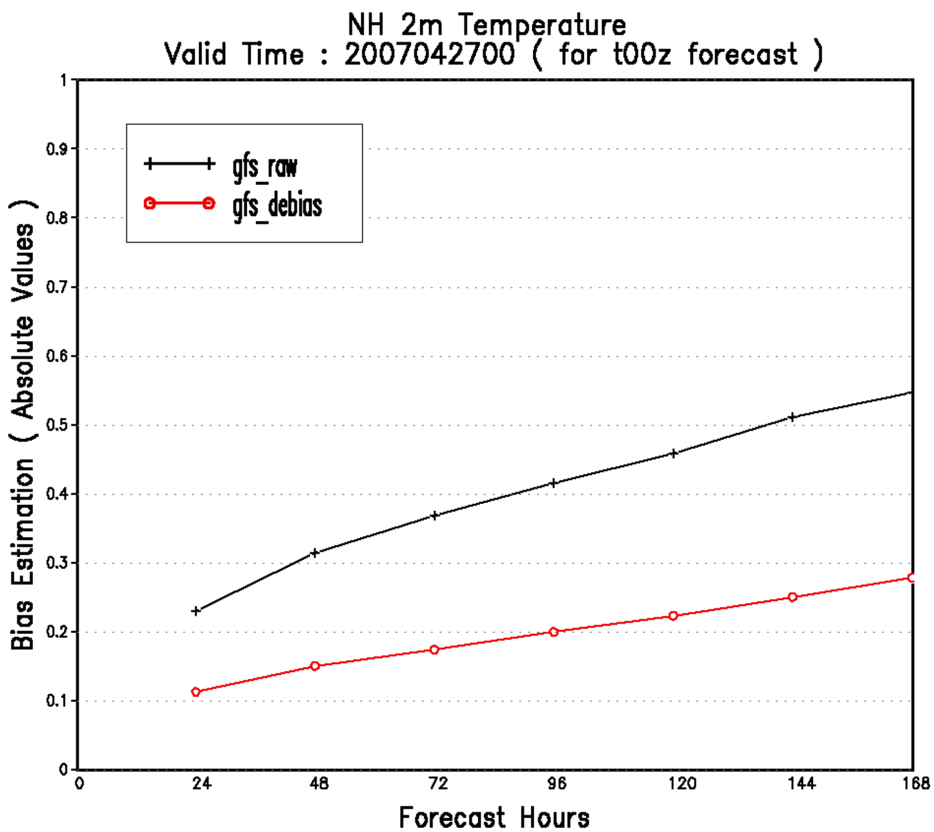
6.4.4 Bias Corrected Forecast:

The new forecast (bias corrected) will be generated by applying decaying average bias (B) to current forecasts at each lead-time and each grid point.

$$F_{i,j}(t) = f_{i,j}(t) - B_{i,j}(t) \qquad (3)$$

6.4.5 Performance:

The performance is estimated by applying bias correction method. The bias is calculated at each grid point through equation (1) for raw forecast (f) and bias corrected forecast (F), then using decaying average method (w=0.02) to get current average bias, taking absolute bias for each grid point, each lead-time to generate domain average absolute error (bias) which smaller value is better (see figure: example for Northern Hemisphere 2 meter temperature, decaying average (*w*=0.2) about 2 months period ended by April 27, 2007).



6.5: Dual-resolution (hybrid)

(Bo.Cui@noaa.gov and Yuejian.Zhu@noaa.gov)

