Predictability, Probabilistic Forecasting and Ensemble Prediction Systems

Yuejian Zhu [1] and Juhui Ma [1 2 3]

[1] Environmental Modeling Center
NCEP/NWS/NOAA
Washington DC, USA

[2] Institute of Atmospheric Sciences
Nanjing University of Information Science and Technology
Nanjing, China

[3] The National Center for Atmospheric Research
Boulder, CO, USA

ABSTRACT

Ensemble techniques have been used to generate daily numerical weather forecasts since the 1990's in numerical centers around the world due to the increase of computation ability. One of the main purposes of numerical ensemble forecast tends to assimilate initial uncertainty (both observation and analysis errors) and forecast uncertainty (model errors) by applying either initial perturbation method, ensemble assimilation, or multi-model/multi-physics method, and stochastic physics. In fact, the mean of ensemble forecasts is offering better forecast than deterministic (or control) forecast after a short lead-time (1-3 days) for the global model application. There is about a 1-2 day improvement in the forecast skill when using ensemble mean instead of a single forecast for longer lead-time. The skillful forecast (65% and above of an anomaly correlation) could be extended to 8-10 days (or longer) by present state-of-the-art analysis and ensemble forecast system. It is most important that ensemble forecast can deliver the probabilistic forecast directly, which is based on probability density function (PDF), instead of a single value forecast from traditional deterministic system to the users. It has long been recognized that the ensemble forecast is not only improving our weather forecast predictability but also offering a remarkable forecast for a future uncertainty, to help us making right decision, such as relative measure of predictability (RMOP), economic value (EV) and probabilistic quantitative precipitation forecast (PQPF). Not surprisingly, the success of ensemble forecast and its wide application are greatly increasing the confidence of model developers and research communities.

Note: This article is mainly expansion of published article "Ensemble Forecast: A New Approach to Uncertainty and Predictability" (Zhu, 2005) which contributes to the lecture notes for WMO/RTC publication.
Corresponding author address: Yuejian Zhu, Environmental Modeling Center/NCEP,
5200 Auth Road, Camp Springs, MD 20746, USA; e-mail: Yuejian.Zhu@noaa.gov

Chapter 3 Predictability, Probabilistic Forecasting and Ensemble Prediction Systems

3.1 Introduction

In the past decade, the methodologies followed at the National Centers for Environmental Predictions (NCEP) of the National Weather Service of the United States, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Canadian Meteorological Center (CMC) of the Meteorological Service of Canada have been developed to simulate the effect of initial and model uncertainties onto the forecast errors (see Table 1, 2, 3). In early studies, the characteristics of these three global ensemble prediction systems (EPS) have been discussed, and the objective evaluations have been taken by using the three ensemble forecasts for a 3-month period, May-June-July 2002 (Buizza et al. 2005). The probabilistic applications, the probabilistic evaluations and the differences between deterministic and ensemble forecast from NCEP EPS system have been presented in past years (Zhu et al. 1996; Zhu et al. 1999; Zhu et al. 2002 and Zhu 2004). In the part of this presentation, the experiments have been done based on global ensemble forecasts (Jun-July-August 2004) from world numerical centers to represent the improvement of numerical models and ensemble techniques over recent years. In additional, many other cases have been studies, too. Meanwhile, synoptic examples of probabilistic quantitative precipitation forecast (PQPF) from three numerical prediction centers have been exhibited side by side to allow us to compare each one. The multi-center ensembles, which are combined by NCEP EPS and ECMWF EPS, are studied to demonstrate the new approach of ensemble method, which is from different initial condition generation methods, different assimilation systems (initial conditions), different forecast models (dynamics and physical parameterizations) and different model resolutions (vertical and horizontal resolutions). The importance of all these studies are not to rank the performance of the ensemble systems, but to identify possible reasons of superior/inferior performance, thus drawing a guideline for the future ensemble development, and improving the ensemble forecast system and predictability.

This paper will discuss the importance of the ensemble forecast in the next section. After that, the methods of ensemble forecast will be briefly reviewed, and the objective evaluations of forecast from improved, state of the art ensemble systems will be presented in terms of deterministic (or/and ensemble mean) and probabilistic (distribution) concepts in section 4. In section 5, the multiple applications of the ensemble forecast will be introduced. The experimental multi-center ensemble forecast will be discussed in section 6 through selected combinations from comparable ensemble systems. In additional to the discussion of the forecast skill in section 4, the effect of model initial condition and model resolutions will be investigated from one-year statistics.

3.2 Why Do We Need Ensemble Forecast?

There are two main reasons to emphasize the importance of ensemble model forecast. One is the forecast error (uncertainty), which comes from each process of numerical weather prediction system, such as observation and data collection (observation system), data assimilation (analysis system) and forecast model (dynamical process, computation, physical parameterization and et al.). Early studies (Lorenz 1969; Lorenz 1982) suggested that the initial error could grow very fast into the different scales no matter how small the initial error. In fact, the forecast error will be increased continually with model integration before it is saturated. The optimum solution to capture and reduce this forecast error (uncertainty) is to use ensemble forecast instead of single (deterministic) forecast. As shown in the figure 1, the initial probability PDF(D) represents the initial uncertainties. From the best estimate of the initial state a single deterministic forecast is performed. This single deterministic forecast fails to predict correctly the future state. However, an ensemble of perturbed forecasts starting from perturbed initial conditions designed to sample the initial uncertainties can be estimate the probability PDF(D+n) at future time (D+n) (Buizza et al., 2001). Figure 2 is a real example for bimodality, trimodality and uncertainties. The advantage of ensemble forecast to estimate the uncertainties is because the ensemble forecast is producing a set of randomly-equally-likely (independent) solutions for the future. The diversity of these solutions, which is called forecast spread, is mostly representing the forecast uncertainty. The relationship of ensemble spread and ensemble mean error (uncertainty) has been discussed in the early study (Zhu 2004) and will be discussed again in this study. The perfect ensemble prediction system is expected to have a similar spread to their mean error (or high correlation between ensemble spread and ensemble mean error) in a long term statistics. How much does the ensemble spread represent forecast uncertainty in the real atmosphere? It can not be answered quantitatively. It depends on the sizes of spread and error, the distribution of the error and et al. In fact, the skill of ensemble forecast is greatly improved when comparing the ensemble mean forecast to a deterministic forecast after a short lead-time. The ensemble mean forecast for a short lead-time is degraded due to introduce initial perturbation (error) for both of NCEP EPS and ECMWF EPS. Another reason is predictability. The atmosphere is a chaotic system as the solutions of the equations for Lorenz 63 model (Lorenz, 1963) shown in figure 3,

$$\frac{dx}{dt} = \delta(y - x)$$
$$\frac{dy}{dt} = \gamma x - y - xz ,$$
$$\frac{dz}{dt} = xy - \beta z$$

where the parameters $\delta = 10, \beta = 8/3, \gamma = 28$. There are two regimes corresponding to positive values of x and y, and negative values of x and y separately. It's hard to predict the occurrence of regime changes which makes the atmospheric system lacking of long term predictability. Knowing the future has always been a practical and spiritual need to people. The ultimate goal of all scientific works has also been successful prediction. The success of our prediction efforts depends on two main factors: (1) our understanding and knowledge of natural processes and (2) the nature of these processes to be predicted. The increasing of forecast predictability is always corresponding to the decreasing of forecast uncertainty. The reduction of forecast error from ensemble forecast is greatly increasing the predictability. In addition, when considering the forecast itself and user community, one of the goals of the United States National Weather Service for 2000-2005 requires to provide weather, water and climate forecasts in probabilistic terms by the year 2005 (NWS 1999), which is the most achievable and practical with ensemble forecast. In the past, there were many methods to generate the probabilistic forecast, but the ensemble model forecasts could achieve this goal easily and accurately. As expected, the probabilistic forecast, such as spaghetti diagram (to describe uncertainty), PQPF (to tell the probabilistic forecast) (Zhu et al. 1998; Zhu and Toth 1999; Zhu 2004 and Zhu 2005), RMOP (related to predictability) (Toth et al. 2001), ensemble spread (similar to spaghetti diagram, but more completely) and et al., is more popular to users and publics in past years.

