A FDDA-BASED OPERATIONAL MESOSCALE ENSEMBLE ANALYSIS AND PREDITION SYSTEM (E-RTFDDA) AND A COMPARISON OF MM5 AND WRF

Yubao Liu[1], Mei Xu, Gregory Roux, Ming Ge, Rong Sheu, Joshua Hacker, Tom Warner and Scott Swerdlin
(National Center for Atmospheric Research, Boulder, Colorado)

1. INTRODUCTION

Aiming at regional weather-critical applications, a mesoscale ensemble analysis and prediction system is developed at NCAR/RAL. This system is built upon the NCAR RTFDDA (real-time four-dimensional data assimilation and forecasting), which is an "observation-nudging" based, multi-scale rapid cycling regional and local scale weather analysis and forecasting system (Liu et al., 2006). RTFDDA is an enhanced MM5 and WRF and it has been operated at 20+ regions across US and other global regions, providing real-time multi-scale current weather analyses and 0 – 48 h forecasts. The mesoscale ensemble system described in this paper is an extension of the RTFDDA system enhanced for probabilistic forecast using ensemble modeling approach. Although there are numerous additions to the RTFDDA, the core data analysis and forecasting engine of the ensemble members are essentially similar to RTFDDA, and thus we referred this ensemble system as to ensemble RTFDDA (E-RTFDDA).

In this paper, the philosophy of system design for completeness of ensemble schemes and flexibility for integration of evolving advances in the ensemble forecast and data assimilations in the research community, and some preliminary results of E-RTFDDA test runs are presented. MM5 and WRF model differences in the E-RTFDDA system are emphasized.

The system has been implemented on and is now being tested for real-time operation on a 272-processor Linux cluster, sponsored by the US Army High Performance Computing Management Program (HPCMP). The software engineering implementation is briefly discussed. Statistical verification of E-RTFDDA is an on-going effort that will not be reported on here because, so far, we only have very short E-RTFDDA forecasts.

2. REVIEW OF SYSTEM REQUIREMENTS

Mesoscale (10 - 2000 km) meteorological processes change rapidly in space and time. They are controlled by synoptic circulations and can be caused or affected greatly by local topography and underlying surface physical properties. Physical processes such as radiative transfer, cloud and precipitation, boundary layer mixing, etc., sometimes play dominant roles in shaping the regional weather and climate. Thus, unlike the global ensemble systems in which attention is mostly focused on initial conditions, where perturbations associated with the fast growing dynamics modes are added, mesoscale ensemble prediction systems need to address the uncertainties associated with other aspects of modeling systems. It is known that relatively large errors in mesoscale models often lead to an unrealistically small spread and large systematic errors in ensemble forecasts.

Apparently, mesoscale ensemble analysis and forecasts need a significant number of members (i.e. ensemble size) