Uncertainty and Ensemble Forecast

Jun Du

SREF Development Team, EMC/NCEP/NOAA

This lecture series provides a comprehensive review and discusses some general principles on ensemble forecasting to give readers a big picture about what is involved in this relatively new and rapidly developing branch of numerical modeling and prediction.

Table of Content

Key Words: ensemble forecast, predictability, uncertainty, probability, deterministic, stochastic, evaluation, products, post-processing, decision making

---

Correspondence to: Jun Du, NOAA/NWS/NCEP/Environmental Modeling Center
E-mail: Jun.Du@noaa.gov

Introduction

In early days, weather forecast was indeed called probability and issued as probabilistic format. For example, Prof. Cleveland Abbe issued the first official "Weather Synopsis and Probabilities" on February 19, 1871 (NRC, 2006). Later on, due to the advancement and new knowledge achieved in the numerical weather prediction (NWP) as well as more observations available, scientists start to use single deterministic value to predict weather (still so today!).

After Lorenz discovered chaotic nature of atmospheric behavior/phenomena in 1960s, some pioneering scientists started to seriously reconsider stochastic approaches in predicting weather and climate. Given the fact that intrinsic uncertainties exist in each steps of a prediction process, we have no way to know the ground truth in an exact fashion. Instead, a complete and faithful description of, say, initial condition and model physics must be in a probabilistic distribution form, which is stochastic in nature within a certain range of uncertainty. As a result of this and the chaotic nature of highly nonlinear numerical models (Lorenz, 1993), there might be a multiple of possible realizations for each forecast. In other words, a complete forecast must also be described in a probabilistic distribution with forecast uncertainty explicitly expressed but not in a single deterministic value!

In 1969, Epstein (1969) first proposed a theoretical Stochastic-Dynamic approach to directly describe forecast error distributions (mean, variance and probability density function) in model equations. Unfortunately, it's unrealistic to integrate such a system with limited computing power since the number of forecast equations required to be solved is huge for a real atmospheric system. Instead, Leith (1972) proposed a more practical Monte-Carlo approach with limited forecast members. Each forecast member is initiated with randomly perturbed, slightly different initial condition (IC). He pointed out that with as few as eight members, the average of member could give a best estimation of a forecast with adequate accuracy although more members might be needed for forecast variance estimation. With an analytical turbulence equation, Leith showed that Monte-Carlo method is a practical approximation to Epstein's Stochastic-Dynamical approach. Leith's Monte-Carlo approach is basically the traditional definition of ensemble forecasting although the content of this definition has been greatly expanded in the last 20 years to include the following: (a) perturbing all uncertain components in a state-of-the-art forecasting system such as physics, numeric and boundary forcing besides perturbing atmospheric ICs (observation and analysis), and (b) flow-dependent IC perturbations with dynamically growing structure rather than statistical, random perturbation (see Part 3).

As computing power increases, operational ensemble forecasting became a reality in early 1990s. Both NCEP and ECMWF (European Center for Medium-Range Weather Forecast) operationally implemented its own global model-based, medium-range ensemble forecast system in December 1992, respectively (Tracton and Kalnay, 1993; Toth and Kalnay, 1993; Mureau et. al., 1993; Molteni et. al., 1996). At the same time, a few people realized that predictability issue is not only relevant to medium-range but also to short-range forecasts and therefore started to research on regional model-based short-range ensemble forecasting (Mullen and Baumhefner, 1994; Mullen and Du, 1994; Brooks et. al., 1995; Du et. al., 1997 and 2000; Mullen et. al., 1999). An operational Short-Range Ensemble Forecasting (SREF) system was under development and evaluation over the North American domain at NCEP since 1995 (Tracton and Du, 1998; Stendsrud et. al., 1999; Hamill and Colucci, 1997 and 1998; Hou et. al., 2001) and became a part of U.S. National Weather Service (NWS) real-time production suite in April 2001 (Du and Tracton, 2001) which is the first real-time operational regional ensemble system among major NWP centers in the world. A time-lagged ensemble forecasting approach was also operationally used for seasonal prediction (9 months) at NCEP from 2004 based on a global model coupled with ocean (Saha et. al., 2006).

Since the initial implementations of NCEP and ECMWF ensemble systems, ensemble approach has been widely accepted and actively pursued at almost all other major NWP centers around the globe such as Houtekamer et. al. (1996), Ebert (2001), Li and Chen (2002), Wang and Kann (2005), Eckel (2005), Chien et. al. (2006), Tennant et. al. (2007), Teixeira et. al. (2007) and Matsueda et. al. (2007). Research on ensemble forecasting also gained its strength since later 1990s and early 2000s and has merged as a hot topic in NWP nowadays (Buizza et. al., 1999a; Mullen and Buizza, 2001; Hansen, 2002; Grimit and Mass, 2002; Bright and Mullen, 2002a and 2002b; Hamill et. al., 2000 and 2004a; Legg and Mylne, 2004; Wandishin et. al., 2005; Eckel and Mass, 2005; Jones et. al., 2007; Yuan et. al., 2005 and 2007c; Jankov et. al., 2007). It's expected that the ensemble-based probabilistic forecasting will play more and more important role in shaping the future of numerical weather prediction practice and service in years to come.

For technical details, related references are provided at the last part (Part 11), so that interested researchers could study further in depth and join the active research community of this challenging frontier. First, the underline scientific reason why ensemble forecasting is needed was discussed in Part 1. The following parts discussed various aspects related to ensemble forecasting including what ensemble forecasting is aiming for (Part 2), how to build an ensemble prediction system (EPS, Part 3.1, 3.2, and 3.3), what products can be derived from an ensemble (Part 4), what is the role of EPS post-processing (Part 5), how to evaluate the quality of an EPS and its forecasts (Part 6). Due to its increasing importance, how to communicate forecast uncertainty and how to use probability information in users' decision-making process were illustrated in Part 7. Recent development of downstream applications using meteorological ensembles as inputs was also mentioned in Part 8. Part 9 listed some major differences between the two forecast paradigms -- "single forecast" vs. "ensemble forecasts". Part 10 mentioned possible future trend of ensemble-related development. Finally, a summary and references are given at Part 11.

1. Why is ensemble forecast needed?

The ultimate goal of science is to predict future. A prediction process has four basic components: data collection (observation), assimilation of observed data into initial conditions to be used by a numerical forecasting model, model integration to project the initial state into future, and the application of the forecasts to real world situations. Intrinsic uncertainties are introduced at each of those steps during a forecast process, for example, instrumental and human error introduced during the process of collecting data; errors introduced during data assimilation process due to mathematical assumptions; imperfect model physics (approximation of real world such as parameterization of sub-grid effects) and numeric (e.g., discontinuity or truncation); and differences in human (both forecasters and end-users)'s interpretation and decision to a same forecast depending on situations (who, what, when and where). All these kind of errors are intrinsic, unavoidable and sometimes even unknown to us in a real world operation.

Due to its highly nonlinear nature, a numerical prediction model of weather, climate and water is chaotic i.e. a tiny difference in initial states could possibly be amplified into significantly large difference in future states in unstable condition (Lorenz, 1963, 1965, 1993; Thompson, 1957). The difference could be as large as that being randomly picked from model climatology. Therefore, such forecasts become meaningless, which indicates that prediction of weather, climate and water events has uncertainty and limit (predictability). Figures 1-2 depict two such examples of many from the U.S.

National Weather Service's daily real-time operational model guidance. Figure 1 shows that two NCEP (National Centers for Environmental Prediction) operational GFS (Global Forecasting System) model's medium-range (16 days) forecasts which were initialized at only 6-hour apart in time predicting two completely opposite large-scale flow at 500 hpa level: one places a strong trough over the East Coast and another over the West Coast of U.S. (more than 3000 km apart)! Similar situation happens at short range too: two NCEP operational regional Eta model's short-range (2.5 days) forecasts which had slight difference only in their initial atmospheric conditions predicted two distinct scenarios: one predicts a deep low pressure system (<976 hpa) while another a high pressure ridge (around 1008 hpa) over a same area (Fig. 2). Situations like these occur not uncommon in real-time operation especially during major high-impact weather events which are often associated with highly unstable atmospheric conditions.

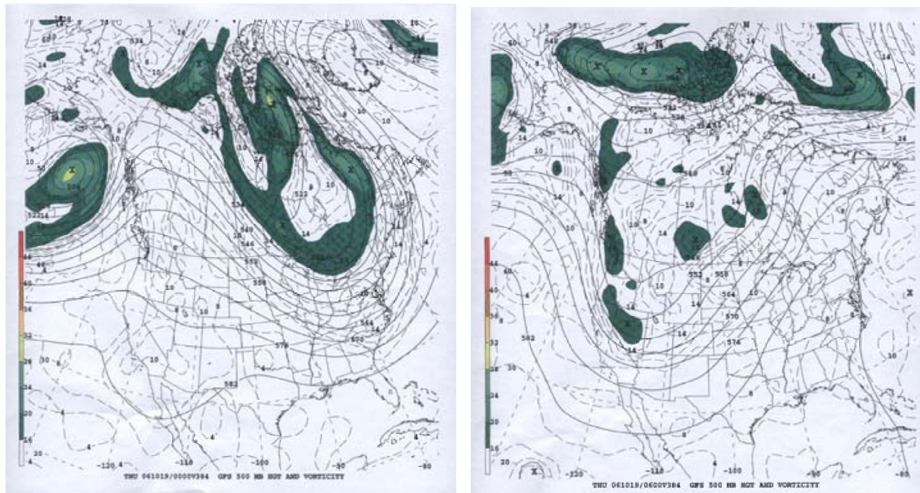

Figure 1 Two consecutive NCEP Global Forecasting System (GFS) 16-day 500-hpa HGT/VORT forecasts with 6-hour apart in initiation time.

Therefore, uncertainty and predictability is a very real and important aspect of a forecast such as weather forecasting. Besides the prediction to an event itself, the uncertainty and predictability of the event also needs to be predicted, i.e. how small initial differences (uncertainties) evolve with time in a model. Without uncertainty quantified, a forecast is incomplete. Ensemble forecasting is a dynamical and flow-dependent approach to quantify such forecast uncertainty (error of the day) and provides a basis to communicate forecast uncertainty and forecast confidence to end-users who can then be best prepared. If reader is interested in observing how small initial differences evolving under various weather situations in a real-time ensemble forecasting system, one could go to the NCEP Short-Range Ensemble Forecasting (SREF) system web page as an example: http://www.emc.ncep.noaa.gov/mmb/SREF/SREF.html.

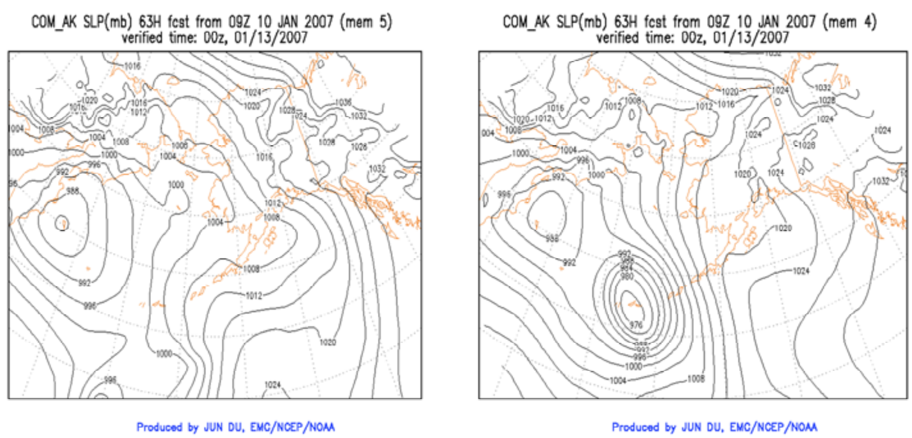

Figure 2 Fig. 2: Two NCEP SREF Eta-member 63-h sea-level pressure forecasts initiated with slightly different initial conditions.

2. What is ensemble forecasting aiming for?

In the current deterministic NWP practice, one wishes to use single model output (Xm) to represent true atmospheric state (X), i.e.

$$X=Xm \qquad (1)$$

Since IC state and model state used in a model actually represent a kind of mean state in a certain degree, a corresponding model forecast (Xm) is, therefore, depicting a mean state too and the equation (1) is never the case in reality. Instead,

$$X=Xm+x0 \qquad (2)$$

is always observed, where x0 is a departure of model forecast from truth. Since exact value of x0 is something we really don't know in prior, we hope to estimate a possible set of solutions Xm+X' to include the truth X or an error distribution X' to have x0 be within the estimated distribution via a certain approach, i.e.

$$X \in Xm+X' \qquad (3)$$

Since the intrinsic unavoidable uncertainties introduced at each of forecast steps are probably random in nature, the error distribution X' must not be a single value but a kind of probabilistic distribution associated with initial uncertainties, instability of flow and the degree of nonlinearity of a modeling system. To reliably estimate and accurately describe this flow-dependent error distribution or forecast uncertainty range X' to have truth be encompassed within Xm+X' is the primary Mission of Ensemble Forecasting. Here, to improve the capability of predicting the error distribution X' needs to improve ensembling technique and strategy, while to improve the accuracy of predicting mean state Xm mainly depends on model and IC qualities. In other words, the ensemble technique is dealing with the random error of a forecast, while the model and IC are dealing with the systematic error of a forecast. A good model and a high quality IC are the basis of ensemble forecasting. Therefore, improving model, IC quality and ensembling technique should be viewed as a whole to advance NWP. By the definition of ensemble mission, one can imagine that ensemble forecasting is most valuable when large forecast uncertainty is around and forecasters don't know what solution to choose from in mainly high-impact events and has minimal value when weather is quiescent and highly predictable (although one still needs ensemble to identify such occasions).

It might be worth pointing out that although ensembling method is gaining popularity in research and operation nowadays, a commonly seen incomplete use or even misunderstanding of the technique is that ensemble is merely used as a tool to improve the accuracy of a single value forecast such as by ensemble averaging all members or constructing a performance-based consensus forecast while the ensemble spread or forecast variance are purposely or mistakenly regarded as meaningless noises to be filtered out. We often heard people are saying that "forecast error can be reduced such and such by using ensemble data". Indeed, due to the nonlinear filtering process, an ensemble mean forecast is statistically or on average more robust and accurate than a single forecast and then an improved forecast, but ensembling technique is not only a tool to improve a single deterministic forecast but more importantly to quantify forecast uncertainty which is the ultimate goal and the core mission of ensemble forecasting as discussed above. Ensemble averaging or other approaches to construct a consensus forecast are only one of the three possible ensemble-based product types: consensus or a most probable solution (mean is the simplest one), spread or forecast variance/confidence, and probability or a distribution (see Part 4).

After this forecast uncertainty issue was realized, the earlier attempt was using statistical approaches such as MOS (Model Output Statistics, Glahn and Lowry, 1972) to address the issue. For example, based on model performance over a long-time period in the past, statistical characteristic of forecast error is obtained for a particular model and the error distribution can then be applied to the model forecast to estimate a probabilistic forecast such as Probability of Precipitation (PoP) (the "Eq. 3" thinking by estimating X'); or, some linear regression equations can statistically be established by either using model outputs (MOS) or observations (Perfect Prog, PP) as predictors to have a best estimate of a variable one wishes to forecast (the "Eq. 2" approach by estimating x0). Apparently, MOS-, POP- and PP-like approach is an important positive development in NWP history to address or reduce forecast uncertainty. However, an intrinsic limitation of any statistical approach is that the estimated error characteristic or distribution represents only the historical performance of a model in general such as model systematic bias but not the error related to the current flow situation (so called "error of the day"). Statistical method also heavily depends on the length of historical data and suffers when model frequently changes (a situation happens all the time in reality). In contrast, ensemble forecasting is a dynamical approach to capture the flow-dependent error of the day since it's derived directly from the current model forecasts of the same day but not from past forecasts and also automatically improves as soon as a base model is improved. Surely, it will be desirable to combine both statistical and dynamical approaches together to portray a best picture of future true state (see Part 5).

3. How to build an ensemble prediction system (EPS)?

How to estimate the uncertainty distribution X' of Eq. (2) in ensemble forecasting? The principle is to describe all possible uncertainty sources in a modeling system as accurately and completely as possible and then to incorporate all those uncertainties (perturbation terms) into the numerical model to be integrated in time to produce a finite size ensemble of forecasts. As an approximation to a theoretically infinite ensemble, the finite size ensemble is then used as the basis to estimate probabilistic distribution and uncertainty or confidence of a forecast. Below is a brief summary of currently existing approaches. Based on the approaches used, EPS could be classified into three general categories: 1-Dimensional, 2-Dimensional and 3-Dimensional systems.

3.1 1-Dimentional EPS

Only IC uncertainty is considered in 1-D EPS by perturbing initial conditions. Three basic properties need to be followed in designing a perturbation: Realism, Divergence and Orthogonality. Realism is that the magnitude of perturbation needs to be within the size of realistic analysis error and should exhibit a realistic spectral distribution over spatial scales such as larger uncertainty in smaller-scale waves (difficult to observe) and smaller uncertainty in larger-scale waves (easier to observe). Divergence is that the perturbation needs to have dynamically growing structure leading members to diverge as much as possible during model integration to cover all possible solutions in a model space. Orthogonality is that the perturbation needs to be orthogonal to each other among members to maximize information content contained in an ensemble, which is especially important for a small size ensemble. Initial conditions or states needed to be perturbed include interior, lower-, upper- and lateral-boundary (if limited-area model) conditions. There are currently five or so different categories proposed for perturbing ICs especially for interior states.

(1) Random Perturbation (Monte Carlo approach): Perturbation is randomly generated based on some kind of statistics (usually a normal distribution) representing typical uncertainty in analysis (such as the average difference between two analyses over a long period of time) (Errico and Baumhefner,

1987; Mullen and Baumhefner, 1994; Du et. al., 1997). Although this type of perturbation represents well the average magnitude of analysis uncertainty (Realism), it is lack of dynamically growing spatial structure and not reflecting "error of the day". As a result, the perturbation growth rate is low and, therefore, the Divergence in solution among members is usually not ideal. Random Perturbation is usually applied in places where one is unsure which other methods work better.

(2) Time-Lagged approach: There are "Direct Time-Lagged" and "Scaled Time-Lagged" two kinds. Direct Time-Lagged approach (Hoffman and Kalnay, 1983) directly pulls multiple forecasts which are initiated from different past times but verified at a same time together as an ensemble (a mixture of old and young forecasts). This method views the error of a past forecast at t=0 (current initial time) directly as IC perturbation which should reflect "error of the day" and has dynamically growing structure leading to larger ensemble spread than random perturbation. The advantage of the method is that the generation of perturbation is absolutely free and there is no need to purposely generate IC perturbation fields for the ensemble, which implies that all operational NWP centers have this type of ensemble automatically. However, a main concern is that the quality (magnitude) of perturbation depends on the age of a forecast since forecast quality usually decreases with lead time. To avoid this weakness, past forecast errors are first scaled down by their "ages" (assuming error growth is quasi-linear) at t=0 to have similar magnitude in all perturbations and then added to or subtracted from the current control analysis to create multiple analyses to initiate an ensemble of forecasts (Ebisusaki and Kalnay, 1983; Kalnay, 2003). This modified version is called Scaled Time-Lagged method and can be simply described by the following equation:

Initial perturbation = scaling x (time-lagged forecast – current analysis) (4)

The latter is not only able to control perturbation size but also doubles the ensemble membership by using both addition and subtraction procedures with very little extra computing cost (which is very important in real-time production) in generating perturbation. Note that this Scaled Time-Lagged method is actually already the same in idea and similar in technical procedure to the Breeding method (see the next paragraph). The Time-Lagged approach has been used in many ensemble research and operation such as NCEP operational seasonal ensemble forecast system (Saha et. al., 2006; Hou et. al., 2001; Lu et. al., 2006; Brankovic et. al., 2006; Mittermaier, 2007). A limitation of the Time-Lagged method is that it cannot create an ensemble with large enough membership size since the number of "good" old forecasts available is limited in reality. Otherwise, the ensemble quality will be severely contaminated if too old forecasts are included to have a large size ensemble. By the way, Time-Lagged forecasts are often used as "initial seeds" to cold start an ensemble which uses other perturbation methods such as Breeding.

(3) Breeding: Different from Scaled Time-Lagged method, Breeding uses two concurrent forecasts (at a recent past time t=-T) rather than a time-lagged forecast and a current analysis to calculate raw perturbation at t=0. The difference is then be scaled down and added to or subtracted from the current control IC (Toth and Kalnay, 1993 and 1997). In this way, one can create as many members as he wants to have a large size ensemble as long as he can create enough initial random seeds or use other approaches (such as borrowing from other existing forecasts like time-lagged ones) in the beginning to initiate or have enough different forecasts to start with (cold start). So, Breeding method can overcome the membership limitation of Scaled Time-Lagged method. Since all the past forecasts used are now concurrent, the scaling factor has no need to be forecast-age based but can be anything related to analysis uncertainty to control perturbation magnitude and/or spatial structure. The perturbation can now be simply described by the equation (5).

Initial perturbation = scaling x (forecast 1 – forecast 2) (5)

Comparing Eqs. (4) and (5), one can also see that the perturbation in Breeding (bred vector) is no longer pure forecast error but the difference between two past forecasts which is a nonlinear extension of Lyapunov vector (Kalnay, 2003). Experience (e.g., NCEP SREF) shows that bred vector becomes mature in structure and leads to large ensemble spread growth after cycling for about two to three days after the cold start. Toth and Kalnay pointed out that the spatial structure of a mature bred vector is not sensitive to scaling period (T) and norm selected and that bred vector reflects analysis error (error of the day) introduced during data assimilation cycle well (Realism). Although the difference between two past forecasts should, theoretically, reflect the error growing structure of the immediate past cycle but not of the forecast period in future (i.e., a looking-backward approach), experiments show that bred vector's growth rate (Divergence) is quite satisfactory in practice and higher than that by using either Monte Carlo or Scaled Time-Lagged methods (Toth and Kalnay, 1993 and 1997). Since this method is simple with no mathematical simplification or assumption (but using full nonlinear primitive-equation based model) and easy to implement, has little cost in computing power and gives good ensemble spread, it's widely used and tested by many such as NCEP ensemble systems (Du and Tracton, 2001; Tracton and Kalnay, 1993) and CMA (China Meteorological Administration). However, it's reported that bred vectors among ensemble members are not orthogonal enough but highly correlated to each other, resulting less optimal in information content contained in an ensemble (Wang and Bishop, 2003; Martin et. al., 2007). As a result of this, the ensemble spread growth (mainly in magnitude but not structure) is closely related to initial amplitude of bred vector. To orthogonize bred vectors, Ensemble Transform (ET) technique is used to make bred vectors more orthogonal to each other by applying a simplex transformation matrix to transform forecast-based perturbations to analysis perturbations (Wei et. al., 2007). Experiment shows that ET technique can improve ensemble performance over classical Breeding method. Therefore, ET has been implemented in the NCEP global ensemble system to enhance the Breeding method (Wei et. al., 2007). Another approach proposed to improve classical Breeding method is called Geometric Breeding which controls spatial correlation of bred vectors among members to make them less correlated to each other (Martin et. al., 2007). Geometric Breeding shows better spread-skill relation than classic Breeding. Since bred vector mainly depicts synoptic scale baroclinic instability but not smaller-scale convective instability (Toth and Kalnay 1993), it is, however, desired to have smaller-scale instabilities included in perturbation for a mesoscale EPS focusing on predicting, say, convective system related heavy precipitation events. Chen et. al. (2003) suggested that differencing two forecasts from a same model but with different versions of a convective scheme (instead of one same version as in classical Breeding) will help to depict convective instability in perturbation and, therefore, enhances ensemble performance in predicting heavy precipitation. On other hand, due to the characteristic that fast, small scale modes quickly saturated, and only slow, large scale modes left during breeding cycle, bred vector is a good candidate to be used in ocean-atmosphere coupled ensemble prediction system which is mainly associated with slow modes (Cai et. al., 2002, Yang et. al. 2006). Recently, Prof. Eugenia Kalnay (2007) reported that bred vector has capability of predicting weather regime transition.

(4) Singular Vector (SV): This method first needs to develop a linear version of a nonlinear model (called Tangent Linear Model, TLM) as well as an Adjoint (Errico, 1997) of the TLM. Then over a desired optimal future time period such as 0-48h, TLM is integrated forward in time and then the Adjoint integrated backward in time to find initial sensitive areas to the forecast, say, at t=48h. This "forward and backward" cycling process needs to be iterated many times to obtain leading singular vectors. Then, a linear combination including rescaling and orthogonal rotating is applied to the vectors to construct a desired number of perturbations. With addition and subtraction of the perturbations to and from a control analysis, an ensemble of forecasts can be formed. Unlike the bred vector, the structure of SV is sensitive to the norm used and the cycling period selected (Errico and

Vukiceric, 1992; Palmer et. al., 1998). ECMWF chooses total energy as norm and 0-48 h cycling window to calculate SVs in their global ensemble system (Buizza, 1994; Palmer et. al., 1998). Obviously, SV is a looking-forward approach rather than looking-backward in time like Breeding does and mathematically optimizes large perturbation growth and orthogonarity to have large ensemble spread and contain more information at a pre-selected targeted forecast time. SV method is widely used and tested in both research and operation such as ECMWF and Canadian Meteorological Service's regional ensemble (Li et. al., 2007). A disadvantage of the method is that it is costly in computing because the number of the iteration of this "forward and backward" integration required is usually about 3 times the number of SVs you want to create (e.g., it needs to integrate about 3x50x2=300 times of 48h-forecast to obtain 50 SVs optimizing at 48h lead time. Therefore, SVs have to be calculated in a much reduced model resolution (comparing to actual forecast model resolution) to save computing time in operation. Another disadvantage is that a finite forecast lead time has to be pre-defined at which SVs are targeted to be optimal in growth. So that SV-based ensemble might not be optimal in performance crossing multiple time ranges such as short and median range at the same time. The linear assumption that perturbation is small enough so that its evolution can be governed by the linear versions (TLM and Adjoint) of a nonlinear model is also a concern in calculating traditional SVs although it retains nonlinear property in some degree by calculating and combining multiple SVs. To overcome the linear assumption, some efforts have been made such as modifying the iteration procedure (Oortwin and Barkmeijer, 1995; Barkmeijer, 1996) and introducing the nonlinear singular vector concept (Mu, 2000) and the conditional nonlinear optimal perturbation (CNOP) method (Mu et. al., 2003; Mu and Duan, 2003; Mu and Zhang, 2006). With simple models, CNOP method showed improved capability of depicting nonlinear features in perturbation comparing to SV method although further studies are needed with full NMP models. Adding moist physics in TLM and Adjoint (Ehrendorfer et. al., 1999) is another step to be closer to reality and showed improved performance. Since SV is mathematically looking forward and focusing on perturbation growing structure at a future time but not related to any immediate past, a reasonable question to ask is that does SV-based perturbation actually reflect error of the day which is introduced by the most recent data assimilation cycle? The following efforts are addressing this kind of concern and yield improved results (Barkmeijer et. al., 1998; Fischer et. al., 1998): e.g, evolving-SV approach by adding the final or evolved singular vector from an immediate previous cycle, say, 48h-period prior to the current model initiation time (analysis time) to the current cycle's SV perturbation; using analysis error covariance in replacing total energy as norm to calculate SVs; and the application of Kalman filter and so on. It's interesting to notice that the resulting SV perturbation is closer to Lyapunov vector or Bred vector by either switching to evolving-SV or using analysis error covariance as norm (Kalnay, 2003; Reynolds and Errico, 1999).

(5) Coupling with Data Assimilation: The simplest version of this method is directly using multiple analyses available to initiate an ensemble of forecasts (Tracton et. al., 1998; Grimit and Mass, 2002). However, the number of available analyses is quite limited, which will restrict ensemble size. By perturbing observations, Houtekamer et. al. (1996) and Houtekamer and Mitchell (1998) purposely generated multiple analyses to initiate their global ensemble system, which shed lights to a new IC perturbation generating approach for ensemble so called Ensemble Transform Kalman Filter (ETKF) approach (Anderson, 1996). The ETKF approach was further explored in detail by Wang and Bishop (2003), Wang et. al. (2004) and Wei et. al. (2006) for generating ensemble. In their studies, ETKF transforms forecast perturbations into analysis perturbations by multiplying a transformation matrix. Using observational information, the magnitude of the analysis perturbations was adjusted before the perturbations are added to a control analysis to initiate an ensemble of forecasts. The transformation matrix used can also guarantee all perturbations to be orthogonal to each other, a desired property for

ensemble forecasting. Although ETKF was not used in the data assimilation procedure to directly output multiple analyses in their studies, ETKF itself is an ensemble-based data assimilation technique (Tippett et. al., 2003; Anderson, 2001; Whitaker and Hamill, 2002; Ott et. al., 2004; Szunyogh et. al., 2004; Hamill, 2006; Zhang et. al., 2004; Wang et. al., 2007). Therefore, ETKF technique has a potential to directly link ensemble prediction and data assimilation (DA) into one unified procedure in an NWP system: ensemble forecast variance provides background error covariance information for DA, while DA provides an ensemble of analyses to initiate an ensemble of forecasts. In such a coupled system, not only EPS can be improved by having more realistic IC perturbations reflecting true error of the day in the analysis, but also the quality of analysis improved by using flow-dependent background error information (Hamill, 2006; Zhang, 2005). Therefore, it's believed that ETKF method has a great potential. This method has been explored to be applied to operation such as regional ensemble systems at UK Met-Office (Mylne, personal communication) and U.S. Navy and Air Force (McLay et. al., 2007). Active research is, however, still going on in this area. For example, it might need a large ensemble size to provide adequate error covariance information for DA, which is very costly in production. A dual-resolution idea has, therefore, been proposed to run such an ETKF-based system to significantly reduce computing cost (Gao and Xue, 2007; Gao et. al., 2007).

It is worth pointing out that all the currently existing IC perturbation generating methods do not work well and yield very little ensemble spread in tropics. This is because the current methods are mainly dealing with slower and larger-scale baroclinic instability which dominates middle- and high-latitude atmospheric motion but not faster and smaller-scale barotropic and convective instabilities which dominate tropical atmospheric motion. Therefore, special consideration needs to be researched for tropics. A simple comparison among various methods can be found in Bowler (2006).

Besides interior initial states, the content of initial condition should also include lower-, upper- and lateral-boundary forcing. The lower boundary forcing is introduced by land and water surface initial parameters such as sea surface temperature, heat and moisture flux, ice and snow cover, soil properties including moisture, temperature and type, surface albedo, roughness and greenness etc.. Out of those, sensitivity of initial soil moisture uncertainty to ensemble prediction in short-range (0-3 days) has been paid particular attention so far. It is found that soil moisture uncertainty plays an important role in convective precipitation during warm season but less important in large-scale precipitation during cool season (Sutton et. al., 2006; Aligo et. al., 2007; Du et. al., 2007a). The sensitivity of some surface variables such as 2-meter temperature to initial soil moisture remains high in both warm and cool seasons although such a sensitivity demonstrates strong diurnal variation related to radiation (much stronger during daytime than nighttime) and geographically preferred regions in particular forecasts (Du et. al., 2007a). Du et. al., 2007a) also reported that the effectiveness of soil moisture perturbation to ensemble forecast depends on perturbation's spatial structure and magnitude: spatially uniform and larger magnitude perturbations produce larger ensemble spread than spatially random and smaller magnitude ones do. In general, comparing with the sensitivity to atmospheric initial condition uncertainty, the sensitivity to land-surface initial parameter uncertainty seems to be secondary for short-range forecasts, which, of course, could be just opposite for longer-range seasonal forecasts which could be more sensitive to lower-boundary forcing uncertainty. Therefore, it's recommended that perturbation in surface forcing needs to be combined with other perturbations (such as IC and physics) in order to build a robust ensemble system. As for upper boundary forcing from space such as solar activity or how the upper boundary being treated in a model, not much attention is yet paid to how sensitive a weather forecast will be or how much ensemble spread could be attributed to due to the uncertainty in its initial description. It needs to be studied quantitatively. For a regional ensemble system, lateral boundary condition (LBC) could play a dominant role in defining ensemble spread of many variables (except for precipitation) if model domain is small (Du and Tracton, 1999; Warner et. al., 1997). Therefore, it is recommended that (a) a large enough model domain should be used in limited-area model (LAM) based ensemble system to avoid negative impact from LBCs; and (b) LBCs need to be perturbed too to ensure diverse ensemble solutions. Currently, a common approach in practice is to use different ensemble members from an available global ensemble system as LBCs for different members in a LAM-EPS such as NCEP SREF (Du et. al., 2004). Nutter et. al. (2004a and 2004b) suggested an approach to compensate spread loss due to LBC. How does the inconsistency or consistency in structure between LBC and internal IC perturbation affect ensemble performance is an issue yet needs to be investigated. It's reported that Spanish Weather Service (INM) has built a LAM-EPS purely based on diversity in LBCs (Garcia-Moya, 2006, personal communication).

3.2 2-Dimentional EPS

Besides IC uncertainty being considered, the uncertainty in model physics and dynamics is also taken into account in a 2-D EPS. There are currently many approaches used on this regard such as multi-model, multi-physics, multi-dynamics, multi-ensemble system and stochastic physics. Based on favorable research results such as Mullen et. al. (1999) and Tracton et. al. (1998), NCEP pioneeringly implemented a "multi-ensemble system" approach-based SREF in operation consisting of two sub-ensemble systems where each of them was based on a different regional model from the very beginning of its development (Du and Tracton, 2001). Currently, NCEP SREF consists of four sub-ensemble systems with four regional models (Du et. al., 2006). Obviously, multi-ensemble system approach is a grand mixture of multi-model, multi-dynamics, multi-physics, multi-IC and multi-LBC etc. methods. Multi-model ensemble system is considered to be an ad hoc approach but has been proven to be very effective and work very well (in both reducing the error of ensemble mean forecast and increasing ensemble spread) in practice (Du et. al., 2003; Mylne et. al., 2002). The simplest version of multi-model ensemble is the so-called Poor-Man ensemble where multiple single forecasts from various available models are just pulled together to form an ensemble if one cannot afford to run his own "normal-cost" ensemble (Wobus and Kalnay, 1995; Ebert, 2001). Multi-model approach has been widely accepted and used nowadays. A recent development of multi-model approach is a mixture of multi-ensemble systems from multi-centers such as TIGGE (THORPEX Interactive Grand Global Ensemble) and NAEFS (North American Ensemble Forecasting System) etc. international efforts, which should be now called Rich-Man ensemble (in addition to his own "normal-cost" ensemble). Obviously, a disadvantage of multi-model approach is the cost to develop and maintain many models if it is run by one institute. In addition, multi-model approach can also be utilized in deterministic scope rather than ensemble scope such as the Florida State University's "superensemble" approach (Krishnamurti, 1999) which is a multi-model based MOS-type approach using linear regression technique. It significantly improves forecast accuracy over the original forecasts by correcting model biases. However, this method provides only a most likely deterministic solution with no forecast variance or uncertainty information attached.

Within one model, an ensemble can be formed by alternating physics schemes from one member to another. This multi-physics approach is found to be effective in predicting convective systems with weak large-scale forcing (Stensrud et. al., 2000; Jankov et. al., 2005). Using multiple convective schemes, Du et. al. (2004) compared the relative roles of multi-physics vs. multi-IC in contributing to ensemble spread in short range (1-3 days). Their result showed that IC uncertainty is a dominant contributor to ensemble spread of large-scale basic fields such as wind, pressure, height and temperature, while physics difference provides extra valuable spread mainly being confined to isolated, smaller-scale storm areas. However, for precipitation and convective instabilities such as CAPE, both

IC and physics diversities are found equally important. Similar result is also seen for fine-resolution (4km) storm-scale ensemble (Kong et. al., 2007). Therefore, it's recommended that both IC and physics diversity should be taken into account at the same time in a mesoscale ensemble to maximize forecast diversity. With the NCEP SREF, it is found that the interaction between IC perturbation and physics diversity indeed greatly enhances ensemble spread during warm seasons when combing IC and physics perturbations together although the impact from physics diversity seems minimal in cold seasons. It is expected that multi-physics might be an effective way to build an ensemble system for convection-dominate tropics. A problem noticed of multi-physics approach by simply alternating different physics schemes is that the extra growth rate in ensemble spread gained initially will soon die out with time and cannot sustain over the entire forecasting length.

A more theoretically sounding and sophisticated version of multi-physics approach is Stochastic Physics. Since part of forecast uncertainty stems from parameterization of sub-grid physical processes (Stensrud, 2007) and truncation etc. imperfections of a model, certain parameter values or relevant terms such as tendency, diffusion and energy can be altered (e.g. via multiplying) by, in a stochastic fashion, a factor to account for those possibly missing effects. Therefore, by applying such stochastic process during model integration, forecast value will be altered accordingly. By repeating this stochastic process many times, an ensemble of forecasts can thus be formed. This stochastic process could either be confined within each member without interaction with other members during the entire model integration ("individualism") or be carried out across different members by interactively exchanging information among them during the model integration ("collectivism"). Although this is a promising method both scientifically and economically, a couple of key issues need to be demonstrated before it's fully convinced to replace the current multi-model approach such as (a) can this method steadily outperform the multi-model based ensemble (in terms of mean error, spread-skill relation and probability reliability etc.), and (b) can the extra spread growth rate injected by stochastic physics be sustained during the entire model integration. Some research was done on this in the past (Hotekamer et. al., 1996; Buizza et. al., 1999b; Bright and Mullen, 2002b; Gray and Shutts, 2002; Shutts, 2004) and more is needed to make it a mature method. ECMWF has implemented a version of this method in their global ensemble system (Buizza et. al., 1999b; Shutts, 2004), while NCEP has a plan to do the same for both its global and regional ensemble systems in near future (Hou and Toth, 2007, personal communication).

It is still not clear and an issue to be investigated that how important multi-dynamics is relative to multi-physics in contributing to ensemble spread. Some expect that physics might be more important than dynamics to forecast diversity. This is a very practical issue at NWP centers such as should one model core or multiple model cores be maintained in an ensemble system. It's always easier and cheaper to maintain only one model dynamic core but with varying physics for ensembling.

3.3 3-Dimentional EPS

History or past memory always sheds lights to our future path if it's interpreted and applied properly and is, therefore, an important aspect of weather forecasting (Cao, 2002). In a 3D-EPS, past-memory or history dimension is also considered besides varying IC and model. Direct Time-Lagged ensemble is a typical approach to bring this history dimension in. The degree of consistency from run to run in immediate past should be a measure of forecast uncertainty: high (less) consistency indicates high (low) predictability of an event. This agrees with forecaster's experience: a forecaster is usually more (less) confident when he sees high-consistency (jumpiness) between runs from cycle to cycle. A main concern of using the past-time dimension is that forecast quality degrades with the age of a forecast: older forecasts perform worse than newer forecasts. However, as model and IC quality improves, this might not be true any more if the past time covered is not too old but only immediate cycles. For example, it's not uncommon to observe that a 48hr model prediction could be more accurate than a 24hr prediction. As frequency of running a model at NWP centers increases (e.g., four times per day is a normal practice for many models at NCEP and even more frequently for special models such as NOAA RUC - Rapid Update Cycle - model which is run every hour), the information contained in those immediate past cycles could be huge and needs to be utilized more cost-effectively. An advantage of this dimension is no sacrifice in model resolution. All members are integrated with the highest possible full resolution as the single high-resolution run with no extra computing cost. However, not much research was done to seriously evaluate how to combine a 2-D ensemble with Direct Time-Lagged ensemble into a new 3-D ensemble to improve overall ensemble performance and provide more useful information to users. The situation might improve when more people realize the importance of this past-time dimension in ensembling.

In real world operation, no single EPS (or single type of ensemble product) is universal and satisfies all needs but multiple scale ensemble systems are needed to serve a variety of forecasting purposes. Those multiple systems should work interactively and seamlessly with each other in some kind of adaptive ways (Subsection 2.7). Each system has its own uniqueness in construction and addresses its own unique problems. For example, a climate EPS focuses on the trend of climate change such as due to greenhouse gas-induced global warming or natural variability issues; a seasonal EPS on month to year scale of dominant weather mode such as warm or cold, wet or dry etc.; a global EPS on median-range 3-14 day's large-scale flow pattern and serves early warming purpose; a regional EPS on short-range 1-3 day's detail weather events with more focusing on surface weather elements; a relocatable storm-scale local EPS on 6-24hr details of a particular individual high-impact storm over a specific region of interest such as severe convective storm outbreak, fire weather, hurricane and disastrous event (natural or human-caused); and micro-scale ensemble such as ensemble cloud, turbulence and PBL (planet boundary layer) schemes. Different scale EPS obviously needs different strategies in perturbing ICs and model. For example, both environment and vortex (structure and intensity) need to be perturbed for hurricane prediction (Zhang and Krishnamurti, 1999; Cheung and Chan, 1999a and 1999b); how and what to perturb in ICs (warm bubbles?), might physics play more important role than IC, and how to assimilate special observations like Doppler Radar data into ICs etc. are all issues need to be studied in a convective storm-scale ensemble; fire weather, dispersion ensemble might focus more on near-surface elements and PBL winds and structures; and longer-range forecasts need to consider ocean-atmosphere coupled EPS which lower boundary forcing such as SST etc. should be important; …, just to mention a few.

A frequently asked question is that how many members are needed in an EPS. Based on Du et. al. (1997) study, 7-10 members are normally enough to obtain most of the benefit from an ensemble, a result confirmed by other studies such as Talagrand (personal communication). However, the answer really depends on what your aim is. For example, membership required might be less for 500hpa height and more for convective precipitation; less for a coarse model resolution system and more for a high model resolution system (by the way, an optimal tradeoff between resolution and membership should be determined by cost-benefit ratio); less for ensemble mean forecast and more for probability distribution; and less for prediction purpose and more for data assimilation purpose and so on. It is also possible that the answer might be quite different from a practical or a theoretical point of view: a finite size ensemble might work sufficiently well in practice but a huge or even infinite size might be required in theory. Since a large amount of computing and other resources is involved in ensemble forecasting related tasks, it always needs a balance between efficiency and elegance (Mullen and Buizza, 2002).

4. What products can be derived from an ensemble forecast?

In general, three types of product can be derived from an ensemble: a most probable single solution or consensus forecast, uncertainty measure and a distribution of all possible solutions. Depending on circumstances, most probable single solution could be represented by simple ensemble mean, median or mode of members. Besides, some efforts are also made on constructing such a consensus forecast through more sophisticated methods such as linear regression, ensemble MOS (Gneiting et. al., 2005), performance-based weighting (Woodcock and Engel, 2005), clustering (Greybush et. al., 2007) and Bayesian Model Averaging (BMA) (Raftery et. al., 2005). Note that the "most probable or best solution" is not measured by a particular single realization but by average over a large number of realizations from a statistically reliable EPS (see Part 6). Advantages of the most probable single forecast are information highly compacted with only one single value, easy to understand and use, less confusing, simple and acceptable to most general users and public besides being more accurate statistically. Disadvantages of ensemble averaging are smoothing out spatial details, overestimating light precipitation area coverage and underestimating heavy precipitation area coverage (Du et. al., 1997), and even misleading in bi-modal or multi-modal situations besides providing no uncertainty information. For example, ensemble mean wind speed is very misleading when large uncertainty exists in wind direction among ensemble members. Therefore, we suggest that ensemble spread should also be used to help us to correctly interpret ensemble mean information: e.g., mean is probably more trustable when it's associated with small spread and less trustable when associated with large spread. Under "large spread" situation, other products should also be looked at for further assessment.

Forecast variance among ensemble members can be used to quantify forecast uncertainty. Standard deviation of members with respect to ensemble mean is usually defined as forecast variance known as ensemble spread. Large (small) spread indicates a low (high) confident forecast. To more insightfully interpret the significance of spread information in a forecast, ensemble spread is often normalized by or compared with climate anomaly (Hart and Grumm, 2001). Given that natural variability of a field is quite different over different geographical zones (e.g., large in mid- and high-latitudes and small in tropics), it is sometimes useful too to normalize spread by an averaged spread over a past time period in space to reflect true predictability of atmospheric motion. Advantages of spread are information highly compacted, easy to understand and ability of distinguishing between phase and intensity or amplitude uncertainty of a weather system if combining with ensemble mean information. A main disadvantage is that spread information doesn't tell how members are actual distributed such as normally distributed or skewed or multi-modal distribution. For some variables such as precipitation, forecast variance is somehow related to their ensemble mean values. Under such circumstances, spread doesn't really reflect truth predictability of an event. By normalizing spread of such variables by their mean value might be helpful. In reality, since an ensemble system is not perfectly designed (see Part 3), spread doesn't perfectly reflect true predictability or perfectly correlate with forecast skill. Therefore, a post-processing or calibration is necessary to derive various versions of predictability or confidence related measurement based on raw spread information combining with other information (e.g., IM et. al., 2006).

As discussed in Section 2.2, a complete forecast should be a distribution rather than a single value. To retain full or maximum information in a forecast, many distributive-type of products can be made from an ensemble such as probability, postal stamp chart (individual members), clustering, spaghetti chart, plume diagram (time evolution of forecast values at a given location), extremes or envelope, and three-represented-value type (such as 10-50-90%). A probabilistic forecast can be in 2-D or 3-D spatial distribution at a given time for a given category or a 1-D distribution of probability over all categories at a given location and time or a 1-D time evolution of a probabilistic value (probogram) at a given location. By the way, it needs to keep in mind that the "probability" derived from a finite ensemble (sampling issue) is, theoretically, not true probability but relative frequency. Main advantage of distributive and probabilistic product is full information contained and conveyed, which greatly reduces the chance of having "surprising storms/events" (both missing and false alarm) in forecasts by providing probabilistic heads-up or heads-down information of less-predictable but high-impact weather events in advance. This important piece of information might be captured by only a few members in the ensemble and, therefore, could be completely filtered out in other types of product such as ensemble mean. On the other hand, since distributive and probabilistic products are not deterministic but a range of possibilities, users are often hesitative to make a decision based on probability. How to use probabilistic information in decision-making will be briefly discussed in Section 2.7. As always, each type of products has its own strength and weakness. 2D-probability is displayed only for a given threshold, so that one needs to look at other thresholds at the same time to get a full picture. In conditional probably, one needs to quilt two or more pieces of information together, e.g., only knowing the probability of precipitation type being snow if precipitation occurs (a given condition) is not enough, he also needs to know what is the chance to have the precipitation based on full ensemble members in order to know the whole story. Displaying all individual members together, known as Postal Stamp chart, is found useful for forecasters to have a qualitative first glance to know how diverse among ensemble members, what kind of extremes could be and what to expect on average. Grouping of members into several regimes (Clustering and Tubing, e.g., Alhamed et. al., 2002; Yussouf et. al., 2004; Marzban and Sandgathe, 2006; Atger, 1999a) will provide useful insight in many situations such as multi-modal and regime transition period. Although, mathematically, one can always cluster members into several categories, the separation among clustered groups is not always physically or meteorologically significant. To ensure the separation of clustered groups is meteorologically significant, clustering technique needs to be carefully designed and the significance of separation between groups needs to be statistically tested besides the necessity to have a sufficiently large ensemble size. Spaghetti chart (a group of contours of a selected value from each member being plotted together) is widely used. Since it gives full solutions of all members, one can clearly picture what the mean, mode, distribution and outlier are. Disadvantage is that it's not always pre-known which contour value is proper to be picked for a particular event since there are only a few values (usually one or two) can be displayed in a spaghetti chart. Envelope display is good to see possible extremes at both low and high ends but lacks of detailed member distribution within. However, for a given location, the time evolution of both extreme values and member distribution can be displayed together as a plume diagram. In practice, to reduce data volume while still possibly to retain critical forecast information, a compromised version of full probabilistic distribution is the three-represented-value type product. Two values represent extremes at low and high ends (e.g., either extreme values themselves or values associated with 90% and 10% probability or similar), while one value represents a most probable solution like median, mode or mean. All the above three types of products (most probable, uncertainty and distribution) can sometimes be combined into one single product such as Fig. 3 (nicely showing uncertainties in both wind speed and direction) although it's not always easy to do so. How to condense abundant ensemble information into a simple product which can be easily understood and used by forecasters and end-users is an issue needs to be further studied. Storm Prediction Center of NCEP developed a lot of multi-variable joint-probability type of products which is found a meaningful way in fire weather, convection and winter storm forecasting (Fig. 4, Bright et. al., 2004 and Weiss et. al., 2007).

Usually, IC perturbations are centered around a control analysis, the control member tends to remain near the center of an ensemble cloud and close to ensemble mean in linear or quasi-linear situation. Hence, ensemble mean often verifies more accurate than other perturbed members, which implies that control member could have different statistical property comparing to other perturbed members. Therefore, a question is that control member should be included or excluded or be weighted more in calculating final ensemble products such as probability, spread and mean? Practically, this question only matters when ensemble size is small (say less than 50) since including or excluding control member should not make much difference for a large size ensemble. Also, in a highly nonlinear situation or longer range time integration, control member could be anywhere within the ensemble cloud. In that case, control member should perform similarly as other perturbed members and, therefore, be included in the calculation without being distinguished. Only in the situation where ensemble size is small and flow is relatively linear, this question becomes relevant. The answer to this question may vary depending mainly on two factors: whether an EPS is biased or unbiased and has sufficient spread or not. If an EPS is unbiased and has sufficient spread or over-dispersive, control member is often more accurate and plays more important role and, therefore, it should be included in calculating probability, be weighted more in calculating ensemble consensus forecast but may be excluded in calculating spread to maximize

Multimeteogram

(uses Confidence Intervals, as opposed to a plume plot that shows data from all members)

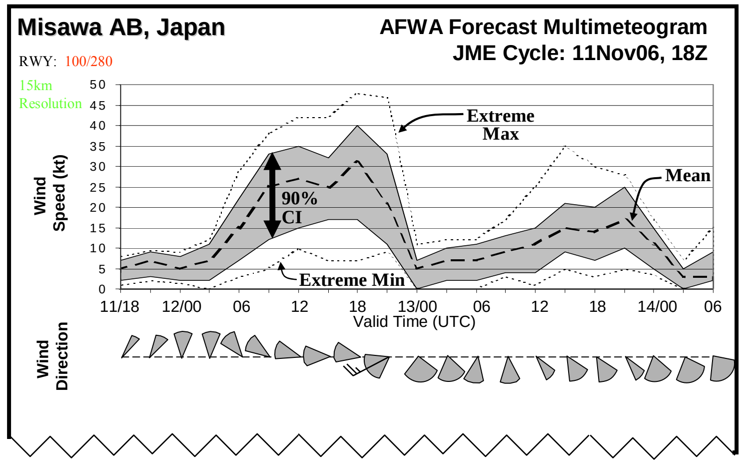

Figure 3 60-h forecast for surface winds at Misawa air force base, Japan. The range of possible wind speed is given with a 90% confidence interval (shaded) and extreme maximum and minimum (dotted) together with ensemble mean value (thick broken line). Wind direction uncertainty and prevailing (mean) direction are also plotted at the bottom. [Courtesy of Maj. F. Anthony Eckel of Naval Postgraduate School]

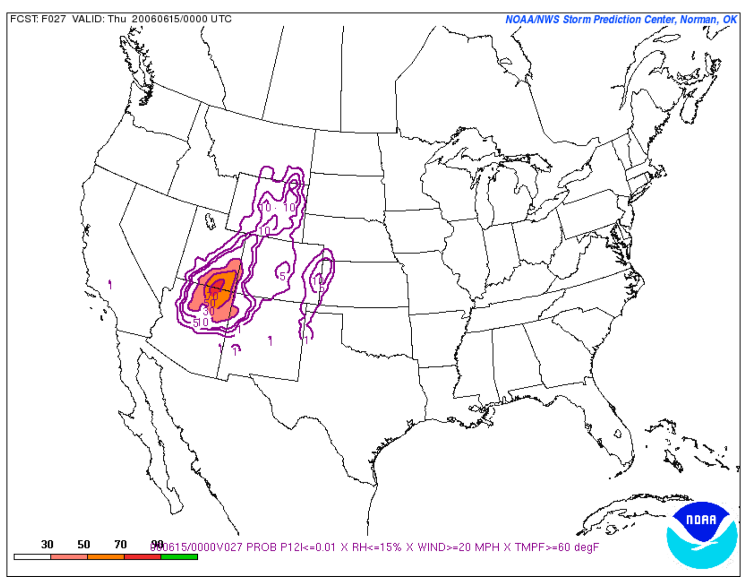

Figure 4 NCEP SREF based Joint or combined probability of Fire Weather Index developed and operationally used by NCEP Storm Prediction Center (SPC). This is a 27-h probabilistic forecast meeting all the following four conditions: 12h precipitation <=0.01 inch, surface relative humidity <=15%, surface wind >=20 miles per hour and surface temperature >=60 degree Ferenhite. [Courtesy of Drs. David Bright and Steve Weiss of SPC]

spread in an under-dispersive ensemble (since the inclusion of control run might reduce spread because the control forecast is probably closer to ensemble mean under linear or quasi-linear situations). Otherwise, if an EPS is either severely biased or under-dispersive, truth is often outside the ensemble cloud. In that case, control member (near the center of ensemble cloud) has less chance to be correct than some perturbed members and, therefore, might not be included in ensemble products to avoid a possibly degraded forecast. By the way, from the above discussion, one might see that the difference between ensemble mean and control member could be used as a measure of nonlinearity: the larger this difference, the higher nonlinear a flow exhibits.

5. What is the role of EPS post-processing?

In reality, model used by ensemble has bias; the uncertainty source of a forecasting system cannot be fully and accurately described by an EPS as well as model spatial resolution has to be compromised due to huge computing cost, which results in a suboptimal ensemble system. The defects of such suboptimal ensemble system include the following: ensemble mean not being better than control and other perturbed individual members, suboptimal spread-skill relation (under- or over-dispersive spread), excessive outliers, unreliable probability and lacking of spatially detail structures etc. Even with perfect IC perturbations, ensemble-based PDF distribution could be woeful at so called "unpredictable spots" as long as model possesses even very small error (Du, 2005). Therefore, post-processing is a necessary and important step to calibrate raw ensemble forecasts. For example, by removing model systematic bias (1st moment), ensemble mean forecast will be more likely to close to the best solution, outlier will be significantly reduced and probability will be more reliable. For multi-model ensemble system, it also ensures no spurious spread introduced when bias of each model is removed before the sub-ensembles are combined into one grand-ensemble. Removing bias is also very important for ensemble-based data assimilation technique. By calibrating 2nd moment (forecast variance), spread-skill relation and the under- or over-dispersive problem of a spread could be improved and remedied. To further improve reliability of a probabilistic forecast, higher-moment such as PDF distribution also needs to be calibrated. In many applications such as hydrology and fire weather, downscaling of a lower-resolution ensemble is necessary to resolve local-scale features. Since ensemble is not perfect in real world, the equal-likelihood property of each member's performance might be violated (see Part 6), i.e., members could perform differently in quality under different weather conditions especially in multi-model or multi-physics based systems. Under these circumstances, some kinds of performance-based weighting to different members might be found useful in practice before combining all members into an ensemble product, which is another kind of post-processing procedures. Removing bias is believed to be important too in searching for "best member". We are always hoping to know in prior which member might verify the best although it's almost impossible since all of them should be equal likely in theory (see Part 6). However, if we can completely remove systematic bias from an ensemble, ensemble median or mean (if distribution not too skewed) should verify the best on average. Thus, for an individual event, we might be able to identify which member is likely to be the best since the member that is closest to the median or mean should also verify close to the best. One might expect that different member could serve as "best member" at different forecast lead time or over different regions or with respect to different weather systems or parameters. In contrast, in a biased EPS, there is no easy reference to be used to identify the best member, where ensemble median or mean should consistently perform poorer than the best member. All those indicate the importance of ensemble post-processing.

There are, in general, two kinds of approaches in post-processing: statistical and dynamical. Since statistical approach is based on past error information, it should work well when bias is relatively constant from day to day and large in size but poorly when bias varies with flow and is small as well as weather regime changes. Statistical approach has many different versions including commonly used running-mean which is an equally-weighted average over a past period of time (e.g., Stensrud and Yussouf, 2003 and 2007; Yussouf and Stensrud, 2006), simple decaying-average which intends to focus more on the most recent past data with equivalently "decreasing weights" with data ages using a Kalman Filter type adaptive algorithm (Cui et. al., 2005), regime-dependent analog approach where weighting depends on flow pattern (Du and DiMego, 2008), linear regression (Krishinamurti et. al., 1999 and 2000; Yuan et. al., 2007a), Artificial Neural Network (Yuan et. al., 2007a and 2007b) and Bayesian Model Average (BMA) etc.. Some of them are more sophisticated than others. The basic idea of BMA is that for each ensemble member, create a probabilistic distribution, then assign a weight to each distribution based on past performance of each member, and finally use weights to combine all distributions into one "master" probabilistic distribution. This BMA approach is gaining its

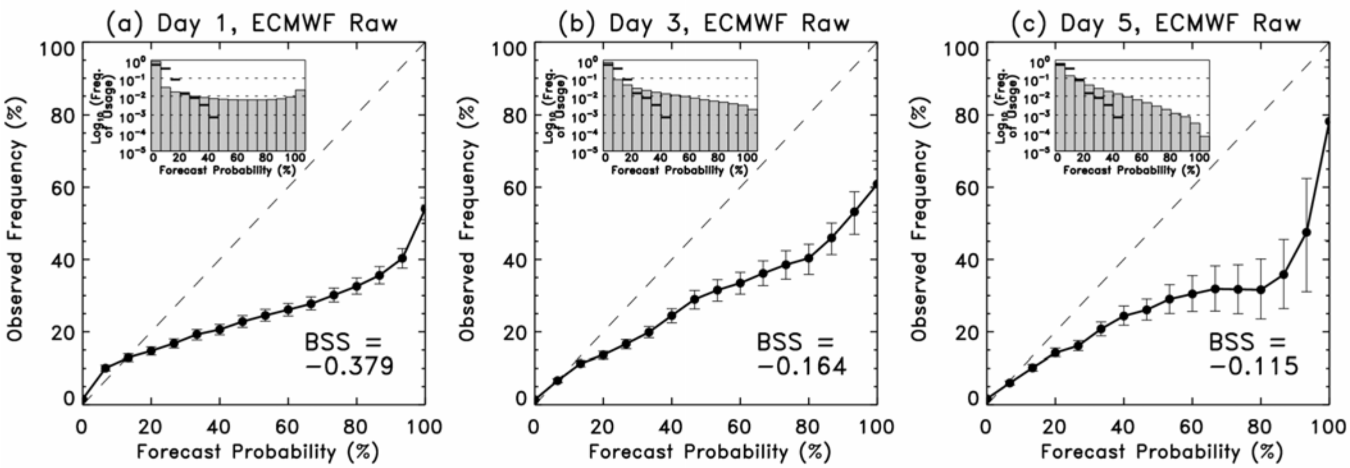

Figure 5 Reliability score of probabilistic forecast of 12-hourly accumulated precipitation >= 5 mm over Continental US at day 1, 3 and 5 leads averaged over 480 case days, based on raw ECMWF global ensemble (15 members). It shows over-confident in the probabilistic forecasts. [Courtesy of Dr. Thomas M. Hamill of Earth System Research Lab/NOAA]

popularity nowadays (Raftery et. al., 2005; Sloughter et. al., 2007; Wilson et. al., 2007). For short-range forecasts (1-3 days), a short data-training period such as 14-30 days might be enough, while for longer-range forecasts (beyond a week), much longer training period might be needed. The length of training period depends on variables too such as shorter for temperature and longer for precipitation. For situations requiring long training period, some special datasets such as hindcast or reforecasting (Hamill et. al., 2004b and 2006) might be purposely generated for post-processing to use. It's reported that using reforecasting data can effectively calibrate probabilistic forecasts to be more reliable (comparing Fig. 5 and 6; Hamill and Whitaker, 2007; Hagedorn et. al., 2007; Hamill et. al., 2007). Post-processing can be applied to 1st moment (mean), 2nd moment (variance) and higher-moment (such as PDF distribution, Eckel and Walters, 1998). For example, statistical dressing and shadowing are ways to increase spread for an under-dispersive ensemble (Roulston and Smith, 2003; Berrocal et. al., 2007; Gilmour and Smith, 1997). Statistical approaches can also be applied for downscaling where topography and other information might also be considered at the same time. A dense observation is surely critical in statistical downscaling and other post-processing.

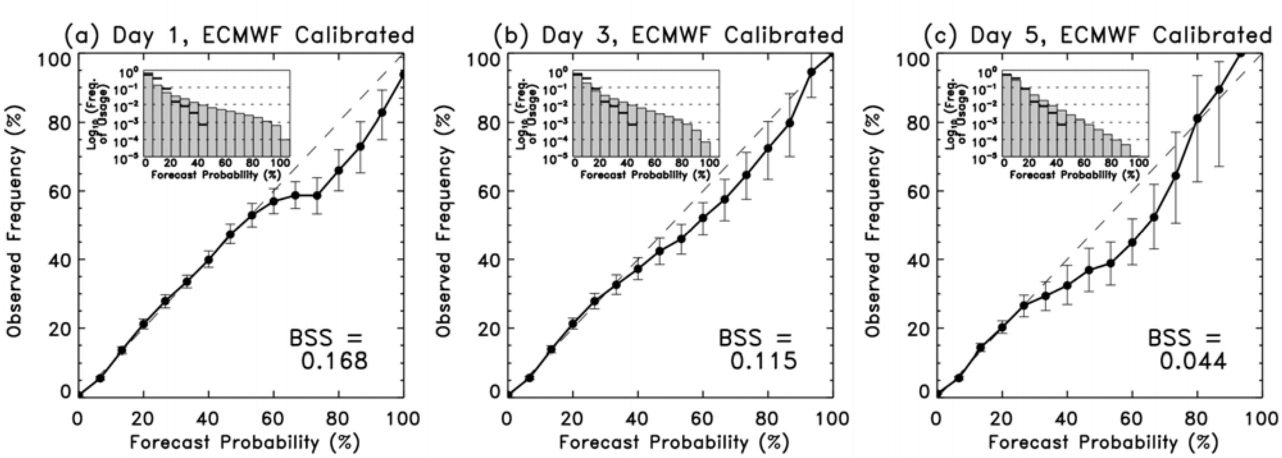

Figure 6 Same as Fig. 5 but for post-calibrated ensemble with a 20-year weekly 15-member reforecasting ensemble data. It shows much more reliable probability comparing to the raw ensemble-based probability (Fig. 5).

In many real world situations, bias varies with flow, and the systematic and random error components are hard to be separated (even such a separation is not physical if mathematically can be done), so that statistical methods don't work well but flow-dependent dynamical approach is desired. There are no widely accepted dynamical methods yet but it remains a land of wildness to be explored in this direction. Based on author's personal experiences, methods could include multi-model based, dual-resolution such as Hybrid Ensembling approach (Du, 2004), spread-error relation and stochastic physics etc.. Dual-resolution is also common for dynamical downscaling. However, for very high-resolution dynamical downscaling, it might be too expensive to run a full-physics downscaling model. Then, an alternative might be to run the very high-resolution downscaling model with no or reduced physics which might be a reasonable assumption for short range. Stochastic physics has potential in reducing bias by simulating various bias effects in model equations. Since a favorable large-scale environmental condition is necessary for an event such as heavy precipitation to occur, careful diagnosis of related environmental dynamical conditions such as moisture convergence, vertical motions and instability might help to calibrate a forecast as a post-processing too (Gao, 2007).

6. How to evaluate the quality of an EPS and its forecasts?

How and what to evaluate is important because it not only gives one a sense of correctness or wrongness but more importantly it could shape how an EPS or a model being developed in a long run. In general, four aspects need to be verified to measure the quality of an ensemble system: equal-likelihood of each ensemble members, superiority of ensemble mean to single control forecast, high spread-skill relation and reliable probability. Those four aspects are related to each other in certain ways. Because all perturbed ICs are supposed to be equal-likely true and all perturbed physics or varying physics schemes or alternative models are also equally plausible, performance of all ensemble members should be, in principle, similar to each other on average. Otherwise, it indicates problems of ensembling technique employed, e.g., either IC perturbation size is too large or alternative models, physics schemes or perturbations added are not really equally plausible. Due to this equal-likelihood property, one can image that it's difficult if not impossible to determine in prior which member is likely to perform the best in a particular forecast.

Due to nonlinear filtering, discrepancies among members (i.e., less predictable elements) are damped or cancelled and only those common features among members (i.e., more predictable parts) are remained during the process of ensemble averaging. This will result in a superior ensemble mean forecast to a single or even higher-resolution control forecast on average. Figure 7 is an example in hurricane track forecasting showing the ensemble mean is close to the observed track. In grid-point verification, smoothing effect of the averaging partially contributes to this superiority but should be in a much less degree merely as a side effect comparing to the nonlinear filtering. It needs to point out that the ensemble averaging can only remove random error but not systematic bias error if an ensemble consists of only one model with one version of physics package. For such an ensemble system, effectiveness of error reduction by ensemble averaging should be measured with respect to the random portion of forecast error but not to the systematic error such as mean-error or the total error such as root-mean-squared error (RMSE) which is a mixture of random and systematic errors. Otherwise, one might be comparing the relative performance of two modeling systems but not two ensembling techniques or strategies, therefore the conclusion drawn might be misleading. However, for a multi-model and/or multi-physics ensemble, bias error could also be reduced by ensemble averaging due to possibly different biases possessed in different schemes or models. Similar to ensemble mean, ensemble median forecast should verify the best too on average. To measure ensemble mean/median forecast accuracy, all methods normally used in evaluating deterministic single forecasts can also be applied such as Threat score, Equitable Threat score, RMSE and correlation etc.

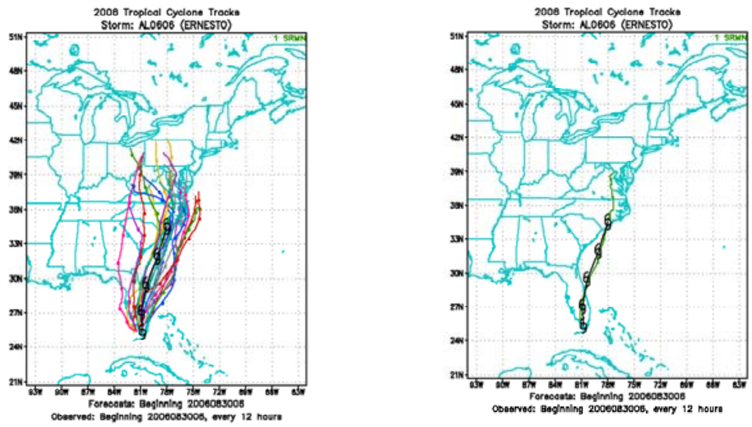

Figure 7 21-member NCEP SREF forecasts of the hurricane Ernesto (2006) track: individual members at the left and ensemble mean at the right with the observed track (hurricane symbols). [Courtesy of Mr. Timothy Marchok of GFDL/NOAA]

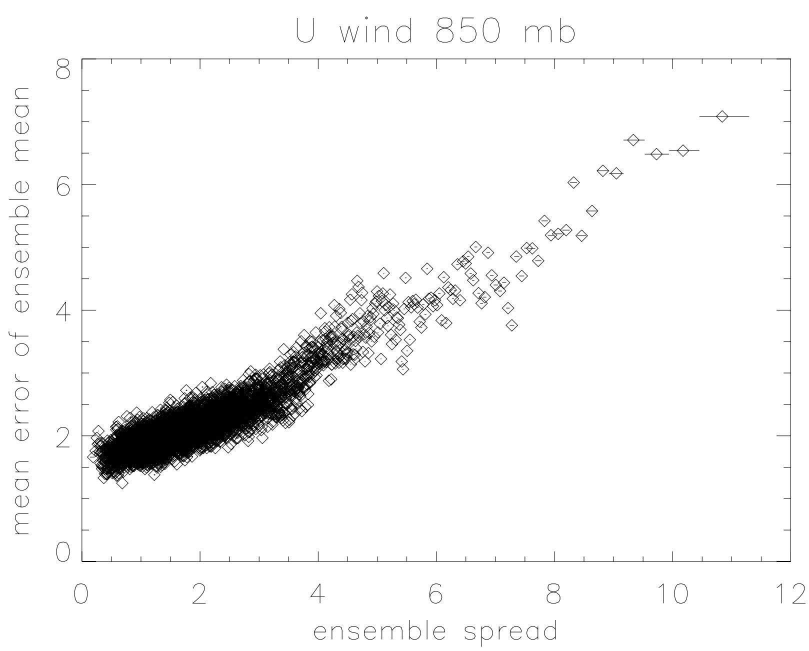

Figure 8 The spread-skill relationship between ensemble spread and mean absolute forecast error, <|forecast – verification|>, of 63-h forecasts of u component of 850-hpa wind of NCEP SREF. Each point represents the mean error of 1000 forecasts with similar values of ensemble spread. [Courtesy of Dr. Mark S. Roulston of UKMO]

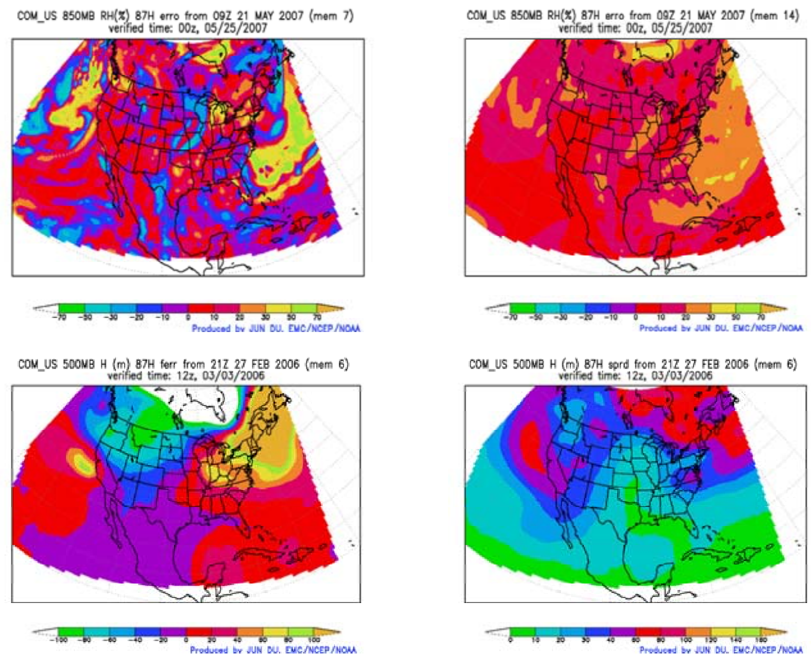

Figure 9 Two snapshots of spatial distribution of 87-h mean-forecast error and spread from NCEP SREF. Upper panel is for 850-hpa relative humidity initiated at 09z, 21 May 2007. Lower panel for 500-hpa height initiated at 21z, 27 Feb, 2006. Left: error; right: spread.

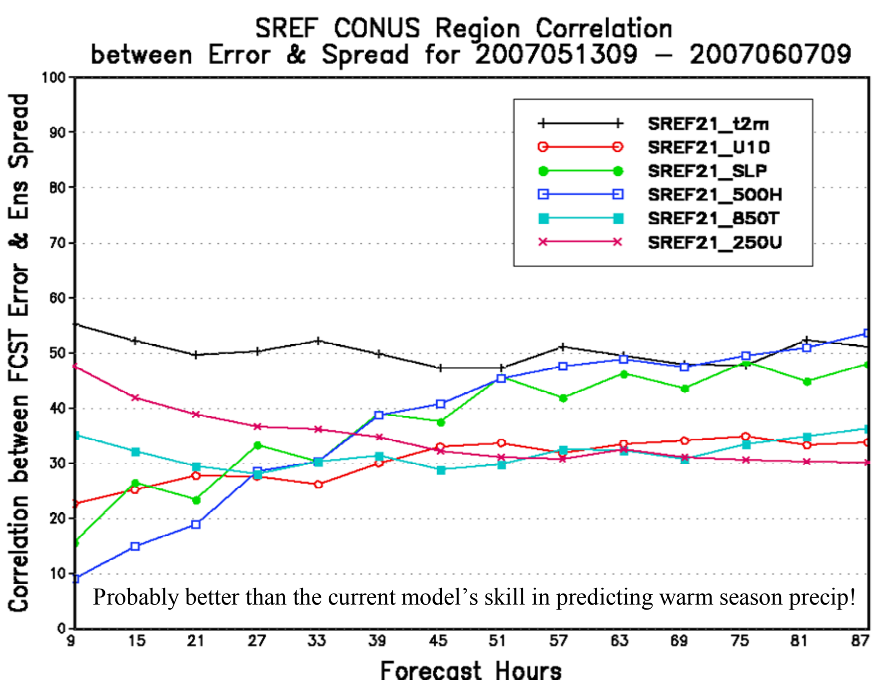

Figure 10 Spatial correlation (%) between ensemble spread and absolute error of mean-forecast of 9z cycle of NCEP SREF averaged over a period from May 13, 2007 to June 7, 2007 for various forecast variables.

For a good ensemble system, ensemble spread should be a good indicator of possible forecast error distribution since spread should reflect the true predictability of a flow. Large (small) spread indicates less (more) predictable event, while less (more) predictable event is more (less) difficult to forecast and should have wider (narrower) error range. Therefore, spread and absolute error of ensemble mean forecast should be positively correlated on average such as shown in Fig. 8, which is called spread-skill relation (Whitaker and Loughe, 1998; Roulston, 2005; Grimit and Mass, 2007). Figure 9 is a snapshot of the spatial distributions of the forecast error of ensemble mean and the ensemble spread from the NCEP SREF, which does show good relation between spread and error in the spatial pattern of large-scale. Figure 10 shows the spatial correlation of various variables averaged over a period of about a month. We can see that such correlation exceeds 50% at day 3.5 for sea-level pressure, 500hpa height and 2m temperature, which is comparable to or even better than the quality of quantitative precipitation forecast by the current state-of-the-art NWP models and, therefore, skillful in providing useful guidance to forecasters. It's also interesting to notice that the spread-skill correlation of sea-level pressure and 500hpa height increases steadily with forecast time and remains below 30% at day 1. This could imply that the current ensemble technique used by NCEP SREF might not be suitable for very short range forecast (0-24hr), which is a subject worth being investigated. Many examine the spread-skill relation by simply comparing two domain-averaged curves of spread vs. RMSE of ensemble mean. By doing so, one needs to keep in mind that the closeness of the spread curve to the error curve is only a necessary condition but not a sufficient condition in measuring true spread-skill relation since their actual spatial patterns might not match to each other although their domain-wide summary statistics does. To avoid this, the comparison between spread and error must be carried out point by point in space. Spatial correlation, as discussed above, is one way to do. Another way is the rank histogram or rank distribution known as Talagrand Diagram/Distribution initially proposed by Talagrand et. al. (1997) and Anderson (1996) and late also documented by Hamill (2001). For a good ensemble system, truth acts just like one of the ensemble members and is not distinguishable from them. In other words, truth has equal chance to be anywhere between any two members. In Talagrand Distribution, all n ensemble members are first sorted out in order from smallest to biggest in value and form n+1 bins including two endings at a given point and time. Then, the chance of truth entering each bin is counted point by point and summed over a region and a period of time. If the resulting distribution of the chance over all bins is flat (--), the truth has the same statistical properties as all ensemble members, then ensemble spread is reliable and reflects true error distribution (perfect spread); if the distribution is in "upside down" U shape, spread over estimates forecast uncertainty (over-dispersive); if the distribution shows L shape, forecasts have high-bias; if shows "reversed" L shape, forecasts have low-bias; and if the distribution is in U shape, it indicates either spread under estimates uncertainty (under-dispersive) or some forecasts have low-bias and some high-bias. Figure 11 is an example of this, showing quite satisfactory spread of 500hpa height from NCEP SREF. Even for a perfect but finite ensemble system with n members, one should always expect [2/(n+1)x100] % (the sum of two end bins) of the time that truth would fall outside the ensemble cloud (outlier) and [(n-1)/(n+1)x100] % chance that the ensemble cloud would encompass the truth. To obtain more reliable assessment out of Talagrand Distribution, Minimum Spanning Tree approach is sometimes used to aid the calculation (Smith and Hansen, 2004; Wilks, 2004; Gombos et. al., 2007).

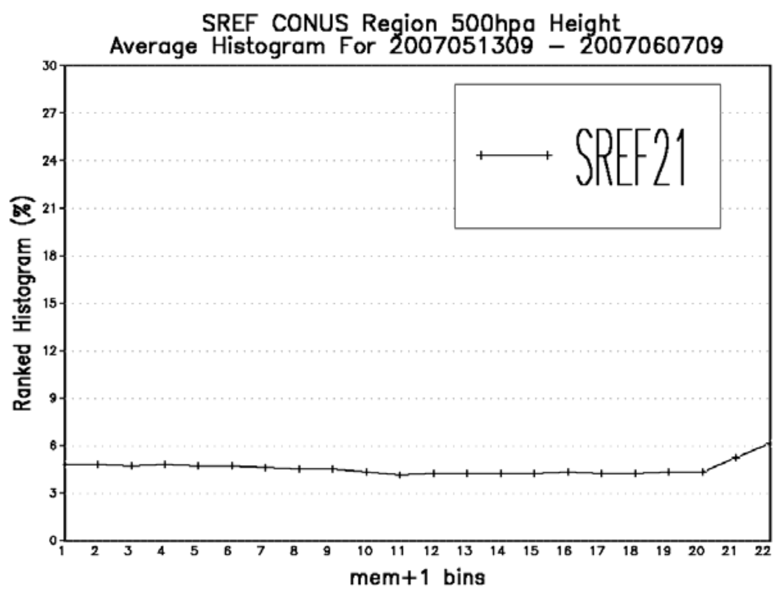

Figure 11 Talagrand distribution of 500-hpa height ensemble forecasts at 87h lead time averaged over a period from May 13, 2007 to June 7, 2007, based on NCEP SREF data. It indicates slight low bias.

There are two attributes to measure the usefulness of a probabilistic forecast: reliability and resolution (Jolliffe and Stephenson, 2003; Roulston and Smith, 2002; Atger, 1999b). In a reliable probabilistic forecast, a probability really means what its surface value says. For example, a perfectly reliable 60% forecast of an event means that in 60 times out of 100 such "60%" forecasts (either spanning in space or in time) the event will actually occur and 40 times not occur. A median forecast

(50%) should verify half of the time. This property can be measured by the Reliability Score (Wilks, 2006). A perfect Reliability Score curve is a diagonal line with x-axis in forecast probability and y-axis in observed frequency of the predicted event. Figure 12 shows an example of it based on NCEP SREF. A reliable probabilistic forecast has, however, no ability to tell which particular probabilistic forecast would verify and which one wouldn't. The ability to distinguish the ones that would occur from those that would not occur is called Resolution. The resolution is related to the sharpness of probability density function (PDF). The sharper a PDF is, the higher resolution or the more skill or the more information a probabilistic forecast has. A perfect deterministic single forecast has perfect resolution (perfect reliability too): full capability of distinguishing "yes" event from "no" event and yes means yes and no means no. Climatology forecast is perfectly reliable but has no resolution. When a PDF distribution becomes as flat as climatology PDF, a probabilistic forecast becomes no skill although it's perfectly reliable. So, one can see that as long as a probabilistic forecast is reliable, the higher the resolution is, the more valuable a probabilistic forecast will be. Reliability reflects how well IC and model physics etc. are perturbed in an EPS. Good spread-skill relation is a basis to have a reliable probabilistic forecast. However, resolution cannot be improved through ensemble technique but only through improvement of model and IC quality themselves. Note that reliability score can be severely contaminated by model systematic bias, while resolution is mainly related to and affected by random error. Therefore, removing model systematic bias can improve reliability but not resolution in a probabilistic forecast. Since ensemble averaging can reduce random error, it improves resolution for a single deterministic forecast. Spatial correlation (to truth) is a way to measure resolution since it reflects random error (not systematic error) for a single forecast. To assess probabilistic forecast accuracy, many other scores are also used such as Brier score (BS, Brier, 1950) for one-category (e.g., rain or no rain) event and Ranked Probabilistic score (RPS, Epstein, 1969; Murphy, 1969 and 1971; Wilks, 2006 and Du et. al., 1997) for multiple Mutually Exclusive and Collectively Exhaustive categories or Continuous Ranked Probability score (CRPS) for continuous variables (Brown, 1974; Unger, 1985; Grimit et. al., 2006). Analogous to RMSE in single forecast verification, BS and RPS are the average deviation between predicted probabilities for a set of events and their outcomes, so a lower score represents higher accuracy. 0 is perfect and 1 (J-1) the worst in BS (RPS), where J is the number of event categories. Similar to RMSE, BS and RPS measure total error and are contributed by both reliability and resolution. Traditionally, BS, RPS and CRPS etc. can be decomposed into reliability, resolution and uncertainty components (Hersback, 2000). Relative Operating Characteristic (ROC) diagram is another tool often used to assess probabilistic forecast using False Alarm Rate (FAR) as x-axis and Hit Rate (HR) as y-axis (Harvey et. al., 1992). An ideal forecast has 0% FAR and 100% HR, while a worst forecast 100% FAR and 0% HR. When HR and

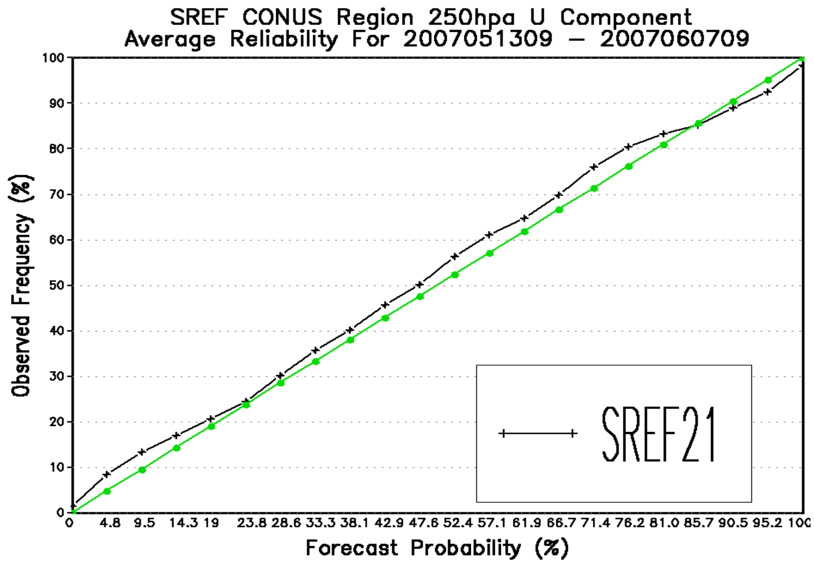

Figure 12 Reliability score of U-component of 250-hpa wind speed probability >= 20m/s at 45h lead time averaged over a period from May 13, 2007 to June 7, 2007, based on NCEP SREF data.

FAR are evenly divided (50% each), it denotes a no-skill forecast shown by a diagonal line in the ROC diagram such as climatological forecast. Area under the ROC curve (AUC) is used to quantify this score: 1 for perfect, 0 for worst and 0.5 for no-skill forecast. Motivated by search for a metric that relates ensemble forecast performance to things that customers will actually care about, Economic Value (EV) score is developed (Richardson, 2000; Zhu et. al., 2002). Calculation of EV is based on the two components of ROC (FAR and HR) as well as a cost-lost ratio which is closely related to customer's dependency on weather. Most of the above scores are sometimes converted into skill-score format with respect to a same score of a reference forecast such as climatology to have a better idea about the relative performance of the forecast. Cautious is needed when using climatology to calculate forecast skill to avoid possible overestimation of skill scores due to possibly different climatologies used (Hamill and Juras, 2007).

Finally, one should keep in mind that no matter how an EPS is verified, a good EPS forecast should demonstrate the following three general properties: the consistency from one cycle to another (probability is found much more consistent than a single-value forecast is), the quality (for single-value forecast) or reliability (for probabilistic forecast) regarding distance between forecast and observation, and the value or benefit realized from action taken by considering the forecast information.

7. How to communicate forecast uncertainty and use probability information in decision-making process?

No matter how accurate a forecast is, a forecast is valuable only until it is correctly understood and used by an end-user to make a decision and take a necessary action upon it (Murphy, 1985). Therefore, the way to effectively and accurately communicate a forecast to end-users is critical since it determines what kind of information a user might get. This is particularly true and important for a forecast in probabilistic form. A same piece of probability information can lead one to take very different actions based on ways of expression. For example, a psychological experiment shows that given two jars, one with 1 red and 9 white balls and another with 10 red and 90 white balls in it respectively, if one can randomly pick one ball (only once) out of any one of the two jars of his own choice and wins an award if the ball he picks is red, it's found that one is more likely to go to the jar with 100 balls to play the gamble hoping more chance (10 instead of 1 red balls!) to pick a red ball although the probability is exactly the same 10% mathematically. In general, what a user gets is often less than what we tell and what we tell is often less than what we know, which indicates rooms for improvement in communicating weather information. How to better convey probabilistic forecast information to end-users is still new to meteorological community and needs to be carefully studied together with scientists in behavioral sciences.

How to apply probabilistic information to decision-making is often most confusing to many people. One might complain what should I do with it if a probability says 50%, half right half wrong? As discussed in the Part 6, given a reliable ensemble system, a 50% forecast means that in 50 times out of 100 such "50%" forecasts of the event will actually occur and 50 times not occur. Thus, this information has significant economical value to a specific business based on its dependence on weather. Table 1 lists possible economic losses and costs involved in a damage-causing weather event under various decisions, where Lu is the loss that cannot be protected against, Lp the loss that can be protected against and C the cost of protection. The benefit of taking action if the event does occur is the difference between L and (Lu + C), while the risk of taking action is wasting the cost C if the event doesn't occur. Therefore, for a reliable P% forecast, the possible benefit and risk are, respectively,

$$\text{Benefit} = P\% \times [L - (Lu + C)] = P\% \, (Lp - C) \qquad (6)$$

$$\text{Risk} = (1\text{-P\%}) \times C \tag{7}$$

For a reliable P% probabilistic forecast	Action taken (as "yes" forecast)	No action taken (as "no" forecast)
Event occur (p%)	Smaller mitigated loss Lu + a cost C with possibility of P%	Bigger total Loss L=Lp+Lu with possibility of P%
Event not occur (1-p%)	No loss but a cost C with possibility of (1-P%)	No cost and no loss with possibility of (1-P%)

Table 1 Potential economic loss and cost involved in a weather forecast by decisions.

Logically, a decision should be made based on the Benefit/Risk ratio. When the ratio > 1.0, one should take action; when the ratio < 1.0, no action; when the ratio close to 1.0, either way might result in similar economic consequence. The Benefit/Risk ratio (related to Lp and C) is strongly user dependent. Figure 13 shows that "user 1" is very sensitive to weather and takes action when probability is close to 20%, "user 3" is not much dependant on weather and takes no action until probability approaches to 80%, "user 2" is moderately dependent on weather and takes action when probability is around 50%, and "user 4" is totally weather independent and takes no action at all no matter how accurate a forecast is. Obviously, this ratio varies also with location such as city or countryside, time such as weekday or weekend, rush hour or non-rush hour, and event significance such as casual or formal activities, political or non-political gatherings etc.. To better serve society and people, meteorologists should work together with individual end-users to carefully develop optimal decision-making tools based on Benefit/Risk ratio to maximize the utility of probabilistic weather forecasting information.

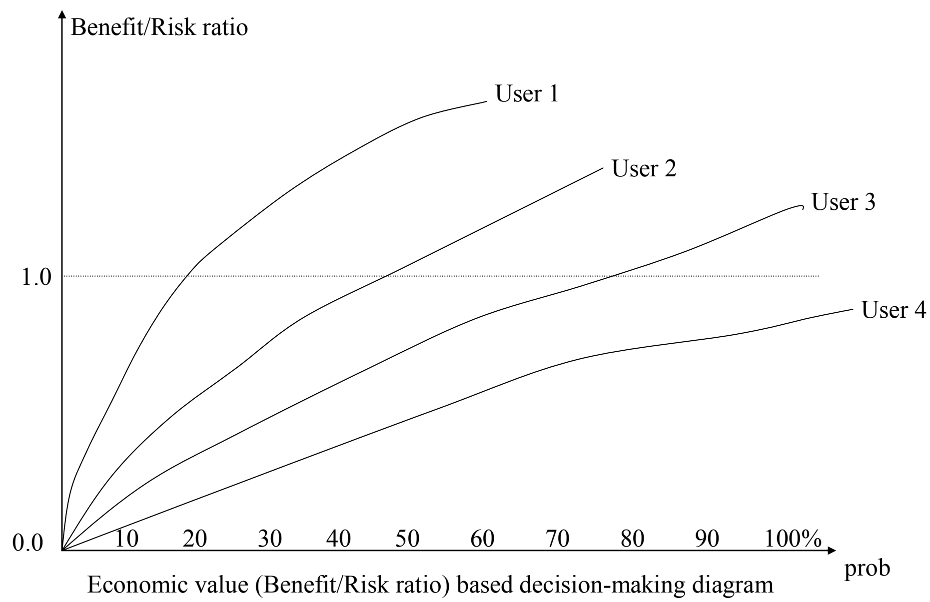

Figure 13 Illustration of an economic value (Benefit/Risk ratio) based decision-making diagram for end-users using probabilistic weather Information.

Please note that besides quantitatively conveying forecast uncertainty such as probability, forecast uncertainty can often be expressed qualitatively too such as via tone of voice, choice of words and even body language especially in TV or radio broadcasting to general public. Some kind of explanation why this particular forecast is to be so uncertain will be very helpful to users in correctly receiving information and making best decisions. For example, a forecaster needs to frankly explain to public that the precipitation type, snow or rain, is very hard to be determined but both are possible for tomorrow's weather because the local area is just near the freezing line (0°C or 32°F temperature zone).

8. What is the impact on downstream applications?

As meteorological ensembles (met-ensemble) become available as part of real-time operation at more and more NWP centers, it's a growing area that many downstream prediction systems which strongly depend on meteorological forecasts as their inputs are also gradually deviating from traditional single-value input paradigm by actively testing how to couple with a met-ensemble system to quantify the forecast uncertainties in their downstream predictions.

Such downstream systems are but not limited to hydrology, air quality, transportation and dispersion, ocean waves, ice drifting, costal storm surge, and electricity generating etc.. Hydrological prediction is sensitive to precipitation information. Uncertainty in precipitation amount and type (liquid or solid) and maybe also in temperature will certainly cause large uncertainty in hydrological prediction of flooding, runoff and stream/river flow in both short term such as flash flood and long term such as snow-melting. Unreliable probability of precipitation forecast caused by model bias and imperfect met-ensemble as well as the mismatch in spatial scale between coarse model resolution of met-ensemble and fine scale of river basin or catchment are main challenges for a hydrological prediction system to correctly use met-ensemble information (Franz, et. al., 2005; Schaake, et. al., 2006). Therefore, downscaling and post-calibration of met-ensemble data is extremely important to hydrological application. Many meteorological fields such as temperature, advection, convection, precipitating process, radiation and especially surface and planetary boundary layer (PBL) properties like turbulence play important roles in controlling air quality and the transportation and dispersion process of pollutant. Those meteorological fields often exhibit large forecast uncertainties. Therefore, predicting air quality and dispersion process must suffer large uncertainty too. This issue has now been paid much attention by air-quality modelers and homeland security dispersion modeling community although how to properly simulate PBL-related uncertainties in a met-ensemble is still an issue yet to be researched. Since predictions in ocean surface wave, ice drifting and coastal storm surge are mainly driven by strong surface wind which possesses large forecast uncertainty, coupling with met-ensemble is also underway in those downstream prediction systems. To quantify such forecast uncertainty, NCEP has already implemented a wind-driven ocean-wave ensemble system in operation (Chen, 2006). Application to electricity generation is also popular in energy companies (Stensrud et. al., 2006). In all those downstream applications, most of them directly use individual met-ensemble members to drive a multiple of downstream predictions (a more expensive way in computing), while some use only forecast variance derived from met-ensemble to quantify uncertainty of a downstream prediction (a less expensive way in computing) such as in dispersion modeling (Warner et. al., 2002).

9. Shifting operational forecast paradigm

From the discussion in Part 2, one can see that there is a fundamental shift in NWP practice and philosophy from the single forecast-based paradigm to the ensemble-based one (Fig. 14). Four main differences are summarized below. First, the former views a forecast as a deterministic process expressed by a single value hoping to have a single best shot, while the latter as a stochastic process (within the range of uncertainty) expressed by probabilistic distribution hoping to address forecast uncertainty alone with a most probable solution. In other words, besides providing a most probable and improved single solution an ensemble also provides extra forecast uncertainty information comparing to a single-value forecast (see Part 4). Therefore, there is nothing to loss but purely gain (complete) when the NWP paradigm shifts from the single-value forecasting to the ensemble forecasting. Besides the omitting of forecast uncertainty, uncertainty in observation and ICs is also ignored in single-forecast based NWP, while, in ensemble-based NWP, observation should also be expressed by a distribution in data assimilation process and multiple analyses would be created to form an ensemble of ICs for model integration to consider uncertainties in both observation and data assimilation procedure. As science and technology advances, ensemble forecasting should become a main tool in weather, climate and water forecasting. In order to smoothly complete this paradigm transition, training and education is an urgent issue. Meteorologists and end-users need to work together to develop strategies of how effectively and correctly using uncertainty and probabilistic information in decision-making process to benefit the entire society the best. The recent study report published by the U.S. National Research Council of National Academy of Sciences - Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts - is a good example of this effort (NRC, 2006). UCAR COMET classroom training courses (http://www.comet.ucar.edu/class/index.html) and local Weather Forecasting Offices' ensemble training workshops are all good formats in this training effort.

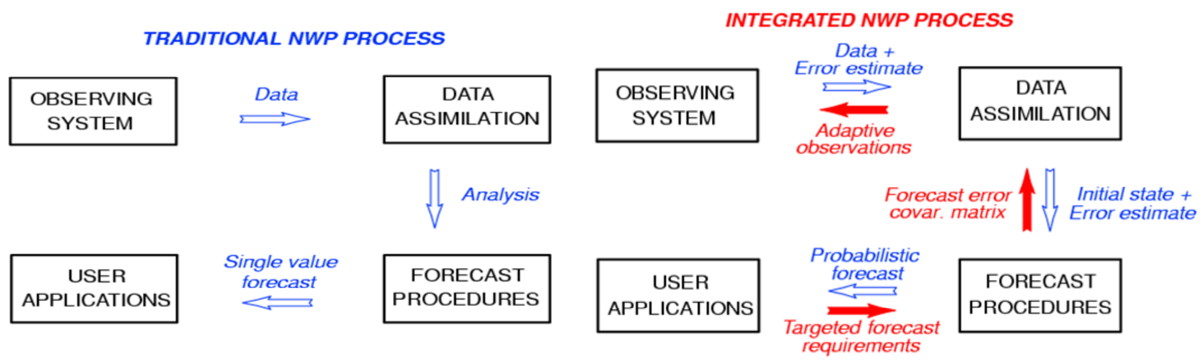

Figure 14 Traditional single-forecast based NWP paradigm vs. new integrated, ensemble-forecast based NWP paradigm.

Secondly, traditional single-forecast based NWP is a one-way system: observation determines model prediction but with no feedback from prediction to observation. However, in reality, error also exists in observation and impacts forecast accuracy in a flow-dependent way, i.e. it may affect a forecast significantly in one time (or over one region or for one weather system) but not much on another time (region or system). Therefore, a two-way system is desired: based on the estimation of potential forecast error over a region of interest, observation over a certain upstream region might need to be adjusted and improved accordingly too. Ensemble forecasting provides a bridge to make this two-way system possible. In an ensemble system with good spread-skill relation (see Part 6), the ensemble should be able to identify those weather systems associated with potential large errors and could also be used to trace the errors back to locate possible source regions in upstream using ensemble spread information. To improve the forecast, extra observations might be made over those source regions being identified. Or opposite action can be considered: less observation such as satellite data can be used in data assimilation when weather is calm to save resources. This process is known as adaptive or target observation technique, which is another new frontier of NWP (Palmer et. al., 1998; Bishop and Toth, 1999; Pu and Kalnay, 1999; Szunyogh et. al., 2000; Bishop et. al. 2001; Majumdar et. al., 2002). Obviously, this two-way approach is more sounding both scientifically and economically. Interactive two-way NWP system has been emphasized as one of the main goals in the on-going international joint research project GIFS (global interactive forecasting system).

Thirdly, although the estimation of forecast error distribution in the single-forecast paradigm is also possible via historical forecast data (see Part 2), it's, however, not flow dependent and doesn't reflect the true predictability or "error of the day", while the dynamical ensemble spread is flow-dependent, does reflect "error of the day", and, therefore, provides more situation-relevant information for better decision-making. No doubt, all the above three changes are reflecting a step forward in science and technology.

Lastly, since it's hard for human to think nonlinearly, forecasters will eventually not be able to keep up with NWP model's thinking someday when model forecast is accurately enough with continuing improvement of model and data quality. At that point, forecasters will have to mainly act like broadcasters or messengers by passively passing a model forecast to public or end-users with not much value added by forecasters' human role if single deterministic forecast is provided. While in the new era of ensemble forecasting, forecaster's human role will remain important and actively play a key role in the process of forecast-making in the following two folds. Although an ensemble provides multiple possible solutions, only one of them will eventually realize in reality. Therefore, a forecaster can act as an interactive "live" post-processing of the raw ensemble forecasts. For example, he might use other available data and the newest observations as well as his experiences to weigh each ensemble solution or filter out some "unlikely" members to possibly narrow forecast uncertainty. At the same time, forecasters should give proper physical interpretation of each distinct possible ensemble solutions to end-users for them to make better decisions and, therefore, provide an enhanced and value-added service to user community.

In one word, the core of switching from single-forecast paradigm to ensemble-forecast paradigm is to provide better service to user community to meet the variety of needs of our customers by producing a more accurate and complete rather than an overly simplified forecast to truly reflect the complex nature of weather, climate and water systems.

10. Future trend of ensemble development

Ensemble forecasting is still in its infant stage and a new frontier of NWP family and has many areas to be yet developed. Below are listed a few potential areas.

(a) Complimentary role of lower-resolution (low-res) ensemble and higher-resolution (hi-res) single forecast. Hi-res single run has smaller-scale detail spatial features and is more accurate for short range in general but lacks of uncertainty information, while low-res ensemble provides uncertainty information but is less accurate for short range and lacks of smaller-scale features. How to best combine low-res ensemble with hi-res single run is an important and practical issue which needs to be further explored (Roebber et. al., 2004; Kong et. al., 2006 and 2007). Hybrid Ensembling approach is one of such efforts, which superposes forecast variances from low-res ensemble on a single hi-res run by adding the difference between hi-res and low-res control forecasts to each ensemble members to improve the overall performance of an ensemble (Du, 2004). This method is found very useful and effective in improving an ensemble and has been operationally implemented for both global and regional EPSs at NCEP. Similar approach could also be applied to ETKF-based data assimilation (Gao et. al., 2007).

(b) Adaptive ensemble systems. Flow-dependent coupling among different scale EPSs might be desired. For example, a three-tier (global, regional and local) flow-dependent adaptive system could work as follows. A regional EPS varies its membership and model spatial resolution based on large-scale flow situation guided by a global EPS: if the global EPS projects less predictable flow or high-impact potential over the region, more members and higher resolution should be used to run the regional ensemble; otherwise, fewer members with coarser resolution could be used by the regional EPS. A global or larger-domain EPS should also determine where to run and which possibly distinct clusters/analyses to initiate a hi-res local EPS focusing on individual high-impact event or local flow for special needs (Molteni et. al., 2001). However, it should be carefully studied how to effectively couple multi-scale systems with each other to provide users maximum useful forecast information: does the smaller-scale EPS act purely as a downscaling tool of the larger-scale EPS by sharing same ICs, LBCs and perturbations or should each system have independent inputs to maximize diversity by using different ICs, LBCs and perturbations? Adaptive or manual-interactively manipulated perturbations in both IC and model configuration depending on specific weather systems or features of your interest, seasons and geographic regions might be found effective too in a local-scale ensemble forecasting (Homar et. al., 2006).

(c) Coupling with data assimilation (DA) process. With ETKF or similar technique, ensemble forecasting and DA can be coupled as one unified NWP component: ensemble provides truly flow-dependent background error to DA, while DA provides multiple analyses to initiate the ensemble model integrations with more realistic IC perturbations truly representing "error of the day" (see Part 3 (5)).

(d) Flow-dependent dynamical post-processing methods are needed as discussed in Part 5. How to combining ensemble approach with traditional statistical approach such as ensemble MOS to produce better forecast guidance should be a fruitful area for exploration too.

(e) Using ensemble in adaptive observation (Bishop and Toth, 1999; Bishop et. al. 2001; Majumdar et. al., 2002; Palmer et. al., 1998). How to effectively use ensemble information for adaptive observation or targeting is still in its infant stage and more research and field experiments are needed. Current limited efforts are mainly focusing on large-scale winter storms of relatively higher predictability. Using regional ensemble to target warm season convective systems and heavy precipitation is obviously important but a more challenging task (Du et. al, 2007b). When forecast error is propagating at a mixture of both phase-speed and group-speed, it's obviously a more complex situation to be investigated in targeting technique.

(f) Ensemble dynamics. A complete and mature theoretical framework is needed for ensemble forecasting which basically doesn't exist at the moment. It's little known and researched about the error dynamics. We need to precisely describe how error evolves and propagates in the governing equations of a model to establish a theoretical ensemble dynamics (Farrell, 1990; Nicolis, 2004; Vannitsem and Toth, 2002).

(g) To better serve user community, broad efforts are needed to understand, communicate and work with specific end-users to develop optimal economic value based decision-making strategies by scientifically incorporating forecast uncertainty information. The newly established American Meteorological Society's Ad Hoc Committee on Uncertainty in Forecasts is a good first step of this.

(h) Last but not least, a highly interactive, flexible and user-friendly visual display software capability needs to be developed for easy manipulating large amount of ensemble data and generating various ensemble products as well as performing ensemble verification. Help from software engineers is obviously necessary on this effort.

11. Summary

Although, given 100% accurate ICs and other conditions, the atmospheric system itself and numerical prediction models should be deterministic in theory. But, in reality, due to intrinsic uncertainties in IC and model configurations plus chaotic nature of nonlinear models, forecast uncertainty and predictability limit is a very real and important property of NWP. Without quantifying uncertainty, a forecast is incomplete. A complete forecast should be in a probabilistic distribution with uncertainty expressed but not in a single deterministic value. Ensemble forecasting is a dynamical approach to quantify forecast uncertainty. It's a relatively new but a rapidly developing branch of NWP and expected to become a main tool in weather, climate and water prediction in near future. Ensemble forecasting is most valuable when large uncertainty is around and forecasters don't know what solution to choose from in mainly high-impact events and has minimal value when weather is quiescent and highly predictable (although one still needs ensemble to identify such occasions). Currently, many major NWP centers around the world have already operationally implemented various ensemble prediction systems as part of their daily production although those EPSs are still in primitive and evolving stage. More and more forecasters and other users are lean toward using ensemble products instead of single deterministic model run nowadays. This trend will certainly continue in years to come. The ensemble-based NWP paradigm is superior to the single-value base forecast by providing flow-dependent uncertainty information besides an improved most probably single solution and taking observation and IC errors into account too. To complete a smooth transition from single-forecast based paradigm to ensemble based one, much effort is needed. Education, coordination and training should play a key role in this transition. The recent U.S. National Research Council report - Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts - is a good example of this effort. The currently going on UCAR COMET classroom training courses and local weather forecasting office's training workshop etc. are all good formats in training and education efforts.

The primary mission of ensemble forecasting is to reliably quantify forecast uncertainty and accurately describe a flow-dependent forecast error distribution to have the truth be always encompassed by ensemble cloud. In general, three types of product can be derived from an ensemble: a most probable single solution or consensus forecast, uncertainty measure and a distribution of all possible solutions. There are still many areas to be explored to maximize the utility of an ensemble, for example, how to better express and convey ensemble information to users in a comprehensive and easily-understandable way; complementary role between higher-resolution single model run and lower-resolution ensemble forecasts; better post-processing of ensemble forecasts including statistical and dynamical approaches, ensemble MOS, usage of re-forecasting or hindcast dataset and downscaling; and economic-value based decision-making process in using forecast uncertainty and probabilistic information.

To better accomplish the ensemble forecasting mission, a 3-dimentional type of EPS is needed by fully capturing all uncertainty sources from IC dimension, model-configuration dimension and history-memory dimension. Certainly, many needs to be further investigated and improved such as how to best couple with data assimilation in IC perturbation generating, stochastic physics perturbation and the value of history-memory dimension. Flow-dependent adaptive multi-EPS across the full spectrum of multi-scales is an area of interest and exploration. Theoretical ensemble or error dynamics is yet to be developed to fully understand how error evolves and propagates in the governing equations of a model.

Uncertainty is the only thing certain in the real world. Downstream application of meteorological ensemble forecasting is high in demanding and a rapidly growing area. Besides driving many downstream prediction systems like hydrology, air quality, storm surge, ocean wave, dispersion, geological prediction and electricity generation etc., adaptive/targeted observation is a special area of application within meteorology itself. Taking uncertainty into picture is a step forward in science and a way to better serve society and people. For the further reading about predictability of weather and climate, readers are referred to Palmer and Hagedorn (2006).

References

Alhamed, A., S. Lakshimivarahan and D. J. Stensrud, 2002. Cluster analysis of multi-model ensemble data from SAMEX. Mon. Wea. Rev., 130, 226-255.

Aligo, E.A., W.A. Gallus and M. Segal, 2007. Summer Rainfall Forecast Spread in an Ensemble Initialized with Different Soil Moisture Analyses. Wea. Forecasting, 22, 299–314.

Anderson, J. L, 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J. Climate, 9, 1518-1530.

Anderson, J. L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev., 129, 2884-2903.

Atger, F., 1999a. Tubing: An Alternative to Clustering for the Classification of Ensemble Forecasts. Wea. Forecasting, 14, 741–757.

Atger F., 1999b. The skill of ensemble prediction systems. Mon. Wea. Rev., 127, 1941–1953.

Barkmeijer, J., 1996. Constructing fast-growing perturbations for the nonlinear regime. J. Atmos. Sci., 53, 2838-2851.

Barkmeijer, J., M. Van Gijzen and F. Bouttier, 1998. Singular vectors and the estimates of analysis-error covariance matrix. Quart. J. Roy. Meteor. Soc., 124, 1695-1713.

Berrocal, V.J., A.E. Raftery and T. Gneiting, 2007. Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts. Mon. Wea. Rev., 135, 1386–1402.

Bishop, C. H. and Z. Toth, 1999. Ensemble transformation and adaptive observation. J. Atmos. Sci., 56, 1748-1765.

Bishop, C. H., B. J. Etherton and S. Majumdar, 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. Mon. Wea. Rev., 129, 420-436.

Bowler, N. E. 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. Tellus, 58A, 538-548.

Brankovic, C., T. N. Palmer, F. Molteni, S. Tibaldi and U. Cubasch, 2006. Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. Quart. J. Roy. Meteor. Soc., 116, 867-912.

Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. Mon. Wea. Rev., 75, 1-3.

Bright, D.R. and S.L. Mullen, 2002a. Short-Range Ensemble Forecasts of Precipitation during the Southwest Monsoon. Wea. Forecasting, 17, 1080–1100.

Bright, D.R. and S.L. Mullen, 2002b. The Sensitivity of the Numerical Simulation of the Southwest Monsoon Boundary Layer to the Choice of PBL Turbulence Parameterization in MM5. Wea. Forecasting, 17, 99–114.

Bright, D. R., S. J. Weiss, J. J. Levit, M. S. Wandishin, J. S. Kain and D. J. Stensrud, 2004. Evaluation of short-range ensemble forecasts during the 2003 SPC/NSSL Spring Program. Preprints, 22nd Conf. on Severe Local Storms, Hyannis, MA, P15.5.

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. DiMego and Z. Toth, 1995. Short-range ensemble forecasting: Report from a workshop, 25-27 July 1994. Bull. Amer. Meteor. Soc., 76, 1617-1624.

Brown, T. A., 1974. Admissible scoring systems for continuous distributions. Manuscript P-5235, The Rand Corporation, Santa Monica, CA, 22 pp. [Available from The Rand Corporation, 1700 Main St., Santa Monica, CA 90407-2138]

Buizza, R., 1994. Sensitivity of optimal unstable structures. Quart. J. Roy. Meteor. Soc., 120, 429-451.

Buizza R., A. Hollingsworth, F. Lalaurette and A. Ghelli, 1999a. Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. Wea. Forecasting, 14, 168–189.

Buizza, R., M. Miller and T. N. Palmer, 1999b. Stochastic representation of model prediction system. Quart. J. Roy. Meteor. Soc., 125, 2887-2908.

Cai, M., E. Kalnay and Z. Toth, 2002. Bred vectors of the Zebiak-Cane model and their application to ENSO prediction. J. Climate, 16, 40-56.

Cao, H, 2002. Memorial dynamics of systems and its applications. Chinese Geology Press, Beijing, China, 192pp. (in Chinese)

Chen, H. S., 2006. Ensemble prediction of ocean waves at NCEP. Proceedings of the 28th Ocean Engineering Conference in Taiwan, National Sun Yat-Sen University, November 2006, pp25-37.

Chen, J., J. Xue and H. Yang, 2003. Impact of physical parameterization schemes on mesoscale heavy rain simulations. Acta Meteorologica Sinica, 61, 203-218.

Cheung, K. W. C. and J. C. L. Chan, 1999a. Ensemble forecasting of tropical cyclone motion using a barotropic model. Part I: perturbations of the environment. Mon. Wea. Rev., 127, 1229-1243.

Cheung, K. W. C. and J. C. L. Chan, 1999b. Ensemble forecasting of tropical cyclone motion using a barotropic model. Part II: perturbations of the vortex. Mon. Wea. Rev., 127, 2617-2640.

Chien, F.C., Y.C. Liu and B.J.D. Jou, 2006. MM5 Ensemble Mean Forecasts in the Taiwan Area for the 2003 Mei-Yu Season. Wea. Forecasting, 21, 1006–1023.

Cui, B., Z. Toth, Y. Zhu, D. Hou and S. Beauregard, 2005. Statistical post-processing of operational and CDC hindcast ensembles. Preprints, 21th Conf. on Wea. Analysis and Forecasting and 17th Conf. on Numerical Weather Prediction. Washington, DC, Aug. 1-5, 2005, Amer. Meteor. Soc., 12B.2.

Du, J., 2004. Hybrid Ensemble Prediction System: a New Ensembling Approach. Preprints, Symposium on the 50th Anniversary of Operational Numerical Weather Prediction, University of Maryland, College Park, Maryland, June 14-17, 2004, Amer. Meteor. Soc., CD-ROM (paper p4.2, 5pp). [available online: http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html].

Du, J., 2005: Impact of Model Error and Imperfect Initial Condition Perturbations on Ensemble-Based Probabilistic Forecasts: UNPREDICTABLE SPOTS. Preprints, 17th Conference on Numerical Weather Prediction/21st Conference on Weather Analysis and Forecasting, Washington DC., Aug. 1-5, 2005, Amer. Meteor. Soc. (paper 15B.6) [available online http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html].

Du, J. and G. DiMego, 2008. A regime-dependent bias correction approach. 19th Conf. on Probability and Statistics, Jan. 20-24, 2008, New Orleads, LA, paper 3.2.

Du, J., G. DiMego, M. S. Tracton, and B. Zhou 2003. NCEP short-range ensemble forecasting (SREF) system: multi-IC, multi-model and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling (edited by J. Cote), Report 33, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 1161, 5.09-5.10.

Du, J., G. Gayno, K. Mitchell, Z. Toth and G. DiMego, 2007a. Sensitivity study of T2m and precipitation forecasts to initial soil moisture conditions by using NCEP WRF ensemble. 22nd WAF/18th NWP conference, Park City, UT, AMS.

Du, J., J. McQueen, G. DiMego, T. Black, H, Juang, E. Rogers, B. Ferrier, B. Zhou, Z. Toth and M. S, Tracton, 2004. The NOAA/NWS/NCEP short-range ensemble forecast (SREF) system: evaluation of an initial condition vs. multi-model physics ensemble approach. Preprints (CD), 16th Conference on Numerical Weather Prediction, Seattle, Washington, Amer. Meteor. Soc.

Du, J., S. L. Mullen and F. Sanders, 1997. Short-range ensemble forecasting of quantitative precipitation. Mon. Wea. Rev., 125, 2427-2459.

Du, J., S. L. Mullen and F. Sanders, 2000. Removal of distortion error from an ensemble forecast. Mon. Wea. Rev., 128, 2427-3351.

Du, J. and M. S. Tracton, 1999. Impact of lateral boundary conditions on regional-model ensemble predicion. Research Activities in Atmospheric and Oceanic Modelling (edited by H. Ritchie), Report 28, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 942, 6.7-6.8.

Du, J. and M. S. Tracton, 2001. Implementation of a 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.

Du, J., R. Yu, C. Cui and J. Li, 2007b. Using Mesoscale Ensemble to Predict Forecast Error Distribution and Target Observation. Acta Oceanologica Sinica, submitted.

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.

Ebisuzaki, W. and E. Kalnay, 1991. Ensemble experiments with a new lagged average forecasting scheme. WMO, Research activities in atmospheric and oceanic modeling. Report 15, 6.31-32.

Eckel, F. A., 2005. Plan for the joint ensemble forecast system (JEFS). Air Force Weather Agency.

Eckel. F. A. and C. F. Mass, 2005. Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting. Wea. Forecasting, 20, 328-350.

Eckel, F. A. and M. K. Walters, 1998. Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. Wea. Forecasting, 13, 1132-1147.

Ehrendorfer, M., R.M. Errico and K.D. Raeder, 1999. Singular-Vector Perturbation Growth in a Primitive Equation Model with Moist Physics. J. Atmos. Sci., 56, 1627–1648.

Epstein, E. S., 1969. Stochastic-dynamic prediction. Tellus, 21, 739-759.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. J. Appl. Meteor., 8, 985–987.

Errico, R., 1982. What is an adjoint model? Bull. Amer. Meteor. Soc., 78, 2577-2591.

Errico, R. and D. Baumhefner, 1998. Predictability experiments using a high-resolution limited area model. Mon. Wea. Rev., 115, 488-504.

Errico, R. and T. Vukicevic, 1992. Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. Mon. Wea. Rev., 120, 1644-1660.

Farrell, B. F., 1990. Small error dynamics and the predictability of atmospheric flows. J. Atmos. Sci., 47, 2409-2416.

Fischer, M., A. Joly and F. Lalaurette, 1998. Error growth and Kalman filtering within an idealized baroclinic flow. Tellus, 50A, 596-615.

Franz, K., N. Ajami, J. Schaake and R. Buizza, 2005. Hydrologic Ensemble Prediction Experiment Focuses on Reliable Forecasts, Eos, 86, No. 25.

Gao S.-T., 2007. Atmospheric mesoscale dynamics and forecasting methods. Meteorological Press, Beijing, China, 215pp. (in Chinese)

Gao, J., J. Du, M. Xue and K. Droegemeier, 2007. An efficient approach of ensemble Kalman filter for data assimilation. Geophys. Res. Lett., submitted.

Gao J. and M. Xue, 2007. An efficient dual-resolution approach for ensemble data assimilation and tests with simulated Doppler radar data. Mon. Wea. Rev., in press.

Gilmour, I. and L. A. Smith, 1997. Enlightenment in shadows. Applied nonlinear dynamics and stochastic systems near the millennium (edited by J. B. Kadtke and A. Bulsara), AIP, 335-340.

Glahn, H. R. and D. A. Lowry, 1972. The use of model output statistics in objective weather forecasting. J. Appl. Meteor., 11, 1203-1211.

Gneiting, T., A. E. Raftery, A. Westveld and T. Goldman, 2005. Calibrated probabilistic forecasting using Ensemble Model Output Statistics and Minimum CRPS estimation, Mon. Wea. Rev., 133, 1098-1118.

Gombos, D., J. A. Hansen, J. Du and J. McQueen, 2007. Theory and Applications of the Minimum Spanning Tree Rank Histogram. Mon. Wea. Rev., 135, 1490–1505.

Gray, M. E. B. and G. J. Shutts, 2002. A stochastic scheme for representing convectively generated vorticity sources in general circulation models. APR Turbulence and Diffusion Note, 285. Met Office, UK.

Greybush, S. J., S. E. Haupt and G. S. Young, 2007. The regime dependence of optimally weighted ensemble model consensus forecasts. Wea. Forecasting, (in revision).

Grimit, E. P., T. Gneiting T, V. J. Berrocal and N. A. Johnson, 2006. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. Quart. J. Roy. Meteor. Soc., 132, 2925-2942.

Grimit, E.P., and C.F. Mass, 2002. Initial Results of a Mesoscale Short-Range Ensemble Forecasting System over the Pacific Northwest. Wea Forecasting, 17, 192–205.

Grimit E. P. and C. F. Mass, 2007. Measuring the Ensemble Spread-Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results. Mon. Wea. Rev., 135, 203-221.

Hagedorn, R., T. M. Hamill and J. S. Whitaker, 2007. Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: 2-meter temperature. Submitted to Mon. Wea. Rev.

Hamill, T. M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Wea. Rev., 129, 550-560.

Hamill, T. M., 2006. Ensemble-based atmospheric data assimilation. Chapter 6 of Predictability of Weather and Climate, Cambridge Press, 124-156.

Hamill, T. M. and S. J. Colucci, 1997. Verification of Eta-RSM short-range ensemble forecasts. Mon. Wea. Rev., 125, 1322-1327.

Hamill, T. M. and S. J. Colucci, 1998. Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. Mon. Wea. Rev., 126, 711-724.

Hamill, T. M. and J. Juras, 2006. Measuring forecast skill: is it real skill or is it the varying climatology?. Quart. J. Roy. Meteor. Soc., 132, 2905-2923.

Hamill, T. M., R. Hagedon and J. S. Whitaker, 2007. Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: precipitation. Submitted to Mon. Wea. Rev.

Hamill, T. M., J. A. Hansen, S. L. Mullen and C. Snyder, 2004a. Meeting summary: workshop on ensemble forecasting in the short to medium range. Submitted to Bull. Amer. Meteor. Soc. [available online: http://www.cdc.noaa.gov/people/tom.hamill/cv.html].

Hamill, T.M., S.L. Mullen, C. Snyder, Z. Toth and D.P. Baumhefner, 2000. Ensemble Forecasting in the Short to Medium Range: Report from a Workshop. Bull. Amer. Meteor. Soc., 81, 2653–2664.

Hamill, T.M. and J.S. Whitaker, 2007. Ensemble Calibration of 500-hPa Geopotential Height and 850-hPa and 2-m Temperatures Using Reforecasts. Mon. Wea. Rev., 135, 3273–3280.

Hamill, T.M., J. S. Whitaker and S.L. Mullen, 2006. Reforecasts: An Important Dataset for Improving Weather Predictions. Bull. Amer. Meteor. Soc., 87, 33–46.

Hamill, T.M., J. S. Whitaker and X. Wei, 2004b. Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. Mon. Wea. Rev., 132, 1434–1447.

Hansen, J.A., 2002. Accounting for Model Error in Ensemble-Based State Estimation and Forecasting. Mon. Wea. Rev., 130, 2373–2391.

Hart, R. E. and R.H. Grumm, 2001. Using Normalized Climatological Anomalies to Rank Synoptic-Scale Events Objectively. Mon. Wea. Rev., 129, 2426–2442.

Harvey, L. O. Jr., K. R. Hammond, C. M. Lusk and E. F. Mross, 1992. The application of signal detection theory to weather forecasting behavior. Mon. Wea. Rev., 120, 863-883.

Hersback, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction system. Wea. Forecasting, 15, 559-570.

Hoffman, R. N. and E. Kalnay, 1983. Lagged average forecasting, an alternative to Monte Carlo forecasting. Tellus, 35A, 100-118.

Homar, V., D.J. Stensrud, J.J. Levit and D.R. Bright, 2006. Value of Human-Generated Perturbations in Short-Range Ensemble Forecasts of Severe Weather. Wea. Forecasting, 21, 347–363.

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

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

Houtekamer, P. L. and H. L. Mitchell, 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 196-811.

Im, J. S., K. Brill and E. Danaher, 2006. Confidence Interval Estimation for Quantitative Precipitation Forecasts (QPF) Using Short-Range Ensemble Forecasts (SREF). Wea. Forecasting, 21, 24–41.

Jankov, I., W.A. Gallus, M. Segal and S.E. Koch, 2007. Influence of Initial Conditions on the WRF–ARW Model QPF Response to Physical Parameterization Changes. Wea. Forecasting, 22, 501–519.

Jankov, I., W.A. Gallus, M. Segal, B. Shaw and S.E. Koch, 2005. The Impact of Different WRF Model Physical Parameterizations and Their Interactions on Warm Season MCS Rainfall. Wea. Forecasting, 20, 1048–1060.

Jolliffe, I. T. and D. B. Stephenson, 2003. Forecast verification: a practitioner's guider in atmospheric science. John Wiley & Sons, Ltd., England, 240pp.

Jones, M.S., B.A. Colle and J.S. Tongue, 2007. Evaluation of a Mesoscale Short-Range Ensemble Forecast System over the Northeast United States. Wea. Forecasting, 22, 36–55.

Kalnay, E, 2003. Atmospheric modeling, data assimilation and predictability. Cambridge University Press, 368pp.

Kalnay, E., 2007. A talk at Arakawa Symposium of 2007 American Meteorological Society annual meeting, San Antonio, TX.

Keil, C. and G.C. Craig, 2007. A Displacement-Based Error Measure Applied in a Regional Ensemble Forecasting System. Mon. Wea. Rev., 135, 3248–3259.

Kong, F., K. K. Droegemeier and N.L. Hickmon, 2006. Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparsion of coarse- and fine-grid experiments. Mon. Wea. Rev., 134, 807–833.

Kong, F., K. K. Droegemeier and N.L. Hickmon, 2007. Multiresolution ensemble forecasts of an observed tornadic thunderstorm system, Part II. Mon. Wea. Rev., 135, 759-782.

Kong, F., M. Xue, D. Bright, M. C. Coniglio, K. W. Thomas, Y. Wang, D. Weber, J. S. Kain, S. J. Weiss and J. Du, 2007. Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA hazardous weather testbed 2007 spring experiment. 22nd Conf. WAF/18th Conf. NWP, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.2.