3.3 Methodologies of Ensemble Forecast

As noted earlier (Buizza et al. 2005), there are two major methods to generate ensemble model forecast around world meteorological centers. One of them is initial perturbation method, which is adding small perturbations to initial analysis, such as NCEP's breed mode method (Toth and Kalnay 1993; Tracton and Kalnay 1993; Toth and Kalnay 1997a) which extends linear concept of Lyapunov Vectors into nonlinear environment to sample subspace of most rapidly growing analysis errors, and ECMWF's singular vector method (Palmer et al 1992; Molteni et al 1996) which is characterized by the fastest growth, and measured using a total energy norm over a finite time interval. Methods for NCEP and ECMWF are assuming forecasting model is perfect, and to assimilate initial (observation and analysis/data assimilation) uncertainty by using small and random initial perturbation. The characteristics of Lyapunov, breeding and singular vectors are listed in table 4.

In May 2006, NCEP GEFS implemented ensemble transform with rescaling (ETR) method to simulate the initial perturbation (Wei et al., 2008). At every 6-hour cycle, both ET and Simplex Transformation (ST) are carried out for all 80 perturbations. ST which is used to ensure members are centered around the analysis is only imposed on 20 perturbations which are used for long forecast. The remaining 60 members are integrated for 6-hour, for cycling. As shown in figure 4, the initial perturbations are centered around the analysis to improve ensemble mean and have simplex structure, not paired. The perturbations have maximum number of effective degrees of freedom which means that the variance will be maintained in as many directions as possible within the ensemble subspace. They are uniformly centered and distributed in different directions. The larger the ensemble, the more orthogonal they become. They become orthogonal if the number of members approaches to infinity. The initial perturbations have flow-dependent spatial structure if the analysis error variance is derived from operational DA system at every cycle. The covariance constructed from the perturbations is approximately consistent with the analysis covariance from the DA if the number of ensemble members is large enough.

In order to improve the initial perturbations of hurricane ensemble forecasts, the hurricane relocation algorithm from the GFS model was modified to be implemented in GEFS in July 2005 (Zhu et al., 2005). Experiment results show that the track spread from individual members is significantly reduced. The new system can be summarized as follows (see figure 5): (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; (4) Add the hurricane perturbation and global ensemble perturbation to the analysis fields to create the model initialization.

At ECMWF, EDA-based perturbations replace evolved singular vectors in the generation of the EPS initial conditions in June 2010, which results in a better spread-skill relationship in the early forecast range over the extra-tropics, and for the whole forecast range over the tropics. The EDA perturbed members are generated by perturbing all observations and the sea-surface temperature field and using the stochastically perturbed parameterization tendency (SPPT) scheme that perturbs the total parameterized tendency of physical processes to simulate random model error (Buizza et al., 2010).

Another set of ensemble forecast is produced by using different numerical models (spectrum model and grid model) and different physical packages in the CMC (Houtekamer and Derome 1995; Houtekamer et al. 1996) before July 2007. The 8 different physical packages have been used in CMC's global EPS. There are, in total, 16 (2 models, 8 different physical packages) ensemble runs from 0000UTC. It is to assimilate initial (by different models) and forecast (by different physical schemes) uncertainties. At ECMWF, Stochastically Perturbed Parametrization Tendencies (SPPT) has been applied to perturb the total parameterized tendency of physical processes with multiplicative noise since 1998 (Buizza et al., 1999). But the random patterns which are constant make the perturbations aren't continuous in space and time. The patterns which produced by the revised SPPT scheme implemented operationally in September 2009 vary smoothly in space and time (Palmer et al., 2009). NCEP GEFS also develops a Stochastic Total Tendency Perturbation (STTP) scheme to simulate model related uncertainties in the recent year (Hou et al. 2010). STTP sample the random errors associated with the total tendency, including both dynamical and physical processes, grid resolved and parameterized components. The general framework with stochastic presentation of model related uncertainties is to add a stochastic forcing term *S* to the conventional tendency *T*, for each ensemble member *i*, i.e.,

$$\dot{X}_i = T_i + S_i \, .$$

The stochastic forcing is linked to the total conventional forcing by sampled from the differences in the conventional tendency between the ensemble members and the control forecast,

$$P_i = T_i - T_0 \, .$$

The *S* terms are formulated by various combinations of the *P* vectors,

$$S_i \sim \sum_j w_{i,j} P_j \, .$$

where *i* and *j* are the index of the ensemble members. The results form experiments indicate that the application of STTP can increase the ensemble spread and reduce systematic error of the ensemble mean forecast. Moreover, it can significantly improve the ensemble based probabilistic forecast, especially in the tropical region.

Meanwhile, in both research and development centers, many other ensemble forecasts have been studied from statistic post processes such as super-ensemble (Krishnamurti et al. 1999), poor-man ensemble (Ebert 2001), Monte Carlo or lagged average forecast (LAF) ensembles for climate study and et al. In the section 6, we will discuss the multi-center ensemble forecast by the combination of NCEP EPS and ECMWF EPS, as well.

3.4 The Skill of Ensemble Forecast

Before we discuss the skill of ensemble forecast, let's review the effect of model initial condition and model resolutions. By running a one-year statistic average (June 1st 2003 – May 31st 2004) of verification scores, the pattern anomaly correlations (PAC) of the NCEP Global Forecasting System (GFS: high resolution control, T254L64 from 0-84 hours, T170L64 from 84-180 hours, T126L64 from 180-384 hours), the NCEP ensemble control (CTL: low resolution control/ensemble control, T126L28 from 0-180 hours, T62L28 from 180-384 hours) and the NCEP 10 ensemble members (5 pairs initial perturbations with the same resolution as ensemble control) are calculated. When comparing GFS and CTL to 10 individual ensemble members, 100% will be awarded if GFS/CTL is better than all individual ensemble members, otherwise, 0% will be given if GFS/CTL is worse than all members, 50% will be added if GFS/CTL is better than 5 of 10 ensemble individual members (randomly). The result is shown in figure 6 for up to 15 days lead-time, 500hPa geopotential height for northern hemisphere latitude band (20-80N). For the short lead-time (0-96 hours), high resolution GFS is best, the individual ensemble perturbation forecasts are far behind the either GFS (due to the resolution and initial error) or CTL (due to the initial error). After a short lead-time (120 hours), the model resolution is not as important as the first 96 hours to improve model forecast skills, as unexpected, CTL is slightly better than GFS from 144 hours to 264 hours lead-time in this experiment period. The differences between GFS/CTL and individual ensemble members are reduced. After 11 days lead-time, the PAC is very low (less than 50%) which will not be considered as a skillful forecast for the synoptic system, and then GFS is better than CTL again. Interestingly, both GFS and CTL are still better than any of individual ensemble members. As noted earlier, the difference between CTL and ensemble members is only initial condition. The difference between CTL and GFS is only resolution. The 50% line is a reference to consider the equality of GFS/CTL and 10 individual ensemble members. Based on the results presented in this section, we should point out that both resolution and initial condition are very important to model forecast. The resolution is playing a key role to success the short-range forecast while the influence of the resolution is much smaller than initial condition for medium-range forecast.