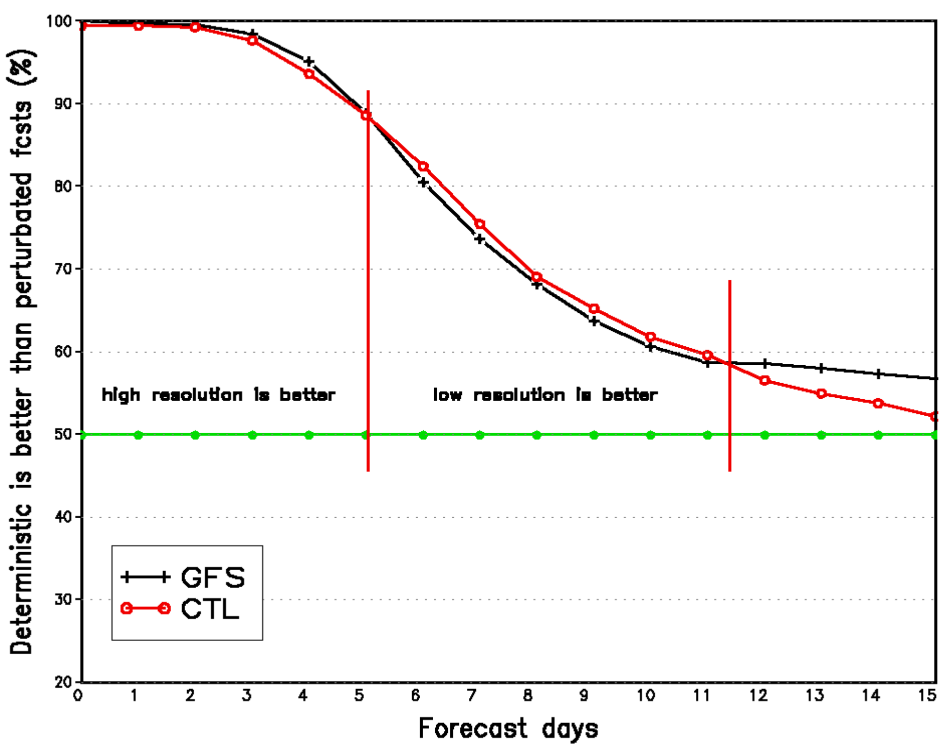
EMC has experimentally summary by combined GFS bias corrected forecast and GEFS bias corrected forecast for the forecast lead-time up to 180 hours. The description of the method is as following:

6.5.1 GFS Bias Corrected Forecast:
Please refer to Section 6.4.

6.5.2 GEFS Bias Corrected Forecast:
Implemented by May 30 2006.

6.5.3 Hybrid GFS and GEFS Bias Corrected Forecasts:

a). Why do we need hybrid? Because GFS performances consistently better than lower resolution (ensemble control at T126L28 resolution) forecast for short lead-time, the example from one-year statistics shows the GFS takes the advantage up to 120 hours (see figure below and reference from Zhu, 2005):



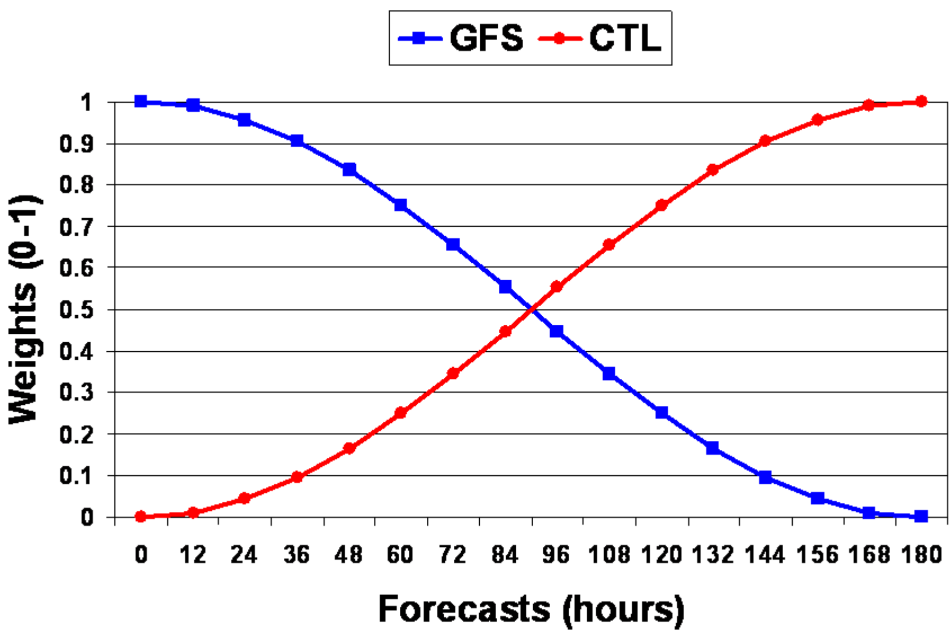
b). In order to combine GFS bias corrected forecasts and GEFS bias corrected forecasts for the first 180 hours, a cosine weighting function has been used to weight GFS and ensemble control (CTL at T126L28 resolution), to have highest weights for GFS at short lead time, to have it smoother or continues to ensemble forecast when the lead-time close to 180 hours. Here is the formula for each ensemble forecast:

$$f_i^*(t) = f_{gfs}(t) \cdot w_{gfs}(t) + f_{ctl}(t) w_{ctl}(t) + (f_i(t) - f_{ctl}(t)) \qquad (i=1,2,\ldots\ n)$$

Where

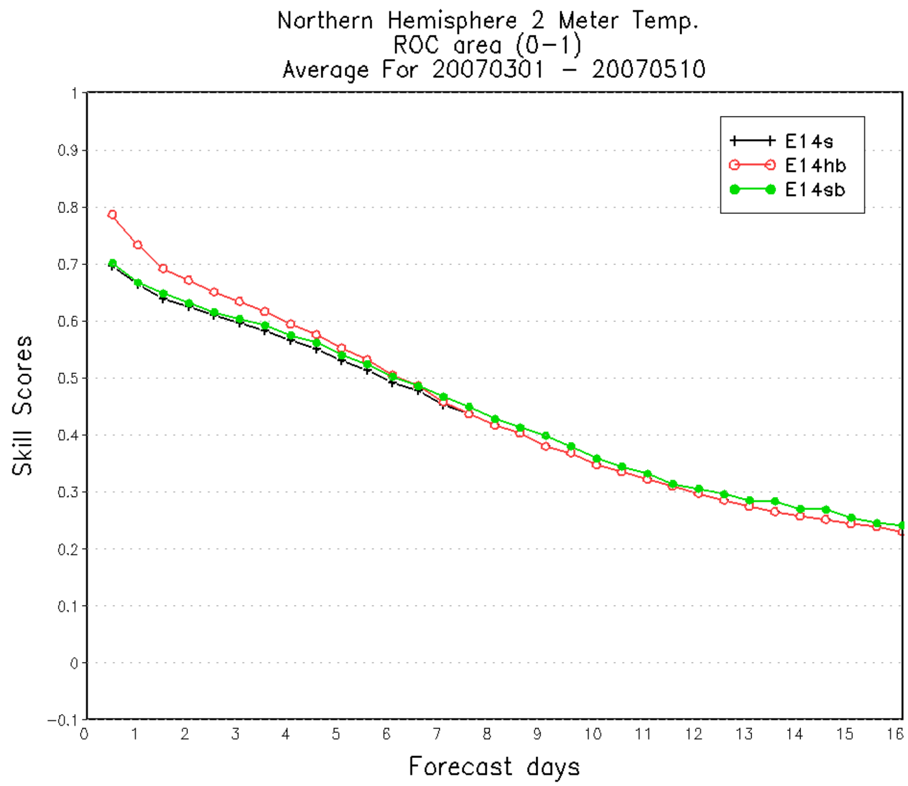
$w_{gfs}(t) = \dfrac{\cos(t)+1}{2}$ and $w_{ctl}(t) = 1 - w_{gfs}(t)$ (t represents forecast hours from 0 to 180)