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

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi and C. E. Williford, 2000. Multimodel ensemble forecasts for weather and seasonal climate. J. Climate, 13, 4196-4216.

Legg, T. E. and K. Mylne 2004. Early Warnings of Severe Weather from Ensemble Forecast Information. Wea. Forecasting, 19, 891-906.

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

Li, Z. and D. Chen, 2002. The development and application of the operational ensemble prediction system at National Meteorological Center. J. of Appl. Meteor. Sci. (in Chinese), 13, 1-15.

Li, X., M. Charron, L. Spacek and G. Candille, 2007. A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. Mon. Wea. Rev., (in press).

Lorenz, E. N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci., 20, 130-141.

Lorenz, E. N., 1965. A study of the predictability of a 28-variable atmospheric model. Tellus, 17, 321-333.

Lorenz, E. N., 1993. The Essence of Chaos. University of Washington Press, Seattle, 240pp.

Lu, C., H. Yuan, B.E. Schwartz and S.G. Benjamin, 2007. Short-Range Numerical Weather Prediction Using Time-Lagged Ensembles. Wea. Forecasting, 22, 580–595.

Majumar, S. J., C. H. Bishop and B. J. Etherton, 2002. Adaptive sampling with ensemble transform Kalman filter. Part II: filed program implementation. Mon. Wea. Rev.

Martin, A., V. Homar, L. Fita, J. M., Gutierrez, M. A.. Rodriguez and C. Primo, 2007. Geometrid vs. classical breeding of vectors: application to hazardous weather in the Western Mediterranean. Geophysical Research Abstracts, 9, European Geosciences Union.

Marzban, C. and S. Sandgathe, 2006. Cluster Analysis for Verification of Precipitation Fields. Wea. Forecasting, 21, 824–838.

Matsueda M., M. Kyouda, H. L. Tanaka and T. Tsuyuki, 2007. Daily Forecast Skill of Multi-Center Grand Ensemble. SOLA, 3, 29-32.

Mclay, J. M., C. H. Bishop and C. A. Reynolds, 2007. The ensemble-transform scheme adapted for the generation of stochastic forecast perturbations. Quart. J. Roy. Meteor. Soc., 133, 1257-1266.

Mittermaier, M. P., 2007. Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. Qart. J. Roy. Meteor. Soc., submitted.

Molteni F, R. Buizza, C. Marsigli, A. Montani, F. Nerozzi and T. Paccagnella, 2001. A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. Quart. J. Roy. Meteor. Soc., 127, 2069-2094.

Mu, M., 2000. Nonlinear singular vectors and nonlinear singular values. Scinece in China (D), 43, 375-385.

Mu M. and W. S. Duan, 2003. A new approach to study ENSO predictability: conditional nonlinear optimal perturbation. Chinese Sci. Bull., 48, 1045-1047.

Mu, M., W. S. Duan and B. Wang, 2003. Conditional nonlinear optimal perturbation and its applications. Nonlinear Processes in Geophysics, 10, 493-501.

Mu M. and Z. Zhang, 2006. Conditional nonlinear optimal perturbations of a two-dimensional quasi-geostrophic model. J. Atmos. Sci., 63, 1587-1604.

Mullen, S. L. and D. P. Baumhefner, 1994. Monte Carlo simulation of explosive cyclogenesis. Mon. Wea. Rev., 122, 1548-1567.

Mullen S. L. and R. Buizza, 2001. Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. Mon. Wea. Rev., 129, 638–663.

Mullen, S. L. and R. Buizza, 2002. The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system. Wea. Forecasting, 17, 173-191.

Mylne, K. R., Evans, R. E., and Clark, R. T., 2002: Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. Quart. J. Roy. Meteor. Soc., 128, 361-384.

Mullen, S. L. and J. Du, 1994. Monte Carlo forecasts of explosive cyclogenesis with a limited-area, mesoscale model. Preprints, 10th Confrence on Numerical Weather Prediction, Portland, Oregon, July 18-22, 1994, Amer. Meteor. Soc., 638-640.

Mullen, S. L., J. Du and F. Sanders, 1999. The dependence of ensemble dispersion on analysis forecast system: implications to short-range ensemble forecasting of precipitation. Mon. Wea. Rev., 127, 1674-1686.

Mureau, R., F. Molteni, and T. N. Parmer, 1993: Ensemble Prediction Using dynamically conditioned perturbations. Quart. J. Roy. Meteor. Soc., 119, 299-323.

Murphy, A. H., 1969: On the "ranked probability score." J. Appl. Meteor., 8, 988–989.

Murphy, A. H., 1971: A note on the ranked probability score. J. Appl. Meteor., 10, 155–156.

Murphy, A. H, 1985. Decision making and the value of forecasts in generalized model of the cost-loss ratio situation. Mon. Wea. Rev., 113, 362-369.

Mylne, K. R., Evans, R. E. and Clark, R. T., 2002. Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. Quart. J. Roy. Meteor. Soc., 128, 361-384.

National Research Council (NRC), 2006. Completing the forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts, National Academy Press, 112pp.

Nicolis, C., 2004. Dynamics of Model Error: The Role of Unresolved Scales Revisited. J. Atmos. Sci., 61, 1740–1753.

Nutter, P., D. Stensrud and M. Xue, 2004a. Effects of Coarsely Resolved and Temporally Interpolated Lateral Boundary Conditions on the Dispersion of Limited-Area Ensemble Forecasts. Mon. Wea. Rev., 132, 2358–2377.

Nutter, P., M. Xue and D. Stensrud, 2004b. Application of Lateral Boundary Condition Perturbations to Help Restore Dispersion in Limited-Area Ensemble Forecasts. Mon. Wea. Rev., 132, 2378–2390.

Oortwijn, J. and J. Barkmeijer, 1995. Perturbations that optimally trigger weather regimes. J. Atmos. Sci., 52, 3952-3944.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil and J. A. Yorke, 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus, 56A, 415-428.

Palmer, T. N., R. Gelaro, J. Barkmeijer and R. Buizza, 1998. Singular vectors, metrics and adaptive observations. J. Atmos. Sci., 55, 633-653.

Palmer, T. and R. Hagedorn (edited), 2006. Predictability of weather and climate. Cambridge University Press, 718pp.

Pu Z.-X. and E. Kalnay, 1999. Targeting observation with the quasi-inverse linear and adjoint NCEP global models: Performance during FASTEX. Quart. J. Roy. Meteor. Soc., 125, 3329-3338.

Raftery, A.E., T. Gneiting, F. Balabdaoui and M. Polakowski, 2005. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Mon. Wea. Rev., 133, 1155–1174.

Reynolds, C.A. and R.M. Errico, 1999. Convergence of Singular Vectors toward Lyapunov Vectors. Mon. Wea. Rev., 127, 2309–2323.

Richardson, D. S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. Quart. J. Royal. Meteor. Soc., 126, 649-668.

Roebber, P.J., D.M. Schultz, B.A. Colle and D.J. Stensrud, 2004. Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations. Wea. Forecasting, 19, 936–949.

Roulston, M. S., 2005. A comparison of predictors of the error of weather forecasts. Nonlinear Processes in Geophysics, 12, 1021-1032.

Roulston, M. S. and L. A. Smith, 2002. Evaluating probabilistic forecasts using information theory. Mon. Wea. Rev., 130, 3297-3319.

Rouston, M. S. and L. A. Smith, 2003. Combining dynamical and statistical ensembles. Tellus A, 55, 16–30.

Saha, S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H.M. Van den Dool, H.L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Peña, S. Lord, G. White, W. Ebisuzaki, P. Peng and P. Xie, 2006. The NCEP Climate Forecast System. J. Climate, 19, 3483–3517.

Schaake, J., K. Franz, A.Bradley and R. Buizza, 2006: The Hydrologic Enemble Prediction Experiment (HEPEX). Hydrology and Earth System Science, hessd-2006-0095.

Shutts, G., 2004. A stochastic kinetic energy backscatter algorithm for use in ensemble prediction systems. Technical Memorandum 449, ECMWF.

Sloughter J. M., A. E. Raftery, T. Gneiting and C. Fraley, 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. Mon. Wea. Rev., 135, 3209-3220.

Smith, L. A. and J. A. Hansen, 2004. Extending the limits of ensemble forecast verification with the minimum spanning tree. Mon. Wea. Rev., 132, 1522-1528.

Stensrud, D. J, 2007. Parameterization schemes: keys to understanding numerical weather models. Cambridge University Press, 479pp.

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

Stensrud, D. J., H. E. Brooks, J. Du, M. S. Tracton and E. Rogers, 1999. Using ensembles for short-range forecasting. Mon. Wea. Rev., 127, 433-446.

Stensrud D. J. and N. Yussouf, 2003. Short-Range Ensemble Predictions of 2-m Temperature and Dewpoint Temperature over New England. Mon. Wea. Rev., 131, 2510-2524.

Stensrud, D. J. and N. Yussouf, 2007. Reliable Probabilistic Quantitative Precipitation Forecasts from a Short-Range Ensemble Forecasting System. Wea. Forecasting, 22, 3–17.

Stensrud, D. J., N. Yussouf, M. E. Baldwin, J. T. McQueen, J. Du, B. Zhou, B. Ferrier, G. Manikin, F. M. Ralph, J. M. Wilczak, A. B. White, I. Djlalova, J. W. Bao, R. J. Zamora, S. G. Benjamin, P. A. Miller, T. L. Smith, T. Smirnova and M. F. Barth, 2006. The New England High-Resolution Temperature Program. Bull. Amer. Meteor. Soc., 87, 491–498.

Sutton, C., T. M. Hamill and T. T. Warner, 2006. Will Perturbing Soil Moisture Improve Warm-Season Ensemble Forecasts? A Proof of Concept. Mon. Wea. Rev., 134, 3174–3189.

Szungyogh, I., E. J. Kostelich, G. Gyarmati, B. R. Hunt, A. V. Zimin, E. Kalnay, D. J. Patil and J. A. York, 2004. A local ensemble Kalman filter for the NCEP GFS model. AMS annual meeting, Seattle, WA, Jan. 11-15.

Szunyogh, I, Z. Toth, R. E. Moss, S. J. Majumdar, B. J. Etherton and C. H. Bishop, 2000. The effect of targeted dropsonde observation during the 1999 winter storm reconnaissance program. Mon. Wea. Rev., 128, 3520-3537.

Talagrand, O., R. Vautard and B. Strauss, 1997. Evaluation of probabilistic prediction systems. Proceedings, ECMWF Workshop on Predictability, ECMWF, 1–25. [Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom.]

Teixeira, J. and C. Reynolds, 2006. The stochastic nature of physical parameterization in ensemble prediction: a stochastic convection approach. Geophysical Research Abstracts, 8, 01548, European Geosciences Union.

Teixeira, J., C. 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.

Tennant, W.J., Z. Toth and K.J. Rae, 2007. Application of the NCEP Ensemble Prediction System to Medium-Range Forecasting in South Africa: New Products, Benefits, and Challenges. Wea. Forecasting, 22, 18–35.

Thompson, P. D., 1957. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. Tellus, 9, 275-295.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. Hamill and J. S. Whitaker, 2003. Ensemble squared root filters. Mon. Wea. Rev., 131, 1485-1490.

Toth, Z. and E. Kalnay, 1993. Ensemble forecasting at NCEP: the generation of perturbations. Bull. Amer. Meteor. Soc., 74, 2317-2330.

Toth, Z. and E. Kalnay, 1997. Ensemble forecasting at NCEP: the breeding method. Mon. Wea. Rev., 125, 3297-3318.

Tracton M. S., J. Du, Z. Toth and H. Juang, 1998: Short-range ensemble forecasting (SREF) at NCEP/EMC. Preprints, 12th Conf. on Numerical Weather Prediction, Phoenix, Amer. Meteor. Soc., 269-272.

Tracton, M. S. and E. Kalnay, 1993. Ensemble forecasting at NMC: practical aspects. Wea. Forecasting, 8, 379-398.

Unger, D. A., 1985. A method to estimate the continuous ranked probability score. Preprints, Ninth Conf. on Probability and Statistics in Atmospheric Sciences, Virginia Beach, VA, Amer. Meteor. Soc., 206–213.

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

Wandishin, M.S., M.E. Baldwin, S.L. Mullen and J.V. Cortinas, 2005. Short-Range Ensemble Forecasts of Precipitation Type. Wea. Forecasting, 20, 609–626.

Wang, X. and C. H. Bishop, 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. J. Atmos. Sci., 60, 1140-1158.

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

Wang, X., T.M. Hamill, J.S. Whitaker and C.H. Bishop, 2007. A Comparison of Hybrid Ensemble Transform Kalman Filter–Optimum Interpolation and Ensemble Square Root Filter Analysis Schemes. Mon. Wea. Rev., 135, 1055–1076.

Wang, Y. and A. Kann, 2005. ALADIN-LAEF (Limited Area Ensemble Forecasting) at ZAMG: Status and Plan. 15th ALADIN Workshop [available online: www.cnrm.meteo.fr/aladin/meetings/Wk2005/WANG.pdf].

Warner, T. T., R. A. Perterson and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basis and potential serious limitation to regional numerical weather prediction. Bull. Amer. Meteor. Soc., 78, 2599-2617.

Warner, T. T., R-S. Sheu, J. F. Bowers, R. I. Sykes, G. C. Dodd and D. S. Henn, 2002. Ensemble simulations with coupled atmospheric dynamic and dispersion models: Illustrating uncertainties in dosage simulations. J. Appl. Meteor., 41, 488–504.

Wei, M., Z. Toth, R. Wobus and Y. Zhu, 2007. Initial perturbations based on the ensemble transform (ET) technique on the NCEP global operational forecast system. Tellus, (in press).

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

Weiss, S. J., J. S. Kain, D. R. Bright, J. J. Levit, G. W. Carbin, M. E. Pyle, Z. I. Janjic, B. S. Ferrier, J. Du, M. L. Weisman and M. Xue, 2007. The NOAA Hazardous Weather Testbed: collaborative testing of ensemble and convective-allowing WRF models and subsequent transfer to operations at the Storm Prediction Center. 22nd Conf. on Wea. Analysis and Forecasting and 18th Conf. on Numerical Wea. Prediction, Park City, UT, 6B.4.

Whittaker, J. S. and T. M. Hamill, 2002. Ensemble data assimilation without perturbed observation. Mon. Wea. Rev., 130, 1913-1924.

Whitaker, J. S. and A. F. Loughe, 1998. The relationship between ensemble spread and ensemble mean skill. Mon. Wea. Rev., 126, 3292-3302.

Wilks, D.S., 2004. The Minimum Spanning Tree Histogram as a Verification Tool for Multidimensional Ensemble Forecasts. Mon. Wea. Rev., 132, 1329–1340.

Wilks, D. S., 2006. Statistical methods in the atmospheric sciences. Academic Press, 627 pp.

Wilson, L.J., S. Beauregard, A.E. Raftery and R. Verret, 2007. Calibrated Surface Temperature Forecasts from the Canadian Ensemble Prediction System Using Bayesian Model Averaging. Mon. Wea. Rev., 135, 1364–1385.

Wobus, R. and E. Kalnay, 1995. Three years of operational prediction of forecast skill. Mon. Wea. Rev., 123, 2132-2148.

Woodcock, F. and C. Engel, 2005. Operational consensus forecasts, Wea. Forecasting, 20, 101-111.

Yang, S-C., M. Cai, E. Kalnay, M. Rienecker, G. Yuan and Z. Toth, 2006. ENSO bred vector in coupled ocean-atmospheric general circulation models. J. Climate, 19, 1422-1436.

Yuan, H., J. Du, J. McGinley, P. Schultz, B. Zhou, C. Lu, Z. Toth and G. DiMego, 2007a. Postprocessing of Precipitation Forecasts for New Configured NCEP Short-Range Ensemble Forecasting (SREF) System, 22nd WAF/18th NWP Conference, Park City, Utah, June 25-29, 2007, Amer. Meteor. Soc., 6B.2.

Yuan, H., X. Gao, S. L. Mullen, S. Sorooshian, J. Du and H. H Juang, 2007b. Calibration of Probabilistic Quantitative Precipitation Forecasts with an Articial Neural Network. Wea. and Forecasting (in press).

Yuan, H., S.L. Mullen, X. Gao, S. Sorooshian, J. Du and H.M.H. Juang, 2005. Verification of Probabilistic Quantitative Precipitation Forecasts over the Southwest United States during Winter 2002/03 by the RSM Ensemble System. Mon. Wea. Rev., 133, 279–294.

Yuan, H., S. L. Mullen, X. Gao, S. Sorooshian, J. Du and H. H Juang, 2007c. Short-Range Probabilistic Quantitative Precipitation Forecasts over the Southwest United States by the RSM Ensemble System. Mon. Wea. Rev., 135, 1685-1698.

Yussouf, N. and D. J. Stensrud, 2006. Prediction of Near-Surface Variables at Independent Locations from a Bias-Corrected Ensemble Forecasting System. Mon. Wea. Rev., 134, 3415-3424.

Yussouf, N., D.J. Stensrud and S. Lakshmivarahan, 2004. Cluster Analysis of Multimodel Ensemble Data over New England. Mon. Wea. Rev., 132, 2452–2462.

Zhang, F., 2005. Dynamics and Structure of Mesoscale Error Covariance of a Winter Cyclone Estimated through Short-Range Ensemble Forecasts. Mon. Wea. Rev., 133, 2876–2893.

Zhang, F., C. Snyder and J. Sun, 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. Mon. Wea. Rev., 132, 1238–53.

Zhang, Z. and T. N. Krishnamurti, 1999. A perturbation method for hurricane ensemble prediction. Mon. Wea. Rev., 127, 447-469.

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