A Stochastic Total Tendency Perturbation Scheme Representing Model-Related Uncertainties in the NCEP Global Ensemble Forecast System

Dingchen Hou[*]*, Zoltan Toth***, *Yuejian Zhu*
Weiyu Yang and Richard Wobus*[*]

Environmental Modeling Center/NCEP/NOAA, Camp Springs, MD, U.S.A.
* Also affiliated with I. M. Systems Group, Inc.
**Current Affiliation: Global System Division/ESRL/OAR/NOAA

Nov. 10, 2010

To be submitted to Tellus

Corresponding author:
Dingchen Hou
Environmental Modeling Center/NCEP/NOAA
W/NP2 NOAA WWB #207
5200 Auth Road
Camp Springs MD 20746
(301)7638000 ext. 7015
email: dingchen.hou@noaa.gov.

Abstract

A stochastic representation of random errors associated with the numerical weather prediction model used in ensemble prediction systems is described and its impacts in the NCEP Global Ensemble Forecast System (GEFS) documented. In the proposed Stochastic Total Tendency Perturbation (STTP) scheme, stochastic forcing terms are formulated by randomly combining the conventional total tendencies of ensemble perturbations and rescaling them to appropriate sizes. The perturbation tendencies are estimated using finite differences with a time interval of 6 hours in its current implementation. Extensive experiments were performed and the results show that the scheme can significantly increase the ensemble spread while reducing outliers and systematic errors in the ensemble mean forecast and improving ensemble-based probabilistic forecasts as well as the ensemble forecast distribution. Forecast improvement is more consistent in the tropics than the extratropics, and more prominent in the cool season than the warm season. In the tropics, STTP improves the forecast by increasing both the statistical reliability and the resolution, while the resolution is hardly affected in the extratropics. In addition, the impact of the scheme is independent of model improvements in formulation or increase in spatial resolution. The scheme has been used in GEFS operational production since Feb. 23, 2010.

1. Introduction

Ensemble prediction systems (EPS) were introduced into operational numerical weather prediction (NWP) in the early 1990s (e.g., Toth and Kalnay 1993; Tracton and Kalnay 1993; Palmer et al. 1993; Molteni et al. 1996) as a practical and successful way of addressing the predictability problem associated with the uncertainty in initial conditions (e.g., Leith 1974). By employing various schemes such as Breeding Vectors (Toth and Kalnay 1993), Singular Vectors (Buizza and Parlmer 1995) and ensemble-based techniques (e.g., Hotekamer and Mitchell, 1998 and Wei et al., 2008) to generate initial perturbations, the EPS systems of the world's major operational forecast centers provide ensemble mean forecasts with an accuracy comparable to or higher than traditional single deterministic forecasts and ensemble-based probabilistic forecasts with information about forecast uncertainty. However, EPS systems based on a "perfect-forecast-model assumption" have a common shortcoming: the root-mean-square (RMS) spread of the ensemble grows significantly slower than the RMS error of the control forecast or ensemble mean forecast (Buizza et al. 1999; Buizza et al. 2005). This has been attributed to the neglect or inadequate representation of the errors associated with the NWP model (Harrison et al. 1999; Vannitsem and Toth 2002). Vannitsem and Toth (2002) suggests that random model errors are introduced throughout the course of numerical integrations and just like errors in initial conditions, project onto fast growing perturbation directions associated with Lyapunov vectors.

Efforts have been made to represent model-related uncertainties in EPS since the mid-1990s. Multi-model and multi-model-version schemes have been employed in both operational systems (e.g., Houtekamer et al. 1996; Du and Tracton 2001) and experimental tests (e.g., Krishnamurti et al. 1999; Stensrud et al. 2000; Hou et al. 2001). These ad hoc schemes employ a procedure where ensemble members differ from each other not only in their initial conditions, but also in

subgrid-scale parameterizations of, for example, horizontal diffusion, convection, radiation, and gravity wave drag. In addition, they may be integrated with different ways of representing orography. While providing practical solutions to the problem of representing model error, the multi-model approach has limitations due to the limited number of available NWP models (versions) and parameterization schemes, a lack of theoretical foundation for their selection, and the requirement for significant resources in maintaining the different models or model versions.

An alternative approach to representing model errors in EPS is to develop formulations that impose stochastic terms on the tendency of the model equations. As a first attempt to simulate the effect of model errors, Toth and Kalnay (1995) inflated the ensemble perturbations during the integration of the ensemble members to increase the ensemble spread, and their preliminary results suggest this approach could lead to a better agreement between ensemble spread and error and the error of the control run, and more skillful ensemble mean. The first operational application of EPS with model errors was the use of a stochastic representation of random errors associated with parameterized physical processes (Stochastic Physical Parameterization, or SPP) in the European Center for Medium-Range Weather Forecasts (ECMWF) EPS (Buizza et al. 1999). By multiplying the total parameterized tendencies by a random number sampled from a uniform distribution between 0.5 and 1.5, and fixed over adjacent 5x5 lat/lon boxes and over 12-hour periods, the scheme has some effect in increasing the ensemble spread and improving the skill of the probabilistic prediction of certain weather parameters. Stochastic backscatter tested with the UK Met Office model (Frederiksen and Davies 1997) and the kinetic energy backscatter algorithm (Shutts 2005; Berner et al. 2009) applied to the ECMWF EPS also had some beneficial impact on forecast skill and synoptic variability.

The SPP approach has received more attention with emphasize on the nature of the parameterization schemes for individual physical processes. Vannitsem and Toth (2002), for example, argue that each component of numerical models must preferably be constructed in such a way that they can represent model-related uncertainties, including those associated with subgrid-scale processes not directly represented in numerical models, when used in an ensemble mode. NWP models used currently at major operational centers have little or no related capabilities. Teixeira and Reynolds (2008) investigated physical parameterizations and pointed out that such parameterizations are inherently stochastic in nature. In fact, the notion of a probability distribution has been central to the boundary layer and turbulence parameterizations (e.g., Golaz et al. 2002), and cloud parameterization methods based on the probability density functions (pdf) of moisture conserved thermodynamics variables have been advocated and implemented in weather and climate prediction models (e.g., Bony and Emanuel 2001; Tompkins 2002; Teixeira and Hogan 2002). Note that most NWP models have been developed and tested in the context of a single unperturbed forecast. In an attempt to reduce random errors, most parameterization schemes, therefore, ignore and suppress the role of stochastic elements. For successful ensemble prediction, however, the representation of such effects becomes imperative. Teixeira and Reynolds (2008) developed a stochastic convection parameterization (SCP) scheme that included the impact of subgrid-scale variability due to deep convection in the Navy Operational Global Atmospheric Prediction System (NOGAPS) ensemble formulation. This is done by multiplying the parameterized tendency due to moist convection by a normally distributed stochastic variable. One open question in the application of SCP or other SPP schemes is how to specify the horizontal, vertical and temporal correlations, which are related to

the scale of the physical processes. Despite interesting work using, for example, concepts associated with cellular automaton (e.g., Palmer 2001), this issue remains to be fully addressed.

Stochastic parameterization of sub-grid scale physical processes is a promising area of research. Such an approach, however, addresses only a part of all model related errors. The stochastic parameterization schemes themselves have their own limitations due to their assumptions and choices, leaving some errors related to the parameterized processes unaccounted for and unrepresented. There also exist grid-scale model errors due to other components of model formulation, for example, temporal and spatial truncation and the design of computational algorithms (Vannitsem and Toth, 2002). For example, Teixeira et al. (2007) investigated the time step sensitivity of NWP models and suggested that truncation error can be a substantial component of total forecast error.

Assessing all sources of model errors including the representation (or lack) of subgrid-scale processes in all model components, and representing them in NWP models will require a sustained effort by the modeling community. Until, and even after this work is completed, there will be a need within ensemble forecasting applications to capture the remaining, residual effects of model related uncertainties. Inspired by this consideration and the pioneering work of Toth and Kalnay (1995), a stochastic representation of total random model error is proposed in this study by imposing stochastic terms on the total tendency of the model equations. The hypotheses behind the proposed scheme are that (a) the deviation of the ensemble members from the control run in their tendencies, or the tendencies of ensemble perturbations in a well-behaved ensemble system with proper initial perturbations, can effectively provide a representative sample of random errors associated with the imperfections of the numerical model used to propagate the ensemble model states, and that (b) model errors will have a significant effect when and where

they project onto existing or developing instabilities in the predicted state of the system. When randomly combined, these tendencies of ensemble perturbations can be used as stochastic forcing terms added to the model tendencies with spatial and temporal correlations consistent with the model dynamics without the need for specifying temporal and spatial correlations. The scheme was tested with the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS), which uses the NCEP Global Forecast System (GFS) as its NWP model.

Full details of the scheme are described in Section 2. The sensitivity of the ensemble forecast to the size of the perturbations and a simple rescaling algorithm are presented in Section 3. The impact of stochastic forcing on the ensemble spread, ensemble mean forecast and ensemble-based probabilistic forecast are discussed in Section 4. Finally, conclusions are summarized in section 5 with further discussions of some related issues.

2.Formulation and implementation of the stochastic perturbation scheme

2.1 Formulation

With subscript i identifying one of the N ensemble members, i=1,2,…,N, (0 is the control GFS forecast) and t the time of the integration, the conventional model equations for an ensemble forecast system running with only initial perturbations can be written as

$$\frac{\partial X_i}{\partial t} = T(X_i, t) \tag{1}$$

where the tendency T is the total tendency or the sum of various terms, including dynamical and physical processes, calculated at grid scale or parameterized for sub-grid scales.

Considering the uncertainty in the model formulation and numerical approximation, a stochastic forcing term S_i should be added to each member, i.e.,

$$\frac{\partial X_i}{\partial t} = T(X_i;t) + S_i(t) \qquad (2)$$

The formulation of the stochastic forcing with the SPP approach generally relates S to the tendency increment due to a particular component of the conventional tendency T, i.e., representing the random error associated with a particular physical process. For example, Buizza et al. (1999) perturbed the parameterized tendency.

We propose to formulate S from the total conventional tendency T. As the perturbations in the initial conditions generated with mature strategies such as Breeding Vectors, Singular Vectors and Ensemble Transform lead to a reasonably (though not perfectly) representative sample of the possible model states, one can assume that the conventional tendencies in the individual ensemble members collectively provide a representative sample of the unknown true value of the total tendency. By comparing the total tendency in each ensemble member against the control forecast, N perturbations in the tendency, i.e.,

$$P_i(t) = T_i(t) - T_0(t) \qquad \text{for i=1,2, …,N} \qquad (3)$$

can be identified and they form a representative sample of the differences between the true value of the tendency and that formulated in the conventional model equation (1). Therefore, these tendency perturbations can be used as the basis in formulating the stochastic forcing S.

As in the SPP approach, random numbers are introduced to address the uncertainty in the total tendency. Although each single P_i, if chosen randomly, can be a valid candidate, a random combination of all N tendency perturbations would be a better choice in hopes that more directions in the phase space would be explored for the ensemble perturbations X_i-X_0 to grow faster and the ensemble spread to be increased. Symbolically, we have

$$S_i(t) \propto \sum_{j=1}^{N} w_{i,j}(t) P_j(t) \qquad \text{for i=1,2, … N} \qquad (4)$$

where the coefficients $w_{i,j}$ are random weights assigned for each P. The stochastic forcing for each ensemble member i corresponds to a different set of random weights $w_{i,j}$, j=1,…N.