The forecast skills could simply be measured in terms of pattern anomaly correlation (PAC) scores (depends on climatological information) and root mean square (RMS) errors of 500hPa (or other levels) geopotential height (or other variables) by considering any deterministic (control) forecast and ensemble's mean. The assessment of the past status (summer of 2002) for three major global ensemble prediction systems was presented by Buizza et al. (2005). The objective evaluation of the present status has been done by using the similar methodologies. The figure 7 shows the northern hemisphere extra-tropic (20N-80N) 500hPa geopotential height PAC scores of three different EPSs mean (solid lines, considering first 10 ensemble members only for each center, using NCEP/NCAR reanalysis data as climatology) comparing to their own deterministic/control forecast (dotted lines) from Jun-August 2004. The verified analysis is used from their own data assimilation system in this experiment. Similar results were obtained by Buizza et al. 2005 except for the improvement of all three systems. The means of ensemble forecasts do not show advantages for first 3 (up to 5, depends on the model and season) days due to introduced initial errors by NCEP and ECMWF. There is similar result of CMC's ensemble for very short lead-time because of using one verified analysis for two different initial conditions (note: only one analysis is available to verification). However, after 3 (up to 5) days lead-time, ensemble means have 6 hours to 24 hours (or longer) advantage than their own deterministic forecast. Unfortunately, there are only 6 days of lead-time available for the evaluation to CMC's control forecast. The differences between ensemble mean and their own deterministic forecast are very similar for three EPSs. Therefore, the improvement of ensemble forecast mostly depends on its analysis and forecast model from these experiment results. The skill of ECMWF's deterministic forecast is slightly better than other deterministic forecasts, and the PAC scores of ECMWF ensemble mean are leading for all of these forecasts, too. Of course, the costs of these three EPSs are slightly different. Less computation times are needed for NCEP's breeding method; in contrast, it is more difficulty to maintain and develop the CMC's method if resources are limited.

When considering skillful forecast, usually defined 65% and above PAC scores based on the synoptic scale forecast (short- medium-range), the NCEP ensemble mean is offering 7 days and 8 hours useful forecast instead of NCEP GFS (deterministic) which has 6 days and 14 hours skillful forecast (see figure 7) by using approximately the same computation resource. There is an 18 hour improvement when considering ensemble mean only in this three-month summer period, which is huge gain compared to the improvement from observation system, data assimilation and forecast model.

The RMS error is another measurement, which does not depend on the climatology. The results of the same period (Jun-August, 2004) for NH 500hPa geopotential height are shown in figure 8. The solid lines are for the ensemble's mean, dotted lines are for their ensemble control (or deterministic forecast). ECMWF's control forecast has a smaller error for first 4-day, after that, ECMWF's ensemble mean is better than ensemble control. It is interesting to note that in the NCEP forecast, either ensemble mean or ensemble control, the forecast errors increase very rapidly in the first 24-hour, after that, the error growth rates are very similar (or close) to ECMWF's. Does this indicate something we need to work on in the future?

Another importance to evaluate ensemble forecast is to use probabilistic methods, such as Rank Probability Score (RPS), Continuous Ranked Probability Score (CRPS), Brier Score (BS), Hitting Rate (HR) and False Alarm Rate (FAR), potential Economic Value (EV), Relative Operating Characteristics (ROC) area and et al (Zhu et al. 1996; Zhu et al. 2002; Toth et al. 2003; Zhu 2004; and Buizza et al. 2005), in which, CRPS and RPS measure the reliability and resolution. The formulas can be written as follows:

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

Where the Heaviside Function H is

$$H(x - x_0) = \begin{cases} 0, x \le x_0 \\ 1, x > x_0 \end{cases}, \qquad \text{and}$$

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

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

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

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. 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. And here is the final formula for decomposition: *BS = Reliability – Resolution + Uncertainty.* The typical application for HR and FAR is the Relative Operational Characteristics (ROC) curve (Toth et al., 2003), or sometimes called the ROC area. Another application is economic value (EV) of a weather forecast by using cost-loss analysis method and considering user reaction (Zhu et al., 2002). The EV is an estimation of forecast resolution, which is the ability of a forecast system to discern sub-sample forecast periods with different relative frequencies of an event.

3.5 Applications of Ensemble Forecast

Many new products have been generated since global ensemble forecasts started. The typical example of early ensemble graphic application is spaghetti diagram (Toth et al. 1997b). Later, the PQPF for different threats (such as 0.1 mm/24 hours, 2 mm/24 hours and so on) has been used for operational application since 1997 in NCEP (Zhu et al. 1998; Zhu et al. 1999; Zhu 2004). The calibrated QPF and PQPF have been implemented in May 4th 2004 which applied bias removed techniques (Zhu 2005). The ensemble mean and spread are the standard products in NCEP since 2000. Recently, precipitation type based probabilistic forecast is implemented for every 6-hours lead time, the products include probabilistic quantitative rain forecast (PQRF), probabilistic quantitative snow forecast (PQSF), probabilistic quantitative freezing rain forecast (PQFF) and probabilistic quantitative ice pellets forecast (PQIF). The values of the relative measure of predictability (RMOP) are calculated globally since 2000 (Toth et al. 2001; and Zhu 2004). Figure 9 shows a synoptic example of probabilistic quantitative precipitation forecast (PQPF) of the Northern American (NA) area for three comparable global ensemble systems (NCEP, CMC and ECMWF EPSs). The initial time is April 26 0000UTC 2004 for NCEP and CMCs, April 26 1200UTC for ECMWFs. The lead times are 12-36 (0-24), 36-60 (24-48), 60-84 (48-72), 84-108 (72-96) and 108-132 (96-120) hours for NCEPs and CMCs (ECMWFs). The contour levels are for 5%, 35%, 65% and 95% respectively. The forecasts of the main future are very close to each other up to 5 days. When verified to the observations (from rain gauge, not shown), all of them make very good forecasts. Through a number of investigations, we are expecting to have more joint ensemble products in the future through Northern American Ensemble Forecast System (NAEFS) project which endorsed by the National Weather Service of United States, the Meteorological Service of Canada and the National Meteorological Service of Mexico in November 2004.

3.6 Multi-Center Multi-Model Ensemble

The Observation, Research and Predictability Experiment (THORPEX) is a major component of the World Weather Research Program under the WMO, which leads to new techniques in observations, data assimilation, forecasting and socioeconomic applications. A key goal of THORPEX is to accelerate improvements in the accuracy of 1-day to 2-week high-impact weather forecasts for the benefit of humanity. The THORPEX Interactive Grand Global Ensemble (TIGGE) project, a main component of THORPEX, is initiated to enhance international collaboration research on multi-model ensemble forecast and pave the way towards operational implementation (Bougeault et al., 2009).

NAEFS is a new weather modeling system run jointly by the Meteorological Service of Canada (MSC) and the U.S. National Weather Service (NWS) to provide numerical weather prediction (NWP) products to weather forecasters in both countries for a forecast period that runs out to 16 days. The NAEFS combines the Canadian global forecast model ensemble and the NWS global forecast model ensemble into a joint ensemble that will create weather forecasts for all of North America. At present, all the national weather agencies in North America are participating in NAEFS - the Meteorological Service of Canada, the National Meteorological Service of Mexico, and the U.S. National Oceanic and Atmospheric Administration NWS. NAEFS provides framework for transitioning research into operations and will be a prototype for ensemble component of THORPEX legacy forecast system-Global Interactive Forecast System (GIFS) (Toth et al., 2005). It is very clear to see there is about more than days gain when comparing CRPS scores of NAEFS products to NCEP or CMC's raw forecasts for Northern Hemisphere 2 meters temperature field in figure 10 and Northern Hemisphere 500hPa geopotential height in figure 11.

Multi-model super-ensemble for weather and climate application has been discussed many years ago (Krishnamurti and et. al. 1999). The study is mainly focused on climate prediction by applying statistic method and training data. After that, a poor man's ensemble has been investigated to predict the PDF of 1-2 days precipitation forecast (Ebert 2001) which using a set of individual model from several operational center. Therefore, the ensemble size is limited. The experiment in this study tends to combine two similar ensemble systems from NCEP and ECMWF. The advantages of this combination could be the improvement of forecast skill, the less computation usage, the large ensemble size and so on.