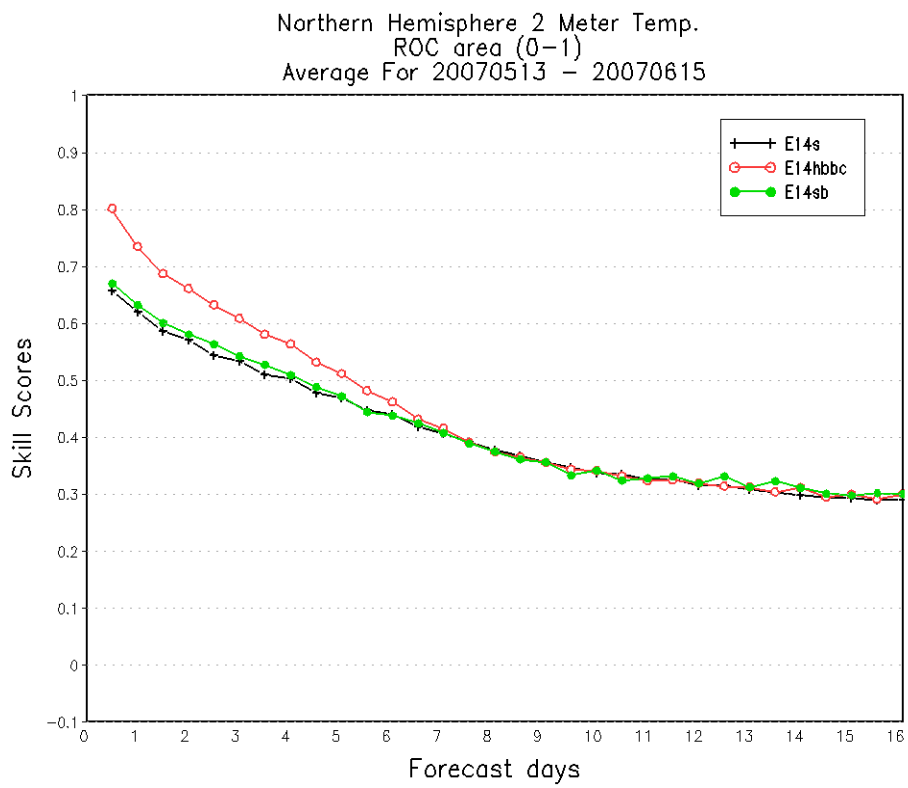
And figure shows weighting function with lead time.



6.5.4 Performance:

The performance is estimated by applying above method for both raw and bias corrected forecast from GFS, ensemble control and ensembles. The improvement depends on the variables. For example, there is significant improvement for Northern Hemisphere 2-meter temperature by applying this method (top figure below). E14s is for raw forecast skill, E14hb is for hybrid forecast skill, and E14sb is for bias corrected forecast during 3 month period. Bottom figure shows one month statistics while comparing E14s, E14sb with E14hbbc which is hybrid bias corrected GFS and ensembles.


Northern Hemisphere 2 Meter Temp.
ROC area (0-1)
Average For 20070301 - 20070510
Skill Scores
Forecast days
E14s
E14hb
E14sb


Northern Hemisphere 2 Meter Temp.
ROC area (0-1)
Average For 20070513 - 20070615
Skill Scores
Forecast days
E14s
E14hbbc
E14sb

6.6: Statistical Down-scaling for NAEFS Ensemble Forecasts

(Bo.Cui@noaa.gov and Yuejian.Zhu@noaa.gov)

Statistical down-scaling method has been applied to NCEP global ensemble from 1.0 degree to National Digital Guidance Database (NDGD-5km) resolution. There are four variables (2-meter temperature, surface pressure, 10-meter u and v) for this application in current NCEP operation. Statistical down-scaling method could be explained as following three main steps.

6.6.1 True or Reference:

High resolution analysis, such as Northern American Real Time Meso-scale Analysis (RTMA) on National Digital Forecast Database (NDFD) grid (5km), could be one of the references for statistical down-scaling.