2.2 Specification of stochastic weights

Using matrix notation and omitting the time t, the relation (4) can be rewritten as

$$S_{NM} \sim W_{NN} P_{NM} \tag{5}$$

where the subscripts indicate the dimensions of the matrix and M is the number of grid points. Apparently, the formulation of the stochastic forcing is very similar to the Ensemble Transform (ET) used in the generation of initial perturbations, but is applied to the perturbations in the tendency of model state instead of the perturbations in the model state itself.

To determine the combination or weighting matrix W, one needs to consider the requirements for the stochastic forcing S. First, as (2) is to be applied to all model state variables with the same set of weights, the stochastic forcing S should be in approximate balance, as are the tendency perturbations P. Second, the S vectors should be orthogonal to each other. Since the P vectors form an approximately orthogonal set, the othogonality in S can be achieved if the W matrix is orthonormal, i.e., the w vectors are normalized and orthogonal to each other. Therefore, the problem 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 as

$$W_{NN}(t) = W_{NN}(t-1) R_{NN}(t) \tag{6}$$

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.,

$$R_{NN}(t) = R^0{}_{NN} R^1{}_{NN}(t-1) \tag{7}$$

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. The orthonormalization is then realized by applying the Gram-Schmidt 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 A^T is the transpose of A, and α the "degree" 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) for each time that the stochastic perturbation scheme is applied, generate a random slight rotation matrix using the same α_1 but a different seed, and use (6) and (7) 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 scalar weighting factor $w_{i,j}$ is seem as random increments (corresponding to R^1) superimposed on a smooth trend in the form of a periodic function of time (corresponding to R^0) with the level of noise (due to the random increment) and the period controlled by α_1 and α_0, respectively. α_0 and α_1 are the only two parameters required to specify W(t). While a higher value of α_1 defines noisier curves, a larger α_0 corresponds to shorter periods. 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 weighting factors that determine the stochastic forcing for ensemble member 10. In this particular case with $\alpha_0=\alpha_1=0.05$ and 6 hour intervals between applications, it can be seen from fig. 1 that the period of the trend is about 6 days and the curves look fairly noisy. For reference, $\alpha_1=0.005$ defines smoother curves while $\alpha_0=0.005$ corresponds to a much longer period (>10 days).

2.3 A finite difference form and its implementation

In principle, the stochastic perturbation scheme can be applied every time step of the model integration. However, a less frequent application is preferable for the sake of reduced computational cost. For this purpose, a finite difference version of Equation (2) is employed in this study. With a specified time interval of application designated as Δt, the stochastic scheme can be implemented by integrating Equation (1) instead of (2) from t-Δt to t, and the modifying the model state variables (X) by using

$$X_i^{'} = X_i + \gamma(t)\sum_{j=1}^{N} w_{i,j}(t)\left\{\left[\left(X_j\right)_t - \left(X_j\right)_{t-\Delta t}\right] - \left[\left(X_0\right)_t - \left(X_0\right)_{t-\Delta t}\right]\right\} \tag{8}$$

for i=1,2, …, N at t=Δt, t=2$\Delta\tau$, … . γ(t) is a scaling factor that varies with time but is uniform across all ensemble members. Its values depend on the choice of time interval Δt and its temporal variation is related to that of the size of the ensemble perturbations. As shown in Section 3, the function will be empirically determined for a fixed Δt.

The quantity defined by the summation in Eq. (8) is referred as a stochastic perturbation (SP) applied to the *i*-th ensemble member and it can be rewritten as

$$SP_i = \sum_{j=1}^{N} w_{i,j}(t)\left\{\left[\left(X_j\right)_t - \left(X_0\right)_t\right] - \left[\left(X_j\right)_{t-\Delta t} - \left(X_0\right)_{t-\Delta t}\right]\right\} \quad \text{for i=1,2, …, N.} \qquad (9)$$

As the weighting matrix W is orthonormal, the size of each SP is determined by the changes of all ensemble perturbations during the past Δt time interval (the quantity in the curly brackets).

The SPs generated by (9) are for all prognostic variables of the model state and they are in approximate balance. As shown as an example in Fig.2, an SP has structures similar to random noise and its size, as 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.

As the stochastic presentation of model errors defined by (1) and (8) adds stochastic forcing by perturbing the total tendency, it is hereafter referred as Stochastic Total Tendency Perturbation (STTP) scheme. Its implementation is straightforward with periodically stopping the N otherwise independent integrations (1), modifying each model state with information from all N model states using (8) and repeating the procedure every Δt hours. This requires all N sets of model states at time level t and t-Δt to be available simultaneously, and can be easily realized if the N ensemble members concurrently run within a single executable. At NCEP, this was made possible by the recent adaption of the Earth System Modeling Framework (ESMF, see e.g., Collins et al. 2005 and Zhou et al. 2007) as the standard programming and operational environment. ESMF provides tools for turning model code into gridded components with standard interfaces and standard drivers with a robust functionality to run them concurrently. Under ESMF environment, all GEFS ensemble members run concurrently with a separate control packages for each member. A special ESMF code module, known as a coupler, was designed to allow information sharing and exchange among all ensemble members and the

application of Eq. (8). With Δt=6hrs, the CPU time consumed by the integration increases only slightly, by less than 5%.

2.4 Experimental Set-up

At NCEP the Operational GEFS has been providing forecasts for up to 16 days of lead time for many years. Major changes in the configuration happened on May 20, 2006, when the resolution of the GFS model was increased to T126L28 throughout the integration, an ensemble transform (ET) technique (Wei et al., 2007) replaced the breeding method (Toth and Kalnay, 1993, 1997) in providing initial perturbations, and the number of perturbed ensemble members was increased from 10 to 14. The ensemble size was increased again, to 20, on March 27, 2007.

In developing and testing the STTP scheme, extensive experiments have been run for various periods for different seasons of 2006, 2007 and 2008, with each test periods consisting of a number of consecutive days but only the daily 0000 UTC cycle of the GEFS forecast being made. In Sections 3 and 4, some results of these experiments will be presented to demonstrate the sensitivity of model integrations to the stochastic forcing, the tuning process of the STTP and its impact on ensemble forecasts. For each case, ensemble forecasts are run with initial perturbations only (IC ensembles) and with both stochastic perturbations and initial perturbations (ICSP ensembles). Both the IC and the ICSP ensembles start with the same initial conditions to simplify the comparison and isolate the effects of the stochastic perturbations. The ensemble configuration will be the same as that of real-time operations, except for increased resolution in some experiments to reflect a future implementation. The initial perturbations will be the same as those used in operations except for changed model resolutions in the higher resolution cases, where the fields for the operational initial conditions are converted to the required resolution.

3. Scaling the stochastic perturbations and fine tuning of the scheme

Preliminary tests and tuning of the STTP were conducted with T126 resolution. In this section, the parameter setting and tuning process is described by presenting experiments for the periods of Aug. 20 to 30, 2006 and Jan. 3-13, 2007, representing two different seasons. The primary focus is on the error (RMSE) of the ensemble mean forecast and the RMS spread of the ensemble for the geopotential height at 500hPa (H500) and temperature at 850hPa (T850), averaged over the Northern and Southern Hemisphere extratropics (NH, 20^o-80^oN; SH, 20^o-80^oS) and the Tropics (TR: 20^oN-20^oS).

3.1 Preliminary parameter settings

In this section, results from a set of sensitivity tests will be presented to investigate the sensitivity of the forecast to the size of the stochastic forcing terms and parameter tuning in the proposed STTP scheme. From Eqs. (6), (7) and (8), it can be seen that the adjustable parameters include constants Δt, α_0, and α_1, as well as rescaling factor $\gamma(t)$ as a function of forecast lead time. To select the time interval between successive applications of (8), some tests are carried out for Δt = 3 and 6 hours. Both settings led to stable responses and their differences in the results are insignificant if the proper values of scaling factor $\gamma(t)$ are specified. Therefore, Δt=6h is used both in this paper and in the NCEP operational implementation, for considerations of saving computational resources. As the next step, the two parameters used to generate the combination coefficient matrix W, α_0 and α_1, are both set to be 0.05 after preliminary tests. The corresponding period in the evolution of individual coefficients is 6 days, close to the typical

temporal scale of synoptic variability in the mid-latitudes, and the level of stochastic noise leads to quasi-optimal responses in the experimental tests.

The scaling factor γ is given primary attention and is discussed in the following sub-sections. This parameter controls the size of the stochastic forcing added to the model states and the forecast is very sensitive to its magnitude and temporal variation. More importantly, with each application of STTP adds perturbations to the model state in each integration, some imbalance will be inevitably introduced into the model states. Therefore, it is necessary to control the size of stochastic perturbations to avoid model shocks and keep continuity in the model output. As ensemble perturbations in the model states and their tendencies grow with time until saturation is reached at about 2 weeks lead time, one can expect the scaling factor to decrease with lead time. Although γ was mentioned as a function of time only, spatial variation can be easily introduced to maximize the positive impact of the proposed scheme.

3.2 A benchmark global scaling

Global scaling, with γ as a function of time only, is a natural choice to start the tuning process of the proposed scheme. The first set of tests compares the IC ensemble and an ICSP ensemble with a temporally invariant $\gamma(t)=0.1$ (ICSP_01). The most striking impact of the stochastic perturbations on the ensemble forecast is on the mean error (ME) and spread. Fig. 3 shows the ME and RMSE of the ensemble mean forecast and the RMS spread of the ensemble for H500, averaged over the Northern and Southern Hemisphere extratropics (20-80 degrees) and the Aug. 2006 period. The IC ensemble, as discussed in earlier sections, is characterized by significant negative mean error increasing with lead time. The stochastic forcing significantly reduces the negative bias, especially over the Southern Hemisphere where the mean error is close to 0 from

day 1 through day 15 with the ICSP_01 ensemble. In this test, the spread increases faster from the beginning of the integration and the difference becomes clear at day 2. While this is partly the motivation to apply STTP, the overgrowth of spread starting from about day 7, apparently resulted from the model imbalance accumulated by the repeated application of STTP during the integration, is not desirable. The spread exceeds RMSE at about day 10 and the anomaly correlation (AC) coefficient is significantly degraded in the week 2 forecast, especially in the tropics (not shown). Apparently, the scaling factor γ should be gradually reduced after about day 5 (t=120h). A piece-wise linear function

$$\gamma(t) = \gamma_0(t) = \pm\{0.1 - \frac{0.1 - 0.01}{384 - 120} \bullet Max(0, t - 120)\} \qquad (10)$$

is used to run another ICSP ensemble (ICSP_PW) and it restored the spread and AC to an acceptable level in both extratropics (Fig. 3). Therefore, Eq. (10) is used as our benchmark for global scaling.