First, comparing two ensemble systems from NCEP and ECMWF, both of them using initial perturbation method and available on 1200UTC initial runs, the overall skills are very comparable to each other. After reviewing NCEP and ECMWF EPSs, the experiments are designed to combine two ensembles by selecting (1) 10 members from NCEP first 6 members and ECMWF first 4 members (verifying against NCEP analysis) and (2) 10 members from ECMWF first 6 members and NCEP first 4 members (verifying against ECMWF analysis) in order to match/compare NCEP 10 members ensemble at 1200UTC cycles. The figure 12 is the PAC scores of ensemble mean for up to 10 days, in Jun-August 2004 of NH extra-tropic (20N-80N) 500hPa geopotential height. The two new combined ensembles (closed cycle and open squares) are better than either NCEP or ECMWF original 10-member ensemble. For example, there are about 8 hours of improvement for a 5-day forecast when comparing NCEP ensemble (cross) to new ensemble (2) (open square). Figure 13 is the same as figure 12 but for ROC area verification. ROC area is calculated based on accumulative hit rate and false alarm rate of 10 climatologically-equally-likely intervals (Zhu et al. 1996, Mason, 2003). Both of the new ensembles (1) and (2) are better than NCEP ensemble for all lead-time. The new ensemble (2) is better than ECMWF ensemble in the all ways except for its longer lead-time. The new ensemble (1) is mixed with a short lead time but is slightly worse than ECMWF ensemble in this experiment. A tentative explanation for this result is that the systematic errors (bias) are still in both forecasts and analyses. The probabilistic skills (include resolution and reliability) of the new ensembles should be improved by removing bias (or related pre-process) before they are combined. It is still questionable for the combined ensemble if the original ensemble systems are very different, such as the system design, forecast skill, and spread. The further studies are required to answer this question.

3.7 Discussion and Conclusion

Let us discuss the relationship between the RMS error of ensemble mean and ensemble spread. The RMS error is the distance measurement from ensemble mean (of forecasts) to truth (analysis). The spread is measuring the distance from ensemble mean (of forecasts) to each individual ensemble member. Apparently, a perfect ensemble forecast would expect the size of ensemble spread equal (or close) to the RMS error, which means the ensemble spread will maximally represent forecast uncertainty. But in fact, our current ensemble prediction systems have less spread than RMS error for medium- and extended-range forecasts (Zhu 2004; Buizza et al. 2005), which means the ensemble forecasts are insufficient to capture reality systematically, or none of them is able to simulate all sources of forecast uncertainties for this chaos system. In the practical, the medium-range forecast could be improved by reducing RMS error and increasing spread. The recent experiments indicate (not show here) that the RMS error could be reduced by using statistic calibration (or bias correction), while the spread could be increased by introducing stochastic process (or other techniques) in the NWP model.

The skill of ensemble forecast relatively depends on the quality of our observing system, the data assimilation system (analysis/initial condition) and forecast model (dynamics, physical process and et al.). When the importance of developing ensemble prediction system is emphasized, it is shouldn't be forgot to pay attention to improve our basic numerical weather prediction (NWP) system, which includes data assimilation and forecast model. The model resolution is a key to make superior forecast for the first 1-6 days. After that, the high resolution doesn't take the much advantage due to lack the predictability by nonlinear interaction, physical parameterization and et al. The initial condition is most important to make good forecast from short-, medium- and extended-range.

It is very difficulty to simulate all possible errors (or uncertainties) perfectly in present EPSs. The multi-model and multi-analysis may be the better one to approach, but the cost for maintenance and development is more expensive by any numerical center. Therefore, a tentative conclusion could be: 1) the effort to improve analysis and forecast model could benefit both of ensemble and deterministic forecast; 2) ensemble post process is another way to enhance a forecast skill by using statistic bias correction; 3) combined multi-center, multi-model ensemble with bias correction could approach the goal closely in the future.

3.8 Future Expansion

The prediction science (Zhu, 2010) is a new way to expand modern weather-climate forecast system. Today, the sciences, the engineering and the arts (Zhu, 2010) are playing different roles, all serving the weather forecast system, to ensure that the public users are receiving accurate, reliable forecasts, to help public making right decisions and protecting their properties. The book COMPLETING THE FORECAST: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts (National Research Council of the National Academic, USA. 2006) is introducing the concepts of modern forecast, describing the details of future uncertainty forecast. There are two concepts are very important to us: 1) forecast uncertainty which is for scientist and developer; 2) uncertainty forecast which is the future forecast we will issue for public. Simply from the study, the forecast has an uncertainty which gets larger when the leading time is increasing (Toth and et. al, 2001; Zhu and et. al, 2002; Zhu, 2005; Toth and et. al, 2007). Therefore, we need to make uncertainty forecast for completing service. In order to understand an uncertainty forecast for public users, we need to have the helps from the public education. For example, where are the uncertainties coming from? What is seamless forecast? Figure 14 shows United States NWS seamless suites of forecast products spanning on climate and weather. A couple of important points should be address here 1) global ensemble forecast system is a linkage between weather and climate, 2) a forecast uncertainty is growing with increasing forecast lead times.

Acknowledgments: The authors would like to thank Dr. Zoltan Toth of ESRL/NOAA for many valueable discussions, to thank Drs. Mozheng Wei, Dingchen Hou, Qingfu Liu, Bo Cui and Jiayi Peng for their contributions. We acknowledge the support of Dr. Stephen Lord, Director of Environmental Modeling Center.

References:

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev*., 127, 1941-1953.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* 125, 2887-2908.

Buizza, R., D. Richardson, T. Palmer, 2001: The new 80-km High-Resolution ECMWF EPS. Newsletter n. 90, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pg 2-9.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: Assessment of the status of global ensemble prediction. *Mon. Wea. Rev*. 133, 1076-1097.

Buizza, R., M. Leutbecher, L. Isaksen, and J. Haseler, 2010: Combined use of EDA- and SV-based perturbations in the EPS. Newsletter n. 123, ECMWF, Shinfield Park, Reading RG2-9AX, UK, pg 22-28.

Bougeault, P., Z. Toth, C. Bishop, Barbara Brown, David Burridge, DeHui Chen, Beth Ebert, Manuel Fuentes, Tom Hamill, Ken Mylne, Jean Nicolau, Tiziana Paccagnella, Young-Youn Park, David Parsons, Baudouin

Raoult, Doug Schuster, Pedro Silva Dias, Richard Swinbank, Yoshiaki Takeuchi, Warren Tennant, Laurie Wilson and Steve Worley, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Met. Soc.,* (in press).

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.* 129, 2461-2480.

Hou, D., Z. Toth, Y. Zhu, W. Yang and R. Wobus, 2010: A Stochastic Perturbation Scheme Representing Model-Related Uncertainties in the NCEP Global Ensemble Forecast System. Manuscript to be published.

Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev*., 123, 2181-2196.

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

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

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

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, 21, 289-307.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus,* 34, 505-513.

Mason, Ian B., 2003: Binary Events, in book of *Forecast Verification: A practitioner's guide in atmospheric science*. Edits.: I. T. Jolliffe and D. B. Stephenson. Wiley, 37-76

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc*., 122, 73-119.

NWS, 1999: Vision 2005: National Weather Service Strategic Plan for Weather, Water, and Climate Services 2000-2005. 24 pp. [Available from NWS, 1315 East-West Highway, Silver Spring, MD 20910].

National Research Council of the National Academic (NRC), 2006: Completing the Forecast: Characterizing and communication uncertainty for better decisions using weather and climate forecasts. [EB/OL]. 2006 [2010.02.09]. http://www.nap.edu

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. ECMWF Research Department Tech. Memo. 188, 45pp.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. *ECMWF Tech. Memo.* no. 598, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. Toth. Z., and E. Kalnay, S. M. Tracton, R. Wobus, and J. Irwin, 1997a: A synoptic evaluation of the NCEP ensemble. *Weather and Forecasting*, 12, 140-153.

Toth, Z., and E. Kalnay, 1997b: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* 125, 3297-3319.

Toth. Z., Y. Zhu, and T. Marchok, 2001: On the ability of ensembles to distinguish between forecasts with small and large uncertainty. *Weather and Forecasting*, 16, 436-477. *Bull. Amer. Meteor. Soc.*, 74, 2317-2330.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and Ensemble Forecasts, in book of *Forecast Verification: A practitioner's guide in atmospheric science*. Edits.: I. T. Jolliffe and D. B. Stephenson. Wiley, 137-163.

Toth, Z., J. Desmarais, G. Brunet, Houtekamer, Y. Zhu, R. Wobus, R. Hogue, R. Verret, L.Wilson, B. Cui, G. Pellerin, B. Gordon, E. O'Lenic, and D. Unger, 2005: The North American Ensemble Forecast System (NAEFS). In Proceedings of the 1st THORPEX International Science Symposium, December 2004, Montreal, Canada.

Toth, Z., O. Talagrand, and Y. Zhu, 2006: The Attributes of Forecast Systems. In book of *Predictability of Weather and Climate*, Ed.: T. N. Palmer and R. Hagedorn, Cambridge University Press, 584-595.

Tracton, M. S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational implementation. *Weather and Forecasting*, 8, 379-398.

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.

Wilks, D. S., 1995: Statistical Methods in the Apmospheric Sciences. Academic Press, New York, 467pp.

Zhu, Y., G. Iyengar, Z. Toth, S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. 15th AMS Conf. on Weather Analysis and Forecasting. Norfolk, VA. *Amer. Meteor. Soc*. J79-82.

Zhu, Y., Z. Toth, E. Kalnay, and S. Tracton 1998: Probabilistic Quantitative Precipitation Forecasts based on the NCEP global ensemble. Special Symposium on Hydrology, Phoenix, AZ. *Amer. Meteor. Soc.* J8-11.

Zhu, Y., and Z. Toth, 1999: Objective Evaluation of QPF and PQPF Forecasts Based on NCEP Ensemble, Preprints, Third International Scientific Conference on the Global Energy and Water Cycle, 16-19 June 1999 Beijing China, 47-48.

Zhu, Y., and Z. Toth, 2001: Extreme weather events and their probabilistic prediction by the NCEP ensemble forecast system. Preprints, Symposium on Precipitation Extremes: Prediction, Impact, and Responses, Albuquerque, NM. *Amer. Meteor. Soc.,* 82-85.

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

Zhu, Y. 2004: Probabilistic forecasts and evaluations based on a global ensemble prediction system. World Scientific Series on Meteorology of East Asia, Vol. 3 - *Observation, Theory, and Modeling of Atmospheric Variability*, 277-287.

Zhu, Y., Z. Toth, R. Wobus, Q. Liu, 2005: July 2005 Upgrade of the NCEP Global Ensemble Forecast System. *NCEP implementation briefing (Science decision).*

Zhu, Y. 2005: Calibration of QPF/PQPF forecast based on the NCEP global ensemble. San Diego, CA. *Amer. Meteor. Soc.,* J3.3.

Zhu, Y. 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Advance in Atmospheric Sciences,* Vol. 22, No. 6, 781-788

Zhu, Y. 2010: Prediction Science. *Transaction of Atmospheric Sciences.* Vol. 33, No. 3,

Table 1. NCEP's GEFS configuration

<table>
<tr><th></th><th>Initial uncertainty</th><th>Model uncertainly</th><th>Resolution</th><th>Forecast length</th><th>Ensemble members</th><th>Daily frequency</th></tr>
<tr><td>1992.12</td><td rowspan="6">BV</td><td rowspan="8">None</td><td>T62L18</td><td>12</td><td>2</td><td>00UTC</td></tr>
<tr><td>1994.3</td><td>T62L28</td><td rowspan="8">16</td><td>10(00UTC)
4(12UTC)</td><td rowspan="3">00, 12UTC</td></tr>
<tr><td>2000.6</td><td>T126L28(0-2.5)
T62L28(2.5-16)</td><td rowspan="4">10</td></tr>
<tr><td>2001.1</td><td>T126L28(0-3.5)
T62L28(3.5-16)</td></tr>
<tr><td>2004.3</td><td>T126L28(0-7.5)
T62L28(7.5-16)</td><td rowspan="5">00, 06, 12, 18UTC</td></tr>
<tr><td>2005.8</td><td rowspan="3">T126L28</td></tr>
<tr><td>2006.5</td><td rowspan="3">ETR</td><td>14</td></tr>
<tr><td>2007.3</td><td rowspan="2">20</td></tr>
<tr><td>2010.2</td><td>STTP</td><td>T190L28</td></tr>
</table>

Table 2. ECMWF's GEFS configuration

	Initial uncertainty	Model uncertainly	Resolution	Forecast length	Ensemble members	Daily frequency

<table>
<tr><td>1992.12</td><td rowspan="2">SVINI</td><td rowspan="3">None</td><td>T63L19</td><td rowspan="8">10</td><td>32</td><td rowspan="6">12UTC</td></tr>
<tr><td>1996.12</td><td rowspan="3">T_L159L31</td><td rowspan="13">50</td></tr>
<tr><td>1998.3</td><td rowspan="10">EVO-SVINI</td></tr>
<tr><td>1998.10</td><td rowspan="7">SPPT</td></tr>
<tr><td>1999.10</td><td>T_L159L40</td></tr>
<tr><td>2000.11</td><td rowspan="2">T_L255L40</td></tr>
<tr><td>2003.3</td><td rowspan="8">00, 12 UTC</td></tr>
<tr><td>2006.2</td><td>T_L399L62</td></tr>
<tr><td>2006.9</td><td rowspan="3">T_L399L62(0-10)/ T_L255L62(10-15)</td><td rowspan="6">15</td></tr>
<tr><td>2008.3</td></tr>
<tr><td>2009.9</td><td rowspan="3">Revised SPPT</td></tr>
<tr><td>2010.1</td><td rowspan="3">T_L639L62(0-10)/ T_L319L62(10-15)</td></tr>
<tr><td>2010.6</td><td rowspan="2">EDA-SVINI</td></tr>
<tr><td>2010.x</td><td>Revised SPPT-SPBS</td></tr>
</table>

Table 3. CMC's GEFS configuration

<table>
<tr><th></th><th>Initial uncertainty</th><th>Model uncertainly</th><th>Resolution</th><th>Forecast length</th><th>Ensemble members</th><th>Daily frequency</th></tr>
<tr><td>1998.2</td><td rowspan="3">PO</td><td rowspan="3">Multi-model and multi-parameterization</td><td>T_L95(SEF)</td><td rowspan="3">10</td><td>8</td><td rowspan="3">00UTC</td></tr>
<tr><td>1999.8</td><td>T_L95(SEF) 1.875°(GEM)</td><td rowspan="2">16</td></tr>
<tr><td>2001.6</td><td>T_L149(SEF) 1.2°(GEM)</td></tr>
<tr><td>2007.7</td><td>EnKF</td><td>Multi-parameterization and stochastic processes</td><td>0.9°(GEM)</td><td>16</td><td>20</td><td>00, 12UTC</td></tr>
</table>

Table 4. Characteristics of Lyapunov, Breeding and Singular vectors

Lyapunov vectors	Breeding vectors	Singular vectors
Linear perturbation evolution	Nonlinear perturbation evolution	Linear perturbation evolution
Fast growth	Fast growth	Fastest growth
Sustainable	Sustainable	Transitional (optimized)
Norm independent	Norm independent	Norm dependent
Spectrum of LLVs	Can orthogonal	Spectrum of SVs

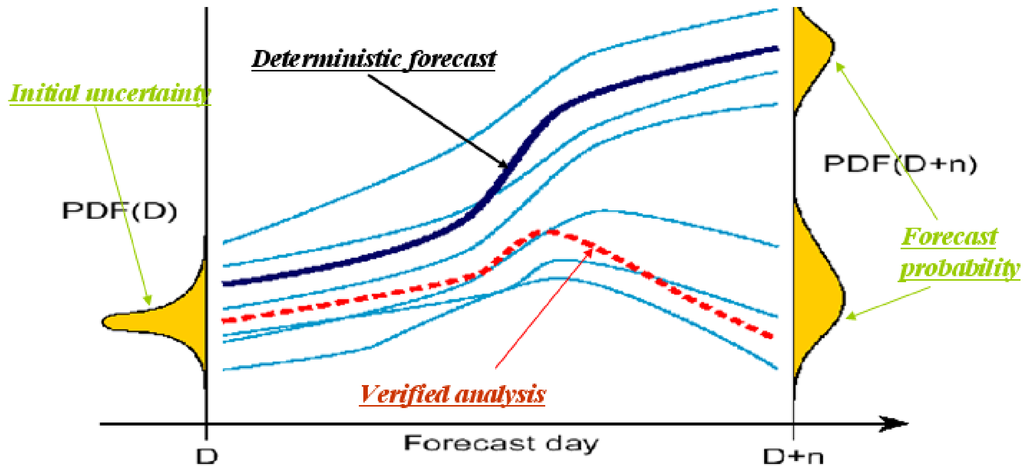

Figure 1. Schematic diagram of ensemble prediction (Buizza et al., 2001).

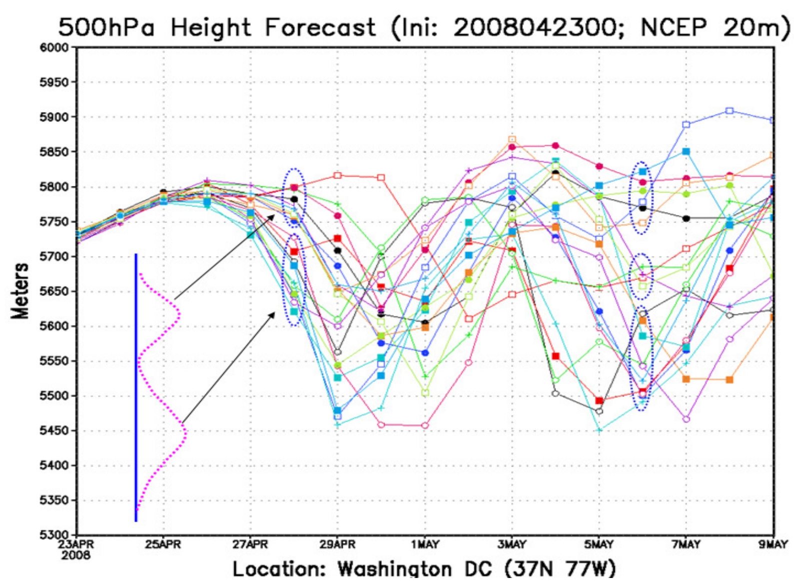

Figure 2. 20-members, 16-day Ensemble forecast from April 23rd 2008 demonstrates a chaotic atmosphere prediction which forms bimodality (at day-5), trimodality (at day-13) and uncertainties. It is for 500hPa geopotential height over Washington DC (37ºN, 77ºW).

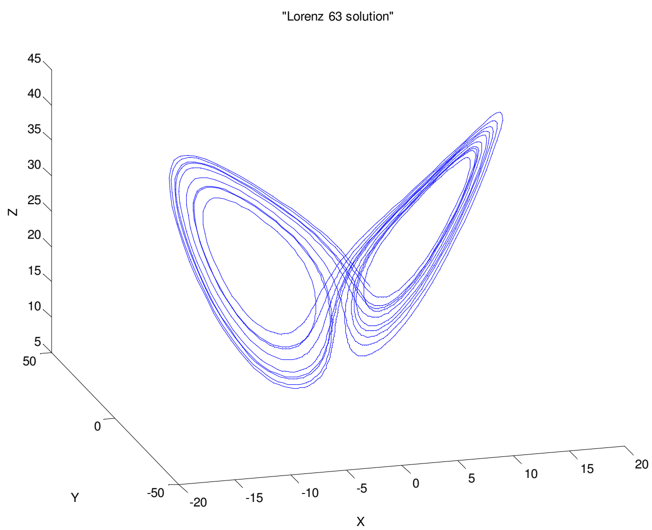

Figure 3. Numerical solution of Lorenz 63 model.

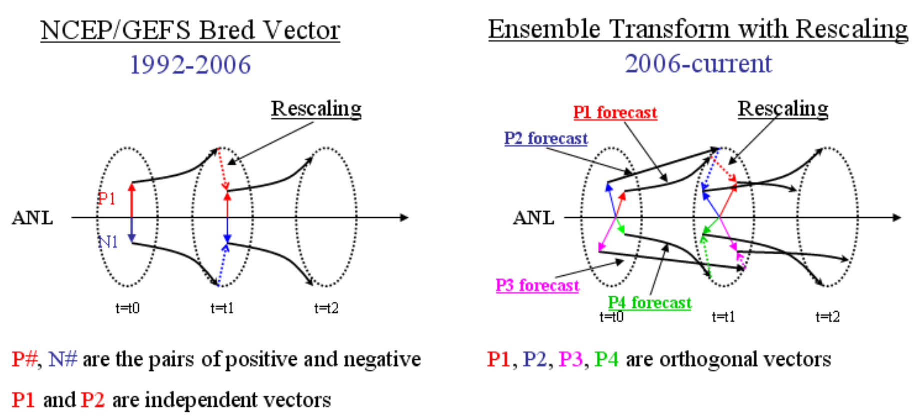

Figure 4. Schematic diagram of breed mode method (1992-2006) and ensemble transform with rescaling method (2006-current) of NCEP Global Ensemble Forecast System.

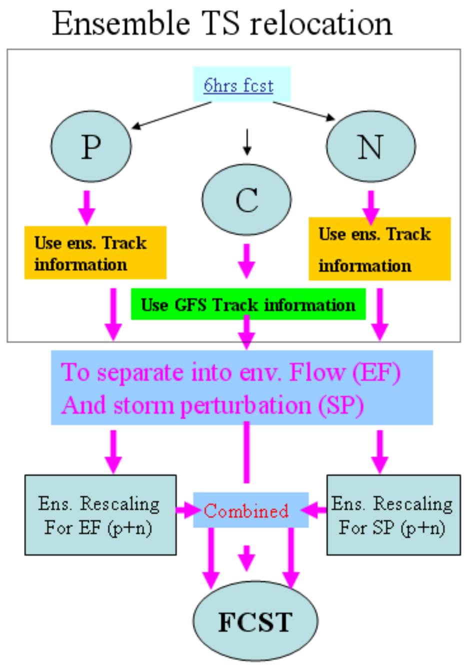

Figure 5. A flow chart of NCEP GEFS hurricane relocation process which was implemented in August 2005.

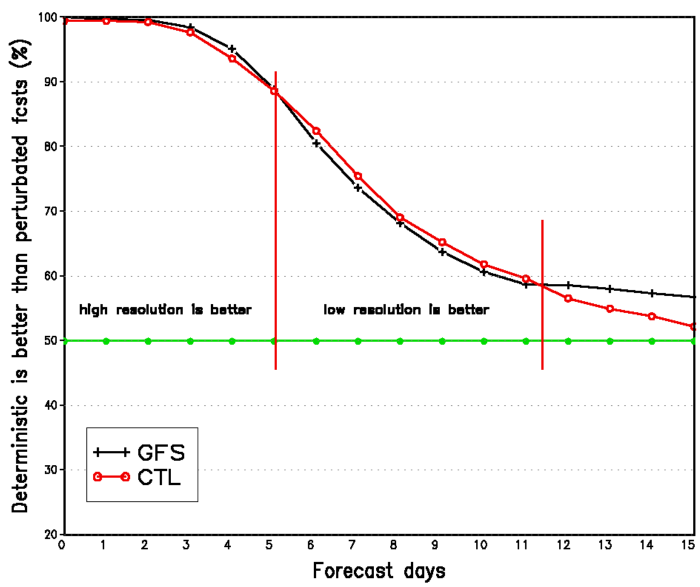

Figure 6. June 1st 2003 – May 31st 2004 (1-year) daily PAC scores for the NCEP/GFS (high resolution control, cross) and the NCEP ensemble control (the same resolution as ensemble members, open circle) are better than (or worse than) NCEP 10 individual ensemble members. The 50% line is a reference (closed circle) to represent GFS/CTL is in medium of ensemble members. Values (PAC scores) refer to the 500hPa geopotential high over northern hemisphere latitude band 20-80N.

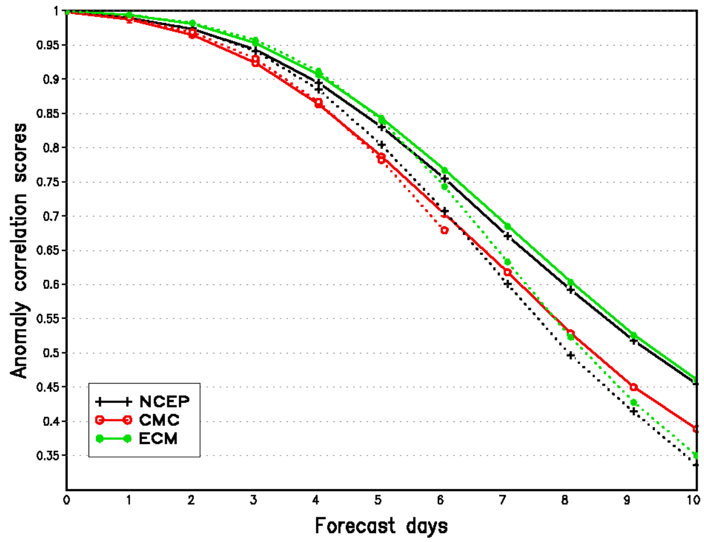

Figure 7. June-August 2004 average PAC scores for the control (dotted lines) and the ensemble mean (solid lines) of the NCEP-EPS (cross), CMC-EPS (open circle) and ECMWF-EPS (closed circle). Values refer to the 500hPa geopotential high over northern hemisphere latitude band 20-80N.

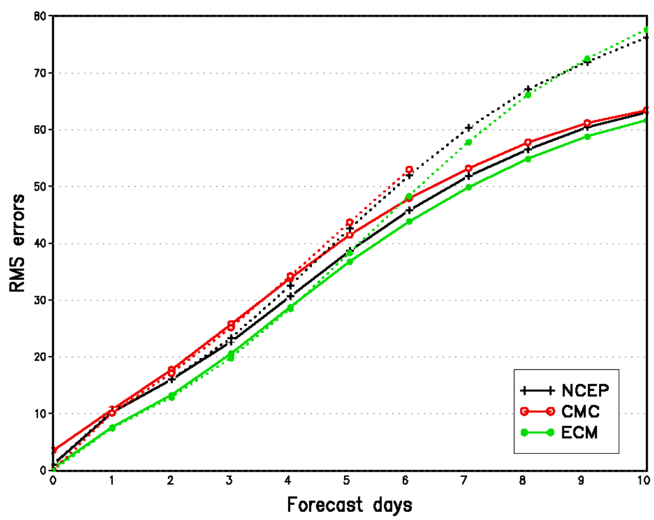

Figure 8. June-August 2004 average RMS errors for the control (dotted lines) and the ensemble mean (solid lines) of the NCEP-EPS (cross), CMC-EPS (open circle) and ECMWF-EPS (closed circle). Values refer to the 500hPa geopotential high over northern hemisphere latitude band 20-80N.

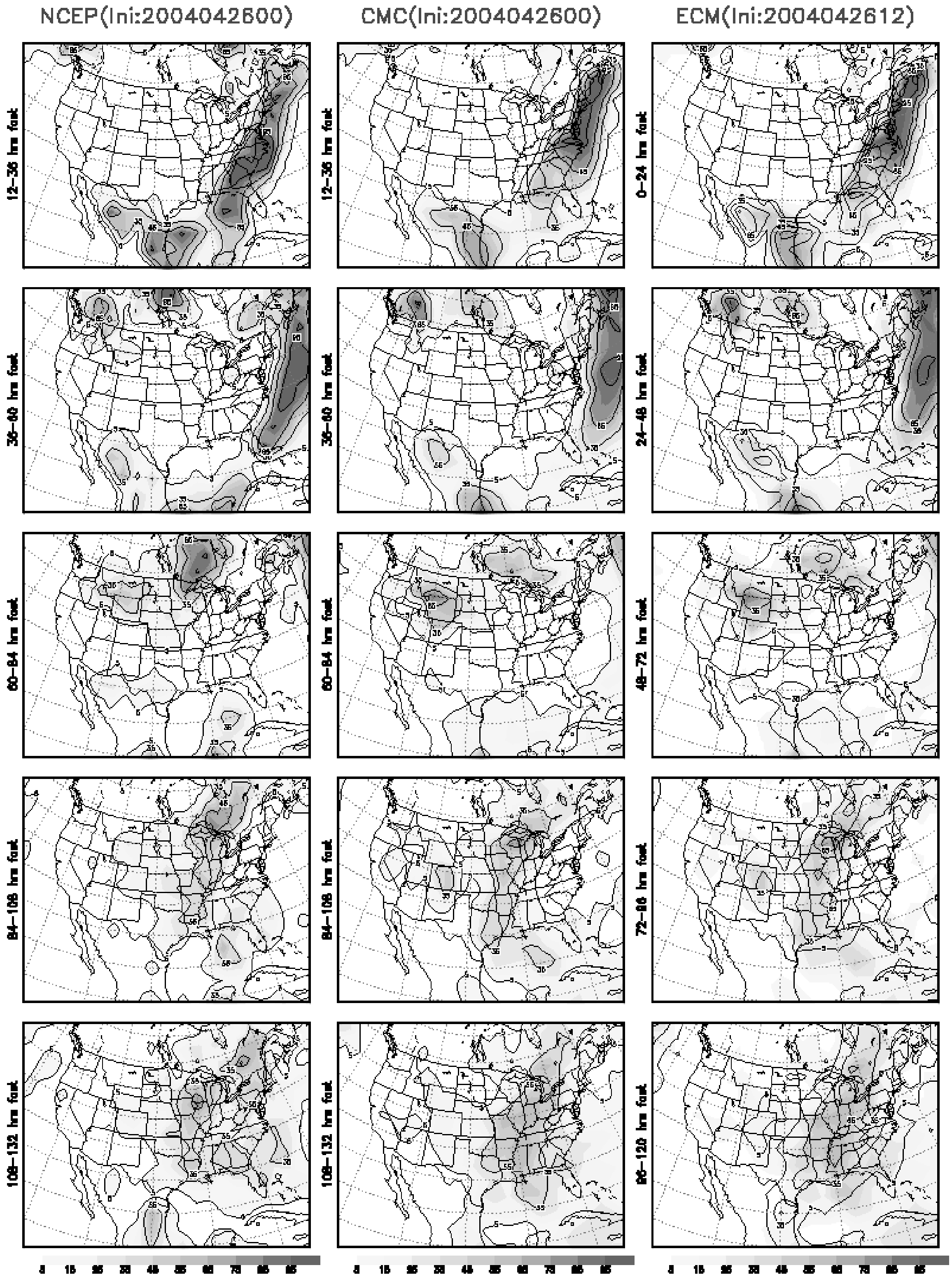

Figure 9. The probabilistic quantitative precipitation forecast (PQPF) for the 24 hours amount is exceeding 6.35mm (or 0.25 inch). The initial times are April 26 0000UTC for NCEP 11 ensemble members and CMC 17 ensemble members, April 26 1200UTC for ECMWF first 11 ensemble members. The gray scale bar indicates the probabilities from 0 to 100 in percentage.