6.6.2 Down-scaling Vector (DV):

In order to get DV, GDAS analysis at 5km resolution need to be generated from 1*1 degree resolution by using bilinear interpolation.

$$DV^{5km}(t_0) = (1-w)DV^{5km}(t_{-1}) + w(GDAS^{5km}(t_0) - RTMA^{5km}(t_0))$$

Where w is the weight to be used as decaying average (see the figure):

6.6.3 Down-scaled Forecasts (DF):

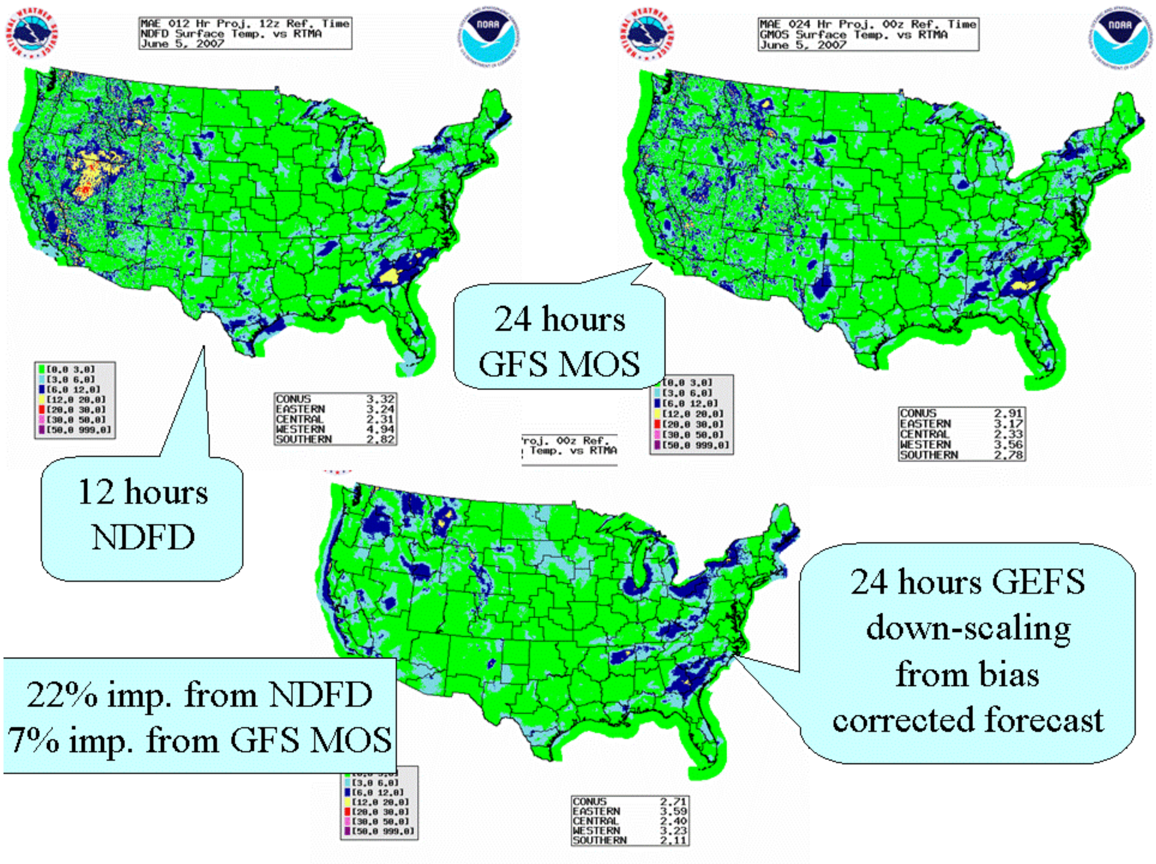
In order to get DF, bias corrected ensemble forecasts (BF) (or 10%, ensemble mode, and 90% probabilistic forecasts) at 5km resolution need to be generated by using bilinear interpolation. Then:

$$DF^{5km}(t) = BF^{5km}(t) - DV^{5km}(t_0)$$

Where t_0 is initial time of forecast, t is forecast lead-time. There are 4 *DVs* available for valid time ($t_0 + t$) at 00UTC, 06UTC, 12UTC and 18UTC.