3.3 Fine tuning to improve tropical forecasts

The result of the ICSP_PW experiment is not satisfactory in the tropics. As shown in Fig. 4, the anomaly correlation coefficient after day 9 is about 10 percentage points lower than that in the IC ensemble, and the spread is still larger than the RMSE of the ensemble mean forecast. Although height in the tropics is not a good variable to represent the quality of the forecast, this degradation suggests that fine tuning is necessary. In order to tune the scheme a number of experiments are performed with parameter α_1 altered and some additional procedures added. When α_1 is reduced from 0.05 to 0.005, the anomaly correlation recovers to the level of the IC ensemble but the spread is further increased by about 50% (not shown). On the other hand, centralizing the scaled SPs (i.e., subtracting their mean from each individual vector) significantly

reduces spread and restores the AC, and simply changing their signs (by taking the negative sign in (10)) further slightly improves AC in both the tropics and extratropics (not shown). This configuration (ICSP_PW, plus the negative sign in (10) and centralization), is denoted as ensemble ICSP_PW_NC, and the results are shown as the short dashed curves in Fig.4.

The results in Sections 3.1-3.3 show that STTP defined by (1) and (8) can significantly change the ensemble forecast. Although too large amplitude of SP leads to a degraded association between the forecast and analysis, properly combining the parameters α_0, α_1, Δt and the scaling function $\gamma(t)$, and applying some special procedure can result in positive impacts such as reduced negative bias and increased ensemble spread. This can be achieved in the extratropics by using a piece-wise linear function (10) for the global scaling factor. Additional procedures, such as changing the sign of the scaled SPs and applying a centralization procedure to them, are necessary for STTP to work properly in the tropics.

3.4 Seasonal and meridional variations of the scaling factor

The global scaling described earlier has the advantage of conserving the original structure of the SPs calculated from (9), but is unable to control the size of the added stochastic tendency on the regional and local scales. As mentioned earlier, adding stochastic perturbations has the effect of reducing the negative bias (regional mean error) and increasing the spread. It would be ideal to control the global scaling function $\gamma(t)$ so that the bias is close to zero and the spread matches the RMSE of the ensemble mean forecast. However, numerous experiments show that STTP reduces negative bias or increases positive bias in almost all cases, and large values of γ lead to a positive bias. This can be seen from Fig. 5 by comparing the IC ensemble with the ICSP_PW_NC, averaged over both extratropics for both the Aug. 2006 and Jan. 2007 periods.

For the IC ensemble, the negative bias is larger in the winter hemisphere than that in the summer hemisphere. This difference is clearer in the Southern Hemisphere where the bias is very small in the summer even without STTP (the IC ensemble). With the γ function specified in (10), i.e., the ICPS_PW_NC ensemble, there is still room to increase the size of the perturbations in the winter hemisphere, both in terms of bias and spread. However, for the SH summer season, the bias is already positive and a further increase in the perturbation size will lead to increased positive bias. Therefore, a better solution is to vary the γ parameter with latitude and season.

Fig. 6 shows the temporal variation in the regional mean error of the 9-day forecast of H500 averaged over the two extratropics for the calendar year 2007, from GEFS operational production (STTP was not implemented). Although the daily variation is noticeable, the dominant mode is the seasonal variation in both hemispheres and they are approximately out of phase, with the peak/valley on about July 1st and January 1st. Based on this observation, the following function is used to specify the γ function

$$\gamma = \gamma_1(\varphi, d)\gamma_0(t)$$
$$\gamma_1(\varphi, d) = 1.0 + 0.2\sin(\varphi)\cos\frac{2\pi d}{364} \qquad (11)$$

where φ is the latitude and d is the Julian day of the year. The result from the experiment using (11), labeled as ICSP_PW_NC_V, is shown in Fig. 5 as the short dashed curves, indicating the slightly reduced impact of SP in the summer and a noticeable increase in the winter.

3.5 The current setting for operational application

The configuration and parameter settings of the ICSP_PW_NC_V ensemble, as the optimal choice resulting from the controlled tuning process, is used both for the rest of this article and for the operational implementation with GEFS at NCEP in Feb. 2010. A more comprehensive tuning

process of STTP requires a through understanding of the structures, evolution and statistical characteristics of stochastic perturbations. Specifically, an adaptive algorithm could be developed to vary the scaling factor γ in 3-dimensional space and objectively specify its values at each grid point.

The specification of $\gamma_0(t)$ in (10) is empirical and more or less arbitrary. Experiments suggest that it is sub-optimal and fine tuning may be necessary in the future, especially when the ensemble setup and model configurations are changed. For this reason, it is replaced with a differentiable function controlled by a number of tunable parameters. A natural choice is one defined using a logistic function

$$\gamma_0(t) = \pm[p_2 + (p_1 - p_2)\left\{1.0 - \frac{1.0}{1.0 + e^{-p_3(t-p_4)}}\right\}] \qquad (12)$$

With parameters p_1=0.100, p_2=0.01, p_3=0.11 and p_4=252 hours in the current setting, the curve is very close to the piece wise linear function (10).

As a summary, for the discussions in sections 4 and 5 of this article and the upgrade of NCEP GEFS in 2010, Equations (1), (6), (7), (8), (9), (11) and (12) are used to implement the stochastic perturbation scheme under the ESMF environment. For further clarity, the parameters used are listed as

(1) $\alpha_0=\alpha_1$=0.05 for R^0 and R^1 in (7);

(2) Δt = 6 hours in (8) and (9);

(3) γ is defined by (11) and (12);

(4) in (12) negative sign is used with p_1=0.100, p_2=0.01, p_3=0.11 and p_4=252 hours;

(5) Centralization (i.e. removing the mean) is applied to the scaled SPs before modifying model states with (8).

4. The impact of SP on ensemble forecasts

In this section, the impact of the STTP scheme is presented in a broader sense, by including the ensemble mean forecast and the ensemble based probabilistic forecast over extended periods based on the most recent GEFS configurations. Emphasis is on the effect of the stochastic forcing on the general quality of the forecast and the improvement over the same GEFS configuration without stochastic forcing. For this purpose, an ensemble is generated with the configuration and parameter settings for the 2010 implementation as described in Section 3.5, and is simply referred to as ICSP. Both the IC and ICSP ensembles are run with T190L28 resolution for the periods of Aug. 20 –Sep. 30, 2006 and Jan. 3-29, 2007, representing two different seasons.

4.1 Outliers

As seen from Section 3, the stochastic parameterization scheme effectively increases the spread of the ensemble forecast. However, whether this increase in spread improves the quality of the ensemble forecast is still a question. To improve the ensemble forecast, one has to reduce the number of outliers, or the grid points where the verifying analysis falls outside the range of the ensemble member forecasts.

Outlier analysis has been used at NCEP (Zhu et al. 1996; Toth et al. 1998) to assess the performance of ensemble forecasts for years. It is based on the analysis rank or Talagrand diagram (Anderson 1996; Talagrand et al. 1998; Toth et al. 2003), which is constructed by accumulating the number of cases over the space and time where the verifying analysis falls within each of the $N+1$ intervals defined by an ordered series of N ensemble members at each grid point. The cases where the verifying analysis falls within one of the two extreme intervals

outside the range of the ensemble are called outliers. The number of outliers should be *2/(N+1)* for a reliable and statistically consistent ensemble forecast. The number of excessive outliers can be obtained by subtracting this expected value from the number of actual outlier forecasts, and is more conveniently expressed as a percentage when divided by the total number of all cases. The resulting Percentage of Excessive Outliers (PEO) is related to the quality of the ensemble forecast, with a positive (negative) value indicating too small (large) a range and/or a strong bias.

The PEO of the T850 forecast averaged over the Northern and Southern Hemispheres and the Tropics for various lead times are shown in Fig. 7. At short lead times, the PEO is high in both the IC and ICSP ensembles. However, during the first week it is reduced more rapidly in the ICSP ensemble than the IC. For a forecast with a lead time longer than 10 days, the PEO in the ICSP is about half of that in the IC for all three regions and both seasons. The reduction in PEO due to STTP is more effective in the winter season than in the summer-fall season, consistent with the variation of the scaling factor defined in (11). In the extratropics, the PEO in ICSP is only 3% or less, with the northern summer as an exception, suggesting that the scaling parameter γ is close to its optimal value and its 3 dimensional spatial variation is necessary for further improvement.

The reduction in the number of outliers associated with an increase in ensemble spread indicates that the STTP scheme has a positive impact on the performance of the ensemble forecast.

4.2 Error of the Ensemble Mean Forecast

It has been noted in Section 3 that the stochastic perturbation scheme tends to reduce negative bias in the ensemble mean forecast. The bias is defined as the difference (error) between forecast

and analysis averaged over a large domain (e.g., NH, SH and TR) and then over a specified period, i.e.,

$$ME = \overline{\langle f - a \rangle} = \langle \overline{(f - a)} \rangle \tag{13}$$

where the angle brackets and the overbar represent the temporal and areal averages, respectively.

In evaluating and calibrating NWP products, systematic error and random error are often treated separately. For example, systematic error is removed by a statistical post-processing procedure at each grid point. In this context, the temporal mean of the error (f-a) is viewed as Systematic Error (SE), i.e.,

$$SE = \langle f - a \rangle \tag{14}$$

and the bias (ME), defined by (13) and shown in Section 3, is the (areal) Mean of the Systematic Error. As the calculation of ME involves cancellations between positive and negative values in the areal average, it is necessary to use a different measure to confirm the impact of the STTP scheme. For this purpose, an error measure, namely the Mean Absolute Systematic Error (MASE) is defined as

$$MASE = \overline{|\langle (f - a) \rangle|} \tag{15}$$

where the areal average (overbar) is applied to the absolute value of the point-wise value of SE. For this reason, MASE is a more meaningful error measure in comparing the two ensembles.

In Figs. 8 a and b, ME and MASE over the NH, SH and TR are plotted for the ensemble mean forecast of the IC (thin line) and ICSP (thick line) ensembles. The reduction in ME and MASE due to stochastic perturbations is seen in all 3 domains, both seasons and for both H500 (Fig. 8a) and T850 (Fig. 8b). For lead times longer than 2 days, the ME reduction is significant. For the 15 day forecast, the ME reduction in the extratropics (one quarter to one third) is more effective than in the tropics (about 10%). For MASE the reduction is smaller, especially in the

T850 forecast over the tropics. Note that the STTP scheme increases ME if it is positive. When this happens, the reduction in MASE may be also very small or negative. One example of this situation is the T850 forecast for lead times of 2-8 days over the tropics in the Jan. 2007 period (Fig. 8b).

To better understand the impact of the STTP scheme on SE, Figs. 9 shows the global distribution of SE for the IC ensemble and its change due to the inclusion of STTP (i.e., the ICSP ensemble) for the H500 and forecast over the August-September period of 2006. In the IC ensemble, SE displays large scale patterns with negative and positive centers of comparable amplitudes, but the negative is slightly larger. Over the tropics, SE is dominantly negative. This explains the negative bias (ME) when averaged over the NH, SH and TR domains. The SE increment due to STTP (lower panel) is dominantly positive, especially in the tropics where no significant negative values appear, leading to the reduction in negative ME for most domains/periods. In addition, there is some negative correlation between the two quantities and the SE increment has a smaller variation than the SE field in the IC ensemble, suggesting that STTP tends to correct the bias in the right direction. The relatively rare cases in which STTP increases positive SE (higher bias) may lead to a negative impact. Nevertheless, the post-processing procedure employed at NCEP to remove the systematic error (Cui et al. 2005) can largely remove the forecast bias and minimizes this negative impact.