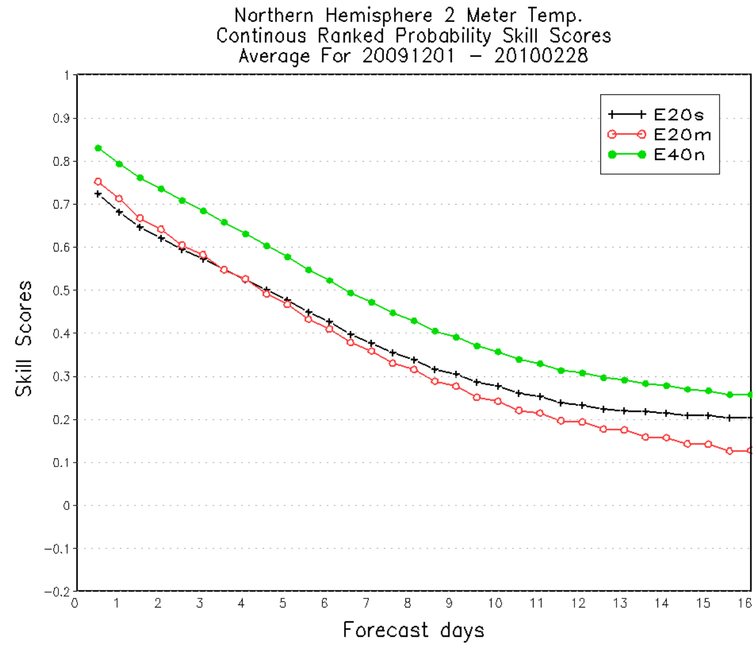

Figure 10. The average CRPS scores of theDecember 1st 2009 – February 28th 2010 for combined Products – NAEFS (E40n - closed circle), NCEP's raw ensemble forecast (E20s - cross) and CMC's raw ensemble forecast (E20m - open circle) . Values refer to the 2 meters temperature over northern hemisphere latitude band 20-80N.

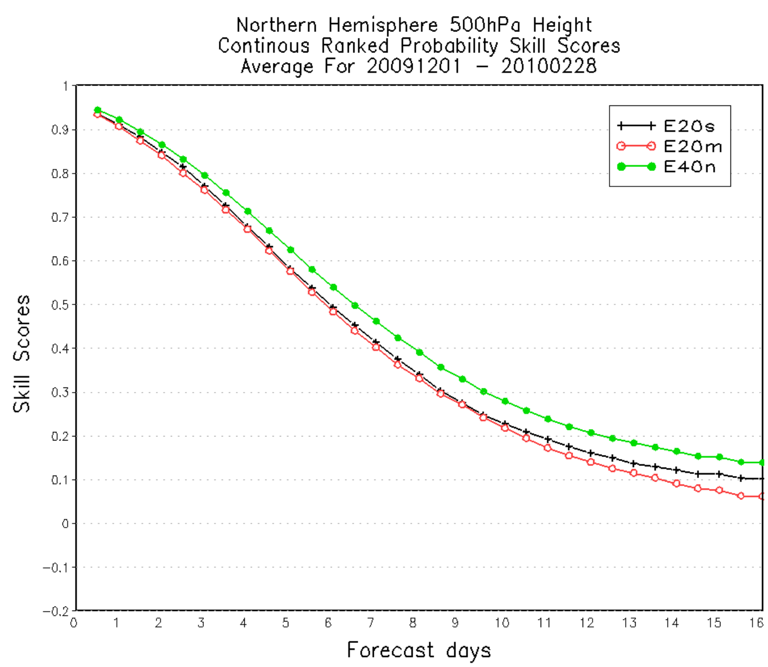

Figure 11. The average CRPS scores of theDecember 1st 2009 – February 28th 2010 for combined Products – NAEFS (E40n - closed circle), NCEP's raw ensemble forecast (E20s - cross) and CMC's raw ensemble forecast

(E20m - open circle) . Values refer to the 500hPa geopotential height over northern hemisphere latitude band 20-80N

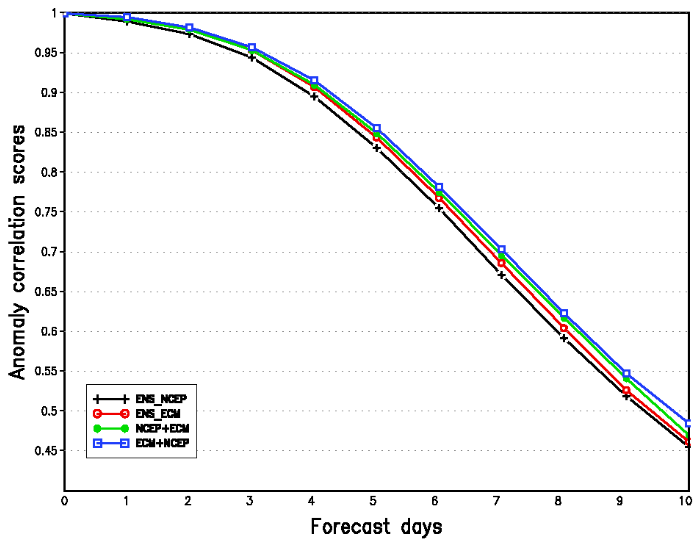

Figure 12. June-August 2004 average PAC scores for 10-member ensemble mean of NCEP (cross, the same as Fig. 2), ECMWF (open circle, the same as Fig. 2), NCEP (6 members) + ECMWF (4 members) (closed circle) and ECMWF (6 members) + NCEP (4 members) (open square). Values refer to the 500hPa geopotential high over northern hemisphere latitude band 20-80N.

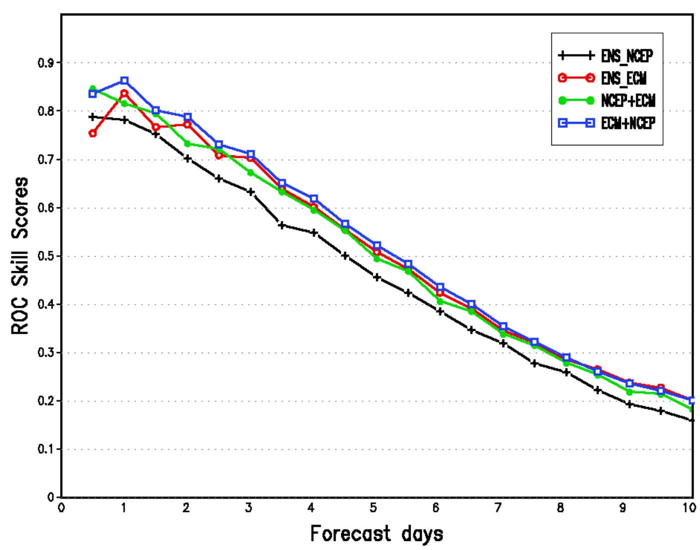

Figure 13. June-August 2004 average ROC skill scores for 10-member ensemble distribution of NCEP (cross), ECMWF (open circle), NCEP (6 members) + ECMWF (4 members) (closed circle) and ECMWF (6 members) + NCEP (4 members) (open square). Values refer to the 500hPa geopotential high over northern hemisphere latitude band 20-80N.

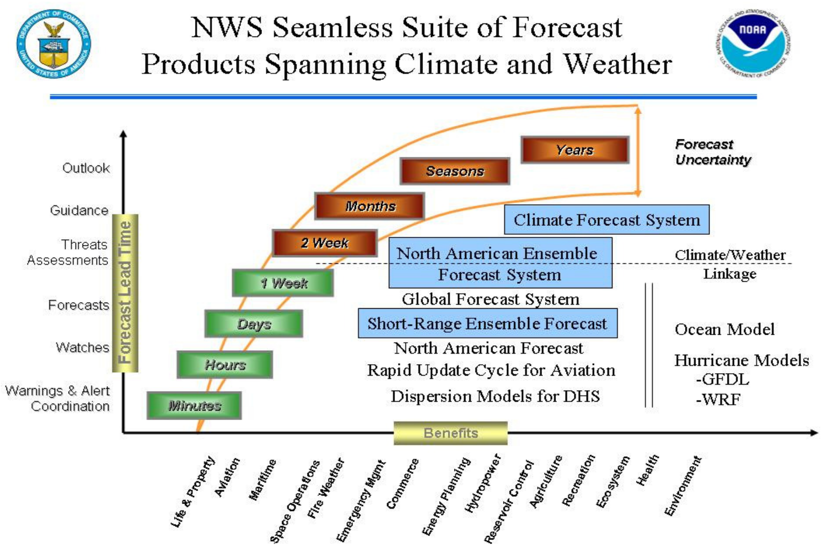

Figure 14. Schematic diagram of NWS seamless suite of forecast products spanning climate and weather