6.6.4 Example of Down-scaling Performance:

There is a example of 24-hour 2-meter temperature down-scaling forecast for NCEP GEFS mean to compare with NDFD 12-hour forecast, and GFS-MOS 24-hours forecast (Figures are provided by MDL/OST). There are more statistics which will be posted on web-page.



Chapter 7: References

Hou, Dingchen, Zoltan Toth, Yuejian Zhu and Weiyu Yang, 2008: Impact of a Stochastic Perturbation Scheme on NCEP Global Ensemble Forecast System. 19th AMS conference on Probability and Statistics. New Orleans, LA, 20-24 Jan. 2008. [Available on line at http://www.emc.ncep.noaa.gov/gmb/ens/ens_info.html] (pdf)

Hou, Dingchen, Zoltan Toth and Yuejian Zhu, 2006: A Stochastic Parameterization Scheme within NCEP Global Ensemble Forecast System. 18th AMS conference on Probability and Statistics. Atlanta, GA, Jan. 29-Feb. 2, 2006. [Available on line at http://www.emc.ncep.noaa.gov/gmb/ens/ens_info.html] (pdf)

Golub, G. H. and C. F. Van Loan, 1996: Matrix Computations (3rd. ed). John Hopkins, ISBN 9780008018-5414-9.

Kurihara, Y., M. A. Bender, R.E. Tuleya and R. J. Ross, 1995: Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev*., **123**, 2791-280.

Liu, Q., T. Marchok, H.-L. Pan, M. Bender and S. Lord: Improvements in Hurricane Initialization and Forecasting at NCEP with Global and Regional (GFDL) models.

NCEP Office Note 472. Corresponding author address: Qingfu Liu, NCEP/EMC, 5200 Auth Road, Camp Springs, MD 20746. E-mail address: Qingfu.Liu@noaa.gov

Lord, S.J., 1991: A bogussing system for vortex circulations in the National Meteorological Center global forecast model. *Preprints of the 19th Conference on Hurricane and Tropical Meteorology*, Miami, FL., 329-330

Wei, M. and Toth, Z. 2003. A new measure of ensemble performance: Perturbations versus Error Correlation Analysis (PECA). *Mon. Wea. Rev*., **131**, 1549-1565.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. 2005. Initial Perturbations for NCEP Ensemble Forecast System. Thorpex Symposium Proceedings for t*he First THORPEX Internal Science Symposium* 6-10 December 2004, Montreal, Canada. The Symposium Proceedings in a WMO Publication 2005, ***WMO TD No.1237***, **WWRP THORPEX No. 6**, 2005. p227-230.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. and Wang, X. 2006a. Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus,* **58A,** 28-44.

Wei, M., Toth, Z., Wobus, R., Zhu, Y. 2006b. Initial Perturbations based on the Ensemble Transform (ET) Technique in the NCEP global Ensemble Forecast System. *US Department of Commerce, NOAA/NCEP Office Note* **No. 453,** 33pp**.**

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the Ensemble Transform (ET) technique in the NCEP global operational forecast system. *Tellus* 60A, 62-79.

Wei, M., Z. Toth, and Y. Zhu, 2009a: Analysis differences and error variance estimates from multi-center analysis data. *The Australian Meteorological & Oceanographic Journal* (in press).

Wei, M., M. Pondeca, …… 2009b: Estimating observation impact signals in NCEP GSI using the Lanczos method. (ready for internal review at EMC).

Zhu, Y., Z. Toth, R. Wobus, Q. Liu, 2005: July 2005 Upgrade of the NCEP Global Ensemble Forecast System. NCEP note.