4.3 Probabilistic Forecast

A review of the commonly used probabilistic forecast verification measures can be found, e.g., in Toth et al. (2003). In this study, the area under the *Relative Operating Characteristic (ROC)* curves, *the Economic Value (EV), Brier Skill Score* (BSS) with the *reliability* and *resolution*

components of Brier Score, and the *Ranked Probability Skill Score (RPSS)* are calculated with the standard NCEP ensemble verification package (Zhu et al. 1996, Toth et al. 1997, Zhu et al. 2002). With this package, these scores are calculated for probabilistic forecasts of events defined by partitioning the range of possible forecast values into a finite number of complete yet exclusive intervals (bins/classes) (Toth et al. 2003). Again, the verification is focused on H500 and T850, both continuous variables. For these two variables, 10 climatologically equally likely bins were determined by Zhu et al. (1996), using the climatological database from the NCEP-NCAR reanalysis (Kalnay et al. 1996).

Fig. 10 shows BSS and the two components of BS for the H500 forecast of the IC and ICSP ensembles averaged over the NH. Note that an improved forecast will have a higher BSS, which in turn requires a lower reliability component and/or a higher resolution component of Brier Score. It can be seen that STTP has positive impacts over the extratropics in both warm and cold seasons with a significant increase in BSS appearing at day 3. In addition, the improvement in BSS is mainly from the decrease in the reliability component of BS. This is also true over the SH (not shown) and for T850 (Fig. 11). In the tropics, the BSS improvement is more consistent and has contributions from both the reliability and resolution components of BS, as shown for T850 in the right panel of Fig. 11.

The variation of the ROC area score with forecast lead time is shown in Fig. 12 for the IC and ICSP ensembles, averaged over the NH, SH and TR. Generally, ICSP has higher scores than IC for all three domains. However, the difference is much larger in the tropics. For the extratropics, the positive impact is stronger in the cold hemisphere. The same observation is also true for RPSS (not shown).

For an ensemble forecast of a continuous variable S, such as geopotential height or temperature, the cumulative distribution function of the predicted quantity can be estimated and denoted by the probability F(s)=p(S<s), where s is any possible value of S. Following Toth et al. (2003), the Continuous Ranked Probability Score (CRPS) can be calculated as

$$CRPS = E\left(\int_{-\infty}^{\infty} [F(s) - H(s-a)]^2 \, ds \right)$$

where H(s-a) is the Heaviside function that takes the value 0 when s-a <0 and is 1 otherwise. It also can be expressed as a skill score, Continuous Ranked Probability Skill Score (CRPSS)

$$CRPSS = 1 - \frac{CRPS}{CRPS_{ref}}$$

In this formula $CRPS_{ref}$ is the CRPS of a reference forecast, which is used as a benchmark for comparison. The reference forecast used in this paper is the climatological forecast.

Fig. 13 depicts the CRPSS for H500 and T850 for the IC and ICSP ensembles averaged over the NH, SH and TR for the northern warm season. Similar to what is seen for BSS and RPSS, the impact of SP on CRPSS is the most significant in the tropics, bringing the useful lead time (CRPSS positive) of the H500 forecast from 7 days to 10 days in the northern summer season. Comparing the two extratropics, the positive impact of STTP is more significant in the SH, with the 5 day forecast skill improved by about 12 hours for both variables. For the NH, the forecast improvement in H500 is clearer than that in T850. For the northern cold season (not shown), forecast improvement over NH is more clearly seen. Again, for the extratropics the positive impact of STTP is stronger in the cold season.

4.4 Impact of STTP and model improvement

The impact of STTP discussed in this section is accessed for the specific configuration of the GEFS and GFS models used for the integration. Apparently, the characteristics of the forecast are strongly dependent on these configurations and this raises the question whether STTP can still be beneficial to the ensemble forecast in the future, with increased resolution and an improved NWP model, and in a broader sense, with other EPSs using different IC generation techniques and NWP models. During the development of the current scheme, both T62L28 and T126L28 were used and the results were very similar to what is presented in this paper at the T190L28 resolution, but the question still needs to be answered with direct comparisons and controlled experiments. For this purpose, two sets of experiments were designed and labeled by changing the horizontal resolution (T126 or T190), selecting the formulation of horizontal diffusion (D4 or D8) and switching STTP on and off (S or null). All these ensembles were run at a fixed vertical resolution of 28 levels as in the operational GEFS configuration. The selection of the horizontal diffusion as an example of NWP model improvement is based on the following observation: The 2007/2008 operational GFS used a fourth order (D4) formulation for T126 and 8^{th} order (D8) for T190 and higher resolutions, and a recent study shows that both single and ensemble forecasts with T126 can be significantly improved if D8 is used (The GFS implemented on Dec 15, 2009 in fact uses D8 for all supported resolutions).

The first set of experiments is made up of 4 ensembles with the labels T126D4 (crosses), T190D8 (open circles), D126D4S (closed circles), and T190D8S (open squares). For simplicity, only the RPSS is plotted in Fig. 14 for the T850 forecasts averaged over the tropics for the period Nov. 1-23, 2007. T190D8 has a higher skill score than T126D4 for all lead times, with the lead time of the 0.5 skill level increased from 2 days to 5 days. This indicates that increasing the horizontal resolution and replacing the 4^{th} order formulation of horizontal diffusion with 8^{th} order

are significant contributors to the forecast improvement. Both runs are further improved by employing the STTP, with the time of a critical RPSS values (0.5 for T190D8 and 0.4 for t126D4) extending from day 5 to about day 7.

The second set of experiments is focused on an ensemble configuration with varying horizontal resolutions in which the model is truncated from T190 to T126 at 180 hours. The labels of these ensembles thus consist of two parts representing the two stages of the model integration, separated by a hyphen. The RPSS scores are shown in Fig. 15 for four experiments. The only difference between T190D8-T126D4 (crosses) and T190D8S-T126D4S (open circles) is that STTP is switched on in the latter. The positive impact of the stochastic perturbations is very similar to what was shown in the previous paragraph. In both cases, the skill has an abrupt drop immediately after the truncation. The other pair of experiments (closed circles and open squares) is identical to the first pair except that the horizontal diffusion in the later stage is 8^{th} order. The abrupt drop of skill is avoided by using an improved model (higher order horizontal diffusion scheme) but the impact of STTP, seen as the difference between the two experiments in the pair, is basically the same as between the D4 cases.

The same observations are made for the number of outliers, ensemble spread, bias of the ensemble mean and most of the probabilistic forecast verification scores such as BSS, ROC and CRPSS. These results suggest that the positive impact of STTP is independent of the particular model version and that it will still be beneficial when the NWP model used for integration is improved and very likely when the scheme is used with other NWP models.

5. Conclusions and Further Discussion

This article describes a new approach for representing random errors associated with the NWP model, and presents a stochastic scheme developed at the National Centers for Environmental Prediction and implemented with the Global Ensemble Forecast System (GEFS). Unlike the stochastic parameterization scheme employed in the European Center for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF EPS) and other stochastic physics schemes, the new scheme is not attached to any specific physical process or any specific component of the tendency. Instead, it tries to sample the random errors associated with the total tendency, including both dynamical and physical processes, grid resolved and parameterized components. For this reason, the scheme is referred as a Stochastic Total Tendency Perturbation (STTP) Scheme.

The new approach proposed in this article is based on the hypothesis that the deviation of the ensemble members in terms of their tendencies, or the tendencies of ensemble perturbations, in a reasonably well-behaved ensemble system can effectively provide a sample of the random errors associated with the imperfections of the numerical model used to propagate the ensemble model states. These tendencies of ensemble perturbations, when randomly transformed in the ensemble space (i.e., combined following certain rules to keep their orthogonality) and scaled to an appropriate size, can be used as the stochastic component of the tendency added to the model equation.

The particular version of STTP scheme, described in this article and implemented in the NCEP Global Ensemble Forecast System (GEFS), employs a finite difference form in estimating the ensemble tendencies. All of the ensemble member integrations are stopped every 6 hours and each model state is compared with its value 6 hour earlier to obtain the perturbation tendency for

the ensemble member considered. These perturbation tendencies are then stochastically combined (in a different manner) to generate the required stochastic perturbations.

For an N-member ensemble system, the random combination coefficients are generated in such a way that the NxN matrix of the coefficients in a specific application of the scheme is orthonormal. By doing this, the random combination is very similar to Ensemble Transform but is applied to the perturbation tendencies (e.g., Wei et al., 2007). The variation of the combination matrix with forecast lead-time is to simulate successive rotations of the N vectors in N-dimensional space, with random rotation superimposed on a steady rotation. For a particular element of the matrix, its temporal variation is a periodic function superimposed by random increments, with the period and level of noise controlled by independent parameters.

The impact of the stochastic perturbations on the GEFS is sensitive to their size and they need to be scaled. To maximize the benefits of STTP, the scaling factors should vary in both space and time. In the current implementation, the quasi-optimal rescaling factor is prescribed as a function of lead-time, latitude and season and extensive experiments were performed to tune the parameters.

Extensive experiments were performed for several extended periods in different seasons, each lasting 3 to 6 weeks, using the most recent operational version of the GFS model with different horizontal resolutions. Results from these experiments can be summarized as the following conclusions:

(1) The inclusion in the model equations of stochastic tendency, represented by the rescaled stochastic perturbations to the total tendency, increases the ensemble spread and properly controlling the perturbation size leads to reduced number of outliers;

(2) The application of the Stochastic Total Tendency Perturbation (STTP) scheme significantly reduces systematic error of the ensemble mean forecast;

(3) With the STTP Scheme, the ensemble based probabilistic forecast can be significantly improved. In particular, the forecast distribution is improved, as shown by increased CRPSS scores;

(4) The forecast improvement, especially in terms of the probabilistic forecast verification scores, is more prominent in the tropics than the extratropics. The improvement in the extratropics is more prominent in the cold season;

(5) The improvement in probabilistic forecast scores over the extratropics is mainly from the improved statistical reliability, while improved statistical resolution is equally important in the tropics;

(6) The positive impact of the stochastic perturbations is not sensitive to increases in horizontal resolution and changes in the formulation of the horizontal diffusion. This suggests that the scheme can still be beneficial in the future with improved NWP model performance.

These impacts of STTP are consistent with the general expectations from any realistic stochastic scheme. As stated in Palmer (2001), the nature of stochastic parameterization schemes predicts that they have the following potential benefits: (1) more complete representation of model uncertainty; (2) reduction in systematic error, due to noise-induced drift; and (3) more accurate estimates of internal climate variability. As these stochastic schemes describe the collective effects of unresolved scales on the resolvable scale but not the exact evolution of such motions, they are more likely to improve forecasts in terms of statistical reliability than statistical

resolution. This consistency suggests that the Stochastic Total Tendency Perturbation scheme correctly sampled some uncertainty associated with the NWP model used in the integration.

The implementation of the Stochastic Total Tendency Perturbation scheme can be improved in a number of different ways. Although the time interval between adjacent applications of the scheme is set to be 6 hours, a different value of 3 hours was tested and the difference in the results is not very large. To use the scheme in data assimilation, it would be necessary to test whether a shorter time interval, e.g., 1 hour or less, can further improve the forecast. The rescaling factors are currently specified empirically and subjectively while an adaptive and objective algorithm is a better choice. Finally, the impact of the scheme on real "weather" variables or near surface quantities, such as 2m temperature, 10m winds and precipitation, is not presented in this article due to the limit of space and the relatively small sample sizes available to this study. However, some of these variables were monitored during the development and testing of the scheme to make sure no serious degradation was tolerated. In Feb. 2010, after stringent user evaluation, the scheme was implemented in the GEFS routine operational forecast at NCEP. The evaluation of these "weather" variables will be performed and reported separately with the large operational data set that will be available.

Acknowledgements: The authors thank Mark Iredell, Henry Juang, Stephane Vannitsem and Stephen Lord for advices on the methodology of this research, and Jim Purser for providing the software used in generating the stochastic combination coefficient Matrix. Thanks are also due to Cecile Penland and Prashant Sardeshmukh for discussions about the results of earlier versions of the Stochastic Total Tendency Perturbation scheme, and Shrinivas Moorthi, Joe Sela and Mike Young for their contributions to its operational implementation under ESMF environment. In the

process of EMC/NCEP internal review, Hua-lu Pan and Shrinivas Moorthi provided valuable comments and suggestions for improving the manuscripts and their help is appreciated. The help of Mozheng Wei and Bo Cui with computation and data processing, and Mary Hart with improving the English of the manuscript are acknowledged. The first author was supported by NOAA THORPEX program and NOAA Climate and Global Change Program under Project GC02-228.

REFERENCES

Berner, J., G. J. Shutts, M. Leutbecher and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. J. Atmos. Sci., **66**, 603-625.

Bony, S., and K. A. Emanuel, 2001: A parameterization of the cloudiness associated with cumulus convection; evaluation using TOGA COARE data. *J. Atmos. Sci.,* **58,** 3158-3183.

Buizza, R., 1997: Potential forecast skill of ensemble prediction, and spread and skill distribution of ECMWF Ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99-119.

Buizza, R., M. J. Miller and T. Palmer, 1999: Stochastic simulation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. Meteorol. Soc.,* **125,** 2887-2908.

Buizza, R. and T. Palmer, 1995: The singular vector structure of atmospheric general circulation. *J. Atmos. Sci.,* **52,** 1434-1456.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076-1097.

Cui, B., Z. Toth, Y. Zhu, D. Hou and R. Wobus, 2005: Bias correction methods ----- adjusting moments. *Preprints, European Geoscience Union General assembly 2005*. Vienna, Austria, European Geoscience Union, EGU05-A-05997.

Collins, N., G. Theurich, C. DeLuca, M. Suarez, A. Trayanov, V. Balaji, P. Li, W. Yang, C. Hill and A. da Silva, 2005: Design and Implementation of Components in Earth System Modeling Framework, The International Journal of High Performance Computing Applications, Volume 19, No. 3, Summer 2005, pp. 341-350.

Du, J. and M. S. Tracton, 2001: Implementation of real-time short range ensemble forecasting system at NCEP: an update. Preprints, *9th Conference on Mesoscale processes,* Ft Lauderdale, Florida, Amer. Meteor. Soc., 355-356.

Frederiksen, J. S. and A. G. Davies, 1997: Eddy Viscosity and stochastic backscatter parameterization on the sphere for atmospheric Circulation Models. *J. Atmos. Sci.,* **54,** 2475-2492.

Golaz, J.-C., V.E. Larson, and W. R. Cotton, 2002: A PDF-based model for boundary clouds. Part I: Method and model description. *J. Atmos. Sci.,* **59,** 3540-3551.

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

Hou, D., E. Kalnay and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.*, **129,** 73-91.

Hou, D., K. Mitchell, Z. Toth, D. Lohmann and H. Wei, 2009: The effect of large scale atmospheric uncertainty on Streamflow predictability. Journal of Hydrometeorology, 10, 717-733.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H.L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124,** 1225-1242.

Houtekamer, P. L. and H.L. Mitchell, 1996: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126,** 796-811.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G.,Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki,W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C.,Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, B. Am. Meteor. Soc., 77, 437–471, 1996.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Wilford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecast from multi-model superensemble. *Science*, **285**, 1548-1550.

Leith, C. E., 1974: Theoretical skill of Monto-Carlo forecasts. *Mon. Wea. Rev.*, **102,** 409-418.

Lin, J. W.-B., and J. D. Neelin, 2002: Considerations for Stochastic convective parameterization. *J. Atmos. Sci.,* **59,** 959-975.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Q. J. Meteorol. Soc.,* **122,** 73-119.

NMC Development Division, 1988: Documentation of the research version of the NMC Medium range Forecasting Model. NMC Development Division, 504pp. [Available from NCEP/EMC 5200 Auth Road Camp Springs, MD 20746.]

O'kane, T. J., and J. S. Frederiksen, 2008: Statistical dynamical subgrid-scale parameterizations for geophysical flows. *Physica Scripta*, **T132,** 014033.

Palmer, T. N., F. Molteni, R. Mureau and R. Buizza, 1993: "Ensemble prediction'. Pp21-66 in Proceedings of the ECMWF seminar on validationof models over Europe: Vol. 1. ECMWF, Shinfield Park, reading, UK.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Q. J. Meteorol. Soc.,* **127,** 279-304.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction system. Quart. J. Roy. Meteor. Soc., **131**, 3079-3102.

Son, J., D. Hou and Z. Toth, 2008: An assessment of Bayesian bias estimator for numerical weather prediction. Nonlinear Processes in Geophysics, 15, 1013-1022.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128,** 2077-2107.

Teixeira, J. and T. F. Hogan, 2002: Boundary layer clouds in a global atmospheric model: simple cloud cover parameterizations. *J. Climate*, **15**, 1261-1276.

Teixeira, J., C. A. Reynolds and K. Judd, 2007: Time-step sensitivity of nonlinear atmospheric models: Numerical convergence, truncation error growth, and ensemble design. *J. Atmos. Sci.,* **64,** 175-189.

Teixeira, J. and C. A. Reynolds, 2008: Stochastic Nature of physical parameterizations in Ensemble prediction: A Stochastic convection approach. *Mon. Wea. Rev.*, **136,** 483-495.

Tompkins, A. M., 2002: A prognostic parameterization for the sub-grid scale variability of water vapor and clouds in large scale models and its use to diagnose cloud cover. *J. Atmos. Sci.,* **59,** 1917-1942.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The Generation of perturbations. *Bull Amer. Meteor. Soc.,* **74,** 2317-2330.

Toth, Z. and E. Kalnay, 1995: Ensemble forecasting with imperfect models. WMO CAS/JSC WGNE report on *Research* Activities in Atmospheric and Oceanic Modeling, WMO/TD-No. 665, 6.30. WMO, Geneva, Switzerland.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP: The Breeding Method. *Mon. Wea. Rev.,* **125,** 3297-3318.

Toth, Z., Y. Zhu, T. Marchok, S. Tracton, and E. Kalnay, 1998: Verification of the NCEP global ensemble forecasts. Preprints of the 12th Conference on Numerical Weather Prediction, 11-16 January 1998, Phoenix, Arizona, 286-289.

Toth, Z., and Y. Zhu, 1997: Probabilistic Weather Forecasts Based on the NCEP ensemble, XXII General Assembly of the European Geophysical Society. Vienna, Austria, 21-25 April 1997, *Annales Geophysicae,* Part II, Supplement 11 to Vol.**15**, C558.

Toth, Z., O. Talagrand, G. Candille and Y. Zhu, 2003: Probability and Ensemble Forecast. *Forecast Verification: A practitioner's Guide in Atmospheric Science.* Jolliffe, I. T. and D. B. Stephenson Ed., John Wiley & Sons Ltd.

Vannitsem, S., and Z. Toth, 2002: Short-term dynamics of model errors. *J. Atmos. Sci.,* **59,** 2594-2604.

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

Zhou, S., V. Balaji, C. Cruz, A. da Silva, C. Hill, E. Kluzek, S. Smithline, A. Trayanov and W. Yang (Authorship in alphabet sequence except the first author), 2007: Cross organization interoperability experiments of weather and climate models with the Earth System Modeling Framework, Concurrency Computat. Pract. Exper., 2007, 19, 583-592, Published online 10 October 2006 inWiley InterScience (www.interscience.wiley.com), DOI:10.1002/cpe.1120

Zhu, Y., G. Iyengar, Z. Toth, S. Tracton, and T. Marcock, 1996: Objective evaluation of the NCEP global ensemble forecasting system. *Preprints of the 15th Conference on Weather Analysis and Forecasting,* 19-23 August 1998, Norfolk, Virginia, J79-J82.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson and K. Mylne, 2002: The economic value of ensemble based weather forecasts. *Bull Amer. Met. Soc.,* **83,** 73-83.

Figure Captions

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.

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.

Fig. 3 Mean error (thick curves) of ensemble mean forecast and ensemble spread (thin) of IC (solid/black curves), ICSP_01 (long dashed/red) and ICSP_PW (short dashed/green), along with root mean square error of the IC ensemble (crosses) and the climatological forecast (black squares), for H500 averaged over the Northern (left panel) and Southern (right panel) Hemisphere extratropics.

Fig. 4 Anomaly correlation (left panel), mean error (right panel, thick curves) of ensemble mean forecast and ensemble spread (right panel, thin curves). IC (solid/black), ICSP_PW (long dashed/red) and ICSP_PW_NC (short dashed/green) ensembles are compared for H500 forecast, averaged over the Tropics and the period of Aug. 20-30, 2006. Also plotted in the right panel are the corresponding RMSE of the IC ensemble (crosses) and the climatological forecast (black squares).

Fig. 5 Mean error (thick curves) of ensemble mean forecast and ensemble spread (thin) of IC (solid/black), ICSP_PW_NC (long dashed/red) and ICSP_PW_NC_V (short dashed/green) ensembles, along with RMSE of the IC ensemble (crosses) and the climatological forecast (black squares), for the H500 forecast averaged over the Northern (left panels) and Southern (right panels) Hemisphere extratropics, for the periods of Aug. 20-30, 2006 (upper panels) and Jan. 3-13, 2007 (lower panels).

Fig. 6 Mean error of NCEP operational 9-day ensemble mean forecast of H500 averaged over the Northern (open squares) and Southern (closed square) Hemisphere extratropics, for the calendar year of 2007.

Fig. 7 Percentage of Excessive Outliers over the expected value, for 850 hPa temperature forecast averaged over the Northern (solid/black curve) and Southern (long dashed/red) Hemisphere extratropics and the Tropics (short dashed/green) for the IC (thin) and ICSP (thick) ensembles. The left/right panel is for the northern summer-fall/winter season.

Fig. 8a Mean Systematic Error (MSE, upper panel) and Mean Absolute Systematic Error (MASE, lower panel) of 500 hPa geopotential height averaged over the Northern (solid/black curve) and Southern (long dashed/red) Hemisphere extratropics and the Tropics (short dashed/green), for the IC (thin) and ICSP (thick) ensembles averaged over the northern summer-fall season.

Fig. 8b Same as in Fig.8a except for average over the northern winter season.

Fig. 9 Systematic Error (SE, upper panel) of the 240h forecast of 500 hPa height, for the IC ensemble (upper panel) and the difference between ICSP and IC ensembles (lower panel), averaged over the northern summer-fall season.

Fig. 10 Brier Skill Score (BSS) and the reliability and resolution components of Brier Score (BS), for 500 hPa geopotential height of the IC (light/black curves) and ICSP (dark/red curves), averaged over the Southern Hemisphere extratropics for the northern summer (left panel) and winter (right panel) seasons.

Fig. 11 As in Fig. 10, except for 850 hPa temperature averaged over the Southern Hemisphere extratropics (left panel) and tropics (right panel) for the northern summer season.

Fig. 12 ROC area of 500 hPa geopotential height, averaged over the Northern (solid/black curves) Southern (long dashed/red) Hemisphere extratropics and the tropics (short dashed/green) for the IC (thin) and ICSP (thick) ensembles, for the northern summer (left panel) and winter (right panel) seasons.

Fig. 13 CRPSS of 500 hPa geopotential height (left panel) and 850 hPa temperature, averaged over the Northern (solid/black curves) Southern (long dashed/red) Hemisphere extratropics and the tropics (short dashed/green) for the IC (light/thin) and ICSP (dark/thick) ensembles, for the northern summer season.

Fig. 14 RPSS of 850 hPa temperature, averaged over the northern tropics and the period Nov. 1-23, 2007, for the ensembles T126D4 (crosses), T190D8 (open circles), T126D4S (filled circles) and T190D8S, described in the text.

Fig. 15 RPSS of 850 hPa temperature, averaged over the northern tropics and the period Nov. 1-29, 2007, for the ensembles T190D8-T126D4 (crosses), T190D8S-T126D4S (open circles), T190D8-T126D8 (filled circles) and T190D8S-T126D8S, described in the text.

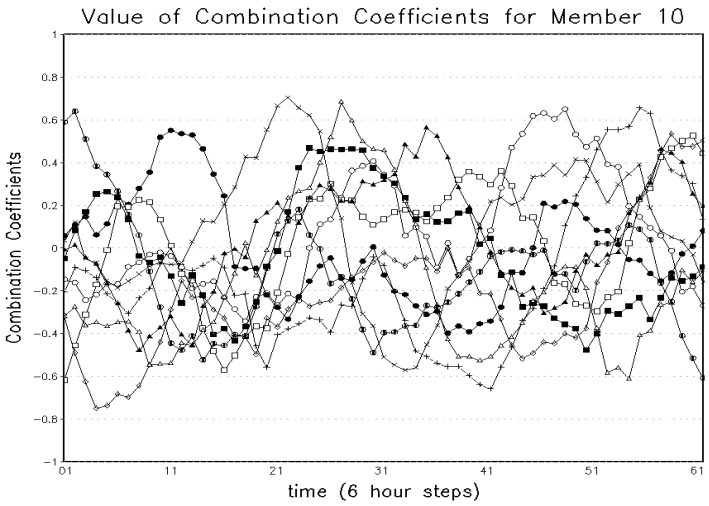

Fig.1 Examples of combination coefficients $w_{i,j}$ as a function of forecast lead time, in a 10 member (N=10) ensemble system. Shown are 10 curves for j=1,2,…,10 with fixed i=10.

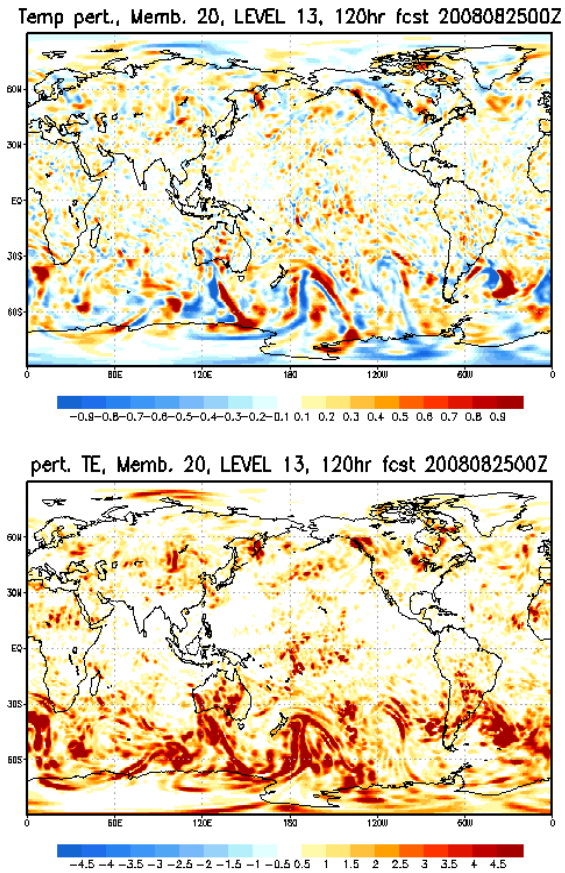

Fig. 2 An example of stochastic perturbations added to modify the model states of the 20-member NCEP Global Ensemble Forecasting System. Shown are (upper panel) the temperature perturbation (unit: K) associated with ensemble number 20, and (lower panel) the corresponding perturbation size defined as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

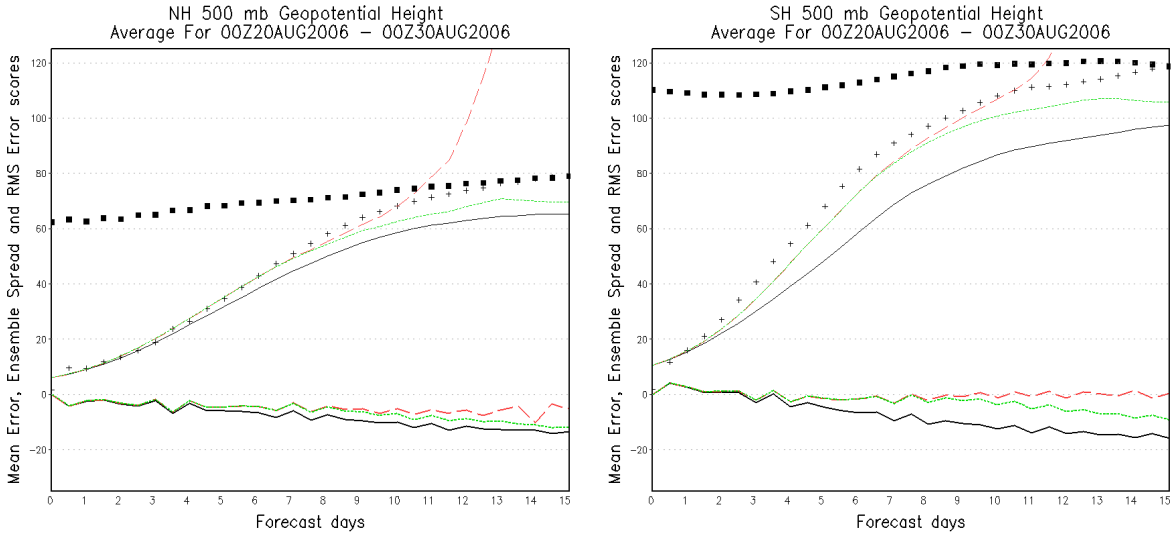

Fig. 3 Mean error (thick curves) of ensemble mean forecast and ensemble spread (thin) of IC (solid/black curves), ICSP_01 (long dashed/red) and ICSP_PW (short dashed/green), along with root mean square error of the IC ensemble (crosses) and the climatological forecast (black squares), for H500 averaged over the Northern (left panel) and Southern (right panel) Hemisphere extratropics.

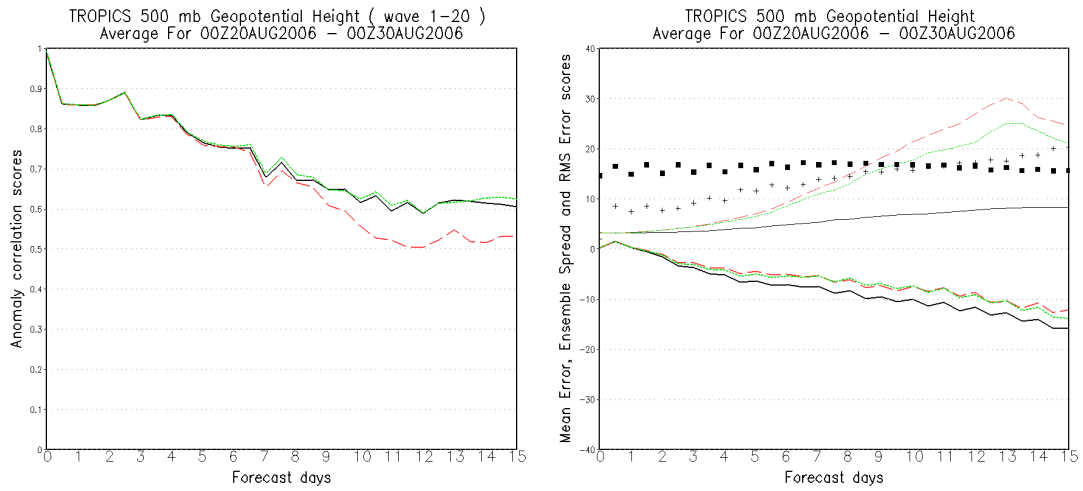

Fig. 4 Anomaly correlation (left panel), mean error (right panel, thick curves) of ensemble mean forecast and ensemble spread (right panel, thin curves). IC (solid/black), ICSP_PW (long dashed/red) and ICSP_PW_NC (short dashed/green) ensembles are compared for H500 forecast, averaged over the Tropics and the period of Aug. 20-30, 2006. Also plotted in the right panel are the corresponding RMSE of the IC ensemble (crosses) and the climatological forecast (black squares).

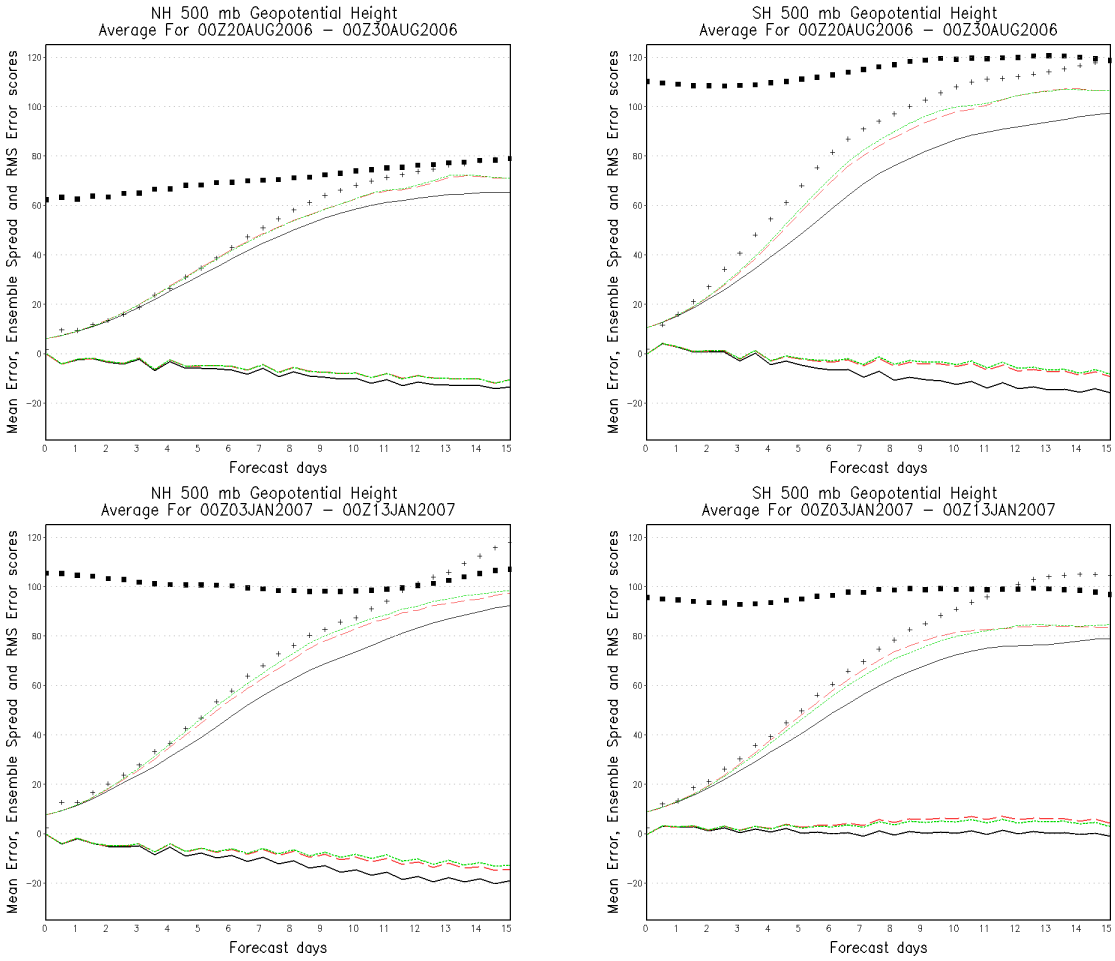

Fig. 5 Mean error (thick curves) of ensemble mean forecast and ensemble spread (thin) of IC (solid/black), ICSP_PW_NC (long dashed/red) and ICSP_PW_NC_V (short dashed/green) ensembles, along with RMSE of the IC ensemble (crosses) and the climatological forecast (black squares), for the H500 forecast averaged over the Northern (left panels) and Southern (right panels) Hemisphere extratropics, for the periods of Aug. 20-30, 2006 (upper panels) and Jan. 3-13, 2007 (lower panels).

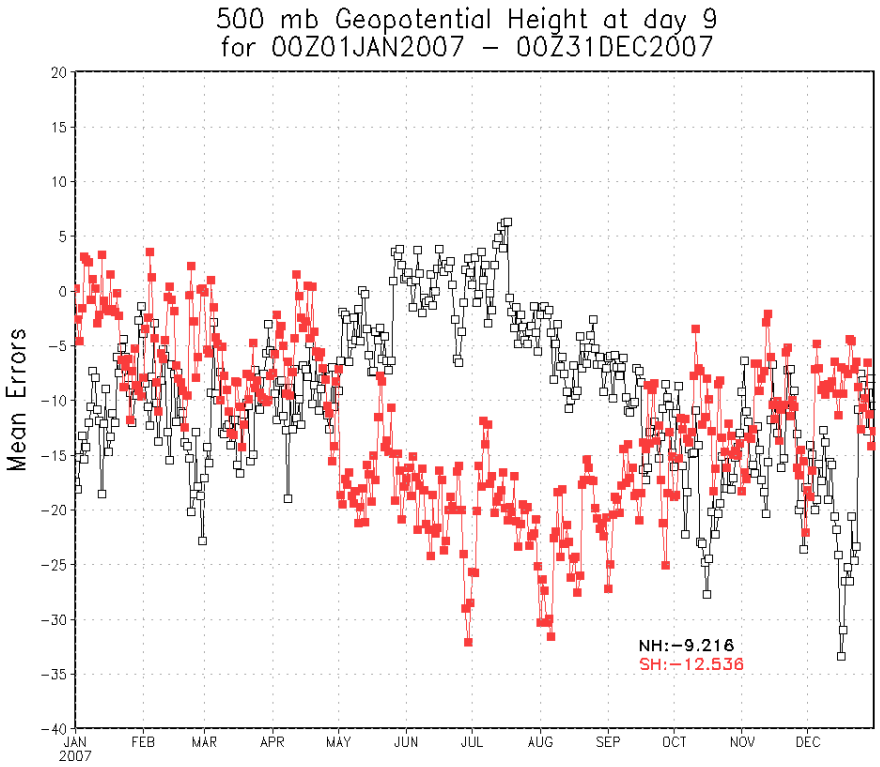

Fig. 6 Mean error of NCEP operational 9-day ensemble mean forecast of H500 averaged over the Northern (open squares) and Southern (closed square) Hemisphere extratropics, for the calendar year of 2007.

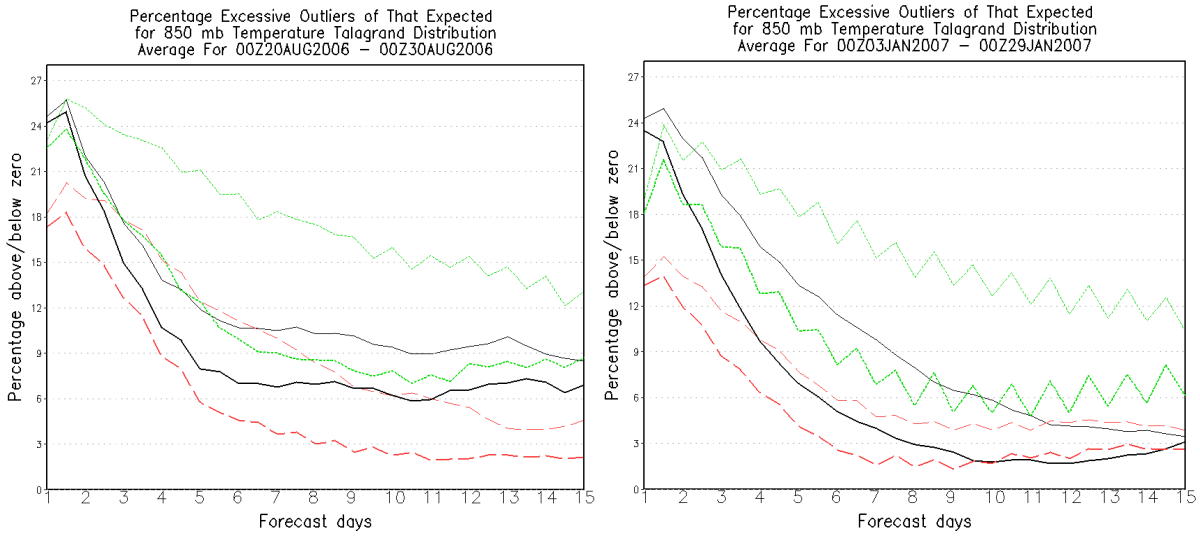

Fig. 7 Percentage of Excessive Outliers over the expected value, for 850 hPa temperature forecast averaged over the Northern (solid/black curve) and Southern (long dashed/red) Hemisphere extratropics and the Tropics (short dashed/green) for the IC (thin) and ICSP (thick) ensembles. The left/right panel is for the northern summer-fall/winter season.

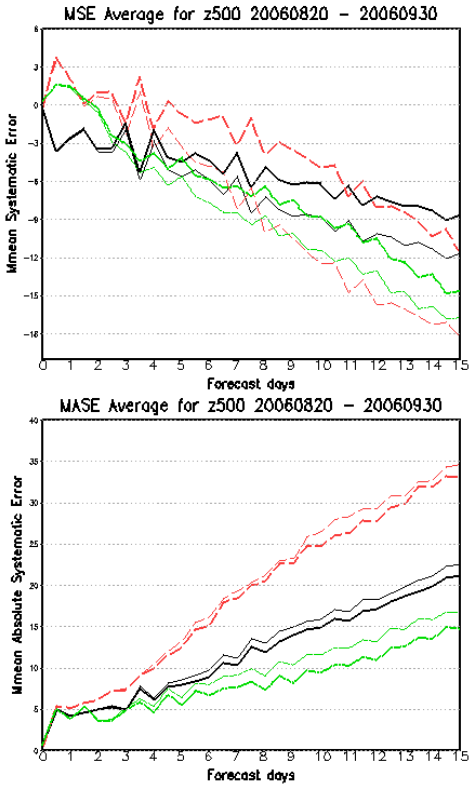

Fig. 8a Mean Systematic Error (MSE, upper panel) and Mean Absolute Systematic Error (MASE, lower panel) of 500 hPa geopotential height averaged over the Northern (solid/black curve) and Southern (long dashed/red) Hemisphere extratropics and the Tropics (short dashed/green), for the IC (thin) and ICSP (thick) ensembles averaged over the northern summer-fall season.

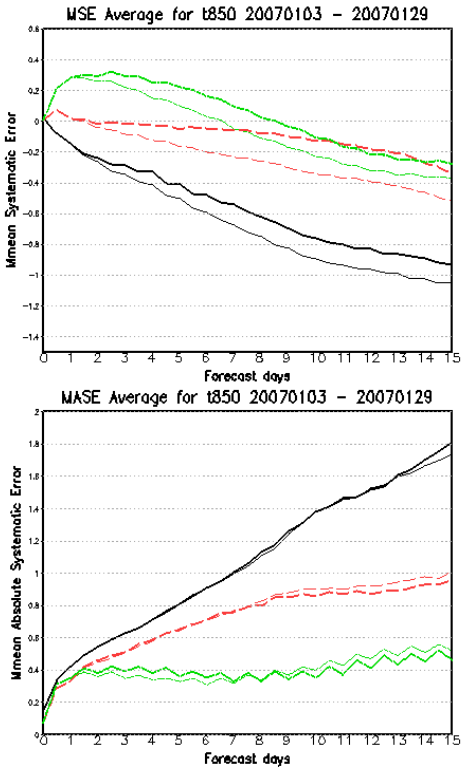

Fig. 8b Same as in Fig.8a except for average over the northern winter season.

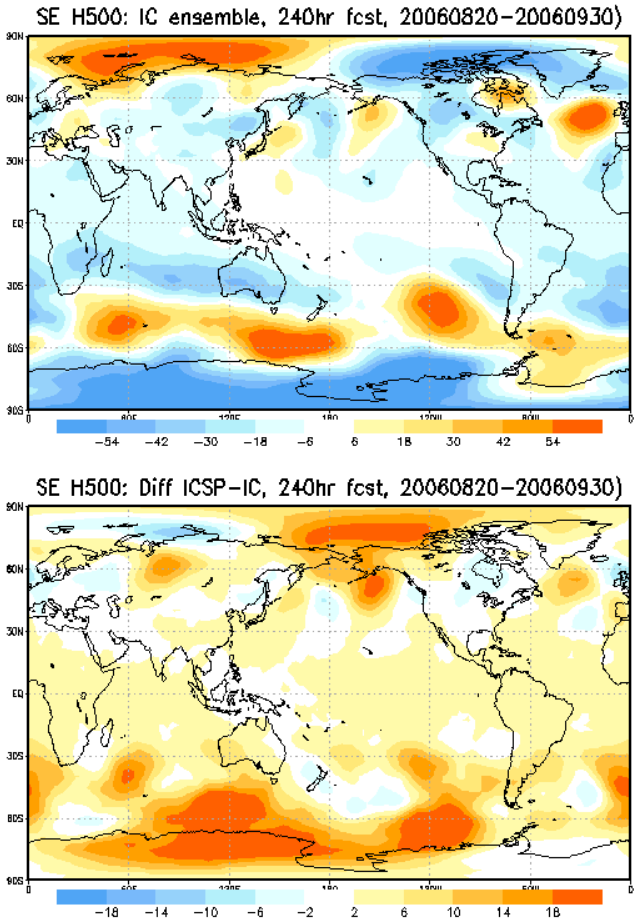

Fig. 9 Systematic Error (SE, upper panel) of the 240h forecast of 500 hPa height, for the IC ensemble (upper panel) and the difference between ICSP and IC ensembles (lower panel), averaged over the northern summer-fall season.

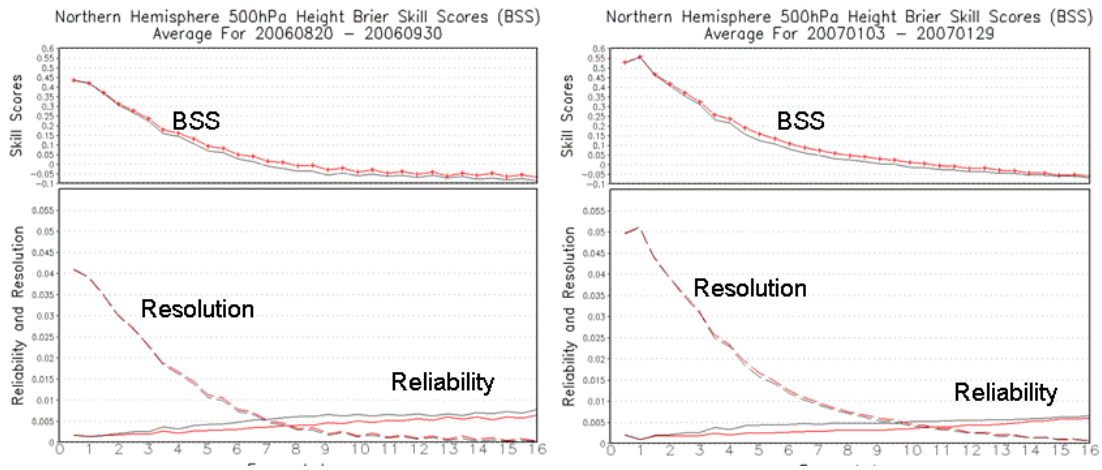

Fig. 10 Brier Skill Score (BSS) and the reliability and resolution components of Brier Score (BS), for 500 hPa geopotential height of the IC (light/black curves) and ICSP (dark/red curves), averaged over the Southern Hemisphere extratropics for the northern summer (left panel) and winter (right panel) seasons.

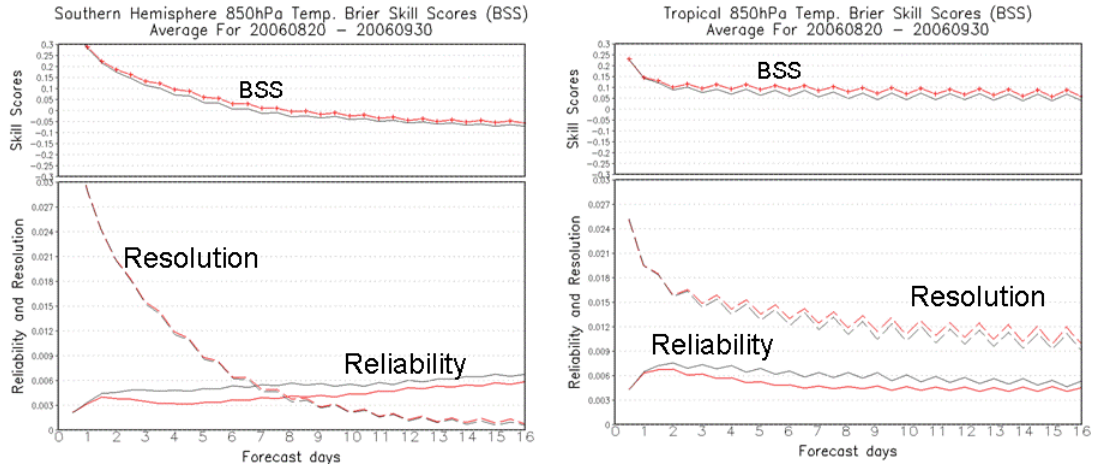

Fig. 11 As in Fig. 10, except for 850 hPa temperature averaged over the Southern Hemisphere extratropics (left panel) and tropics (right panel) for the northern summer season.

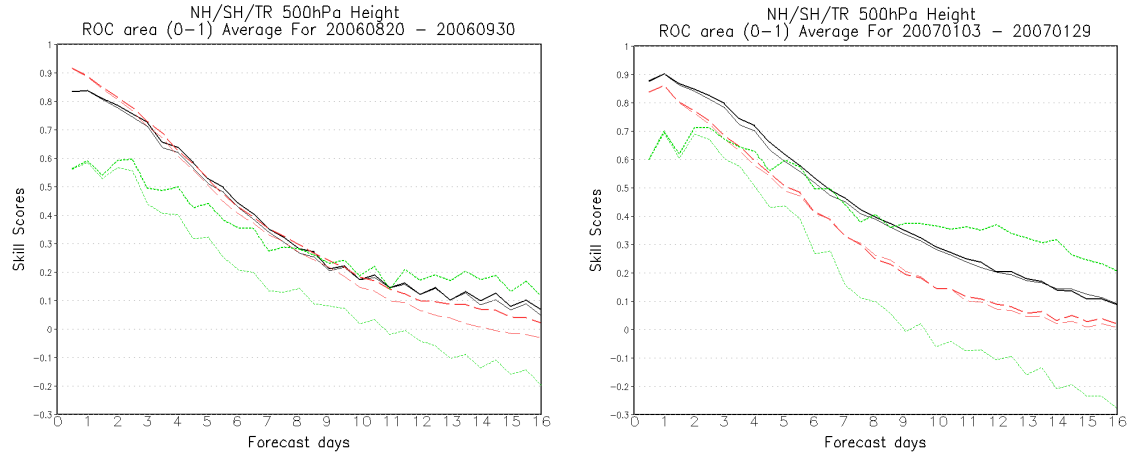

Fig. 12 ROC area of 500 hPa geopotential height, averaged over the Northern (solid/black curves) Southern (long dashed/red) Hemisphere extratropics and the tropics (short dashed/green) for the IC (thin) and ICSP (thick) ensembles, for the northern summer (left panel) and winter (right panel) seasons.

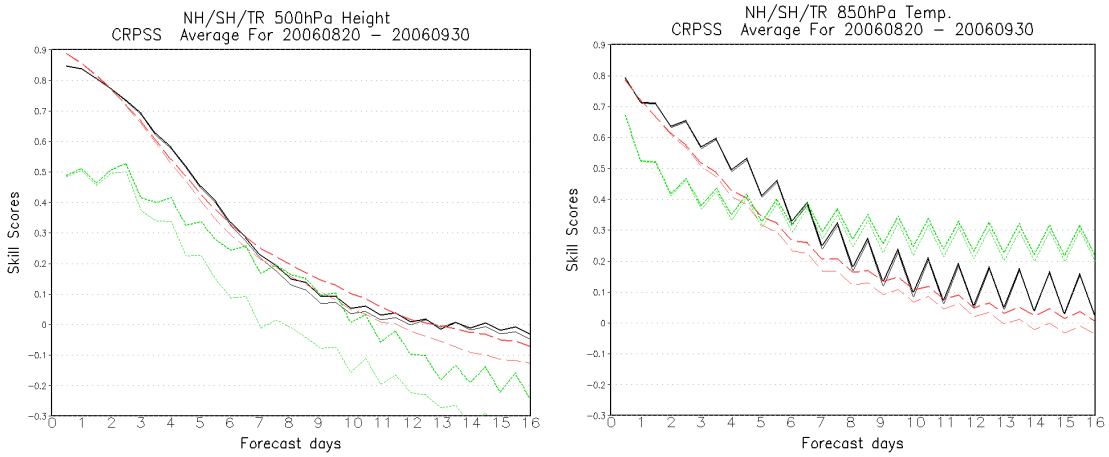

Fig. 13 CRPSS of 500 hPa geopotential height (left panel) and 850 hPa temperature, averaged over the Northern (solid/black curves) Southern (long dashed/red) Hemisphere extratropics and the tropics (short dashed/green) for the IC (light/thin) and ICSP (dark/thick) ensembles, for the northern summer season.

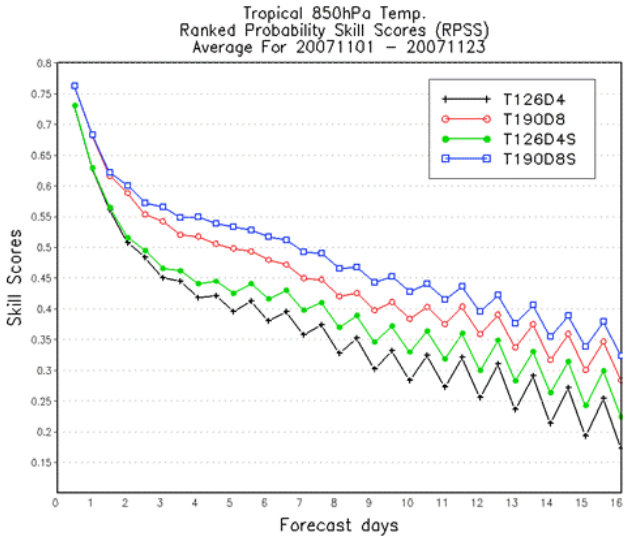

Fig. 14 RPSS of 850 hPa temperature, averaged over the northern tropics and the period Nov. 1-23, 2007, for the ensembles T126D4 (crosses), T190D8 (open circles), T126D4S (filled circles) and T190D8S, described in the text.

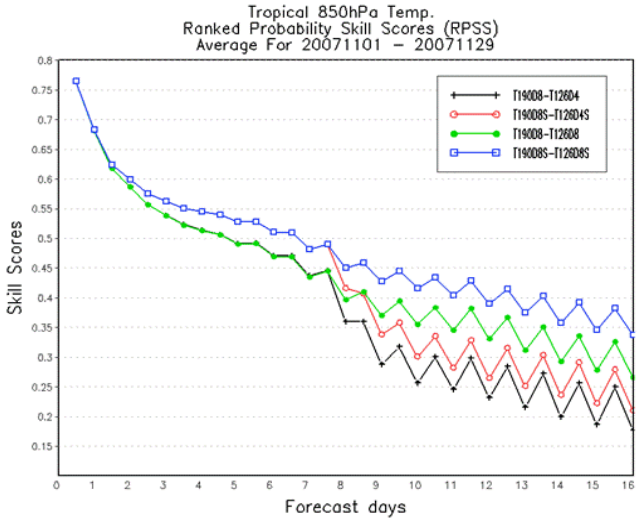

Fig. 15 RPSS of 850 hPa temperature, averaged over the northern tropics and the period Nov. 1-29, 2007, for the ensembles T190D8-T126D4 (crosses), T190D8S-T126D4S (open circles), T190D8-T126D8 (filled circles) and T190D8S-T126D8S, described in the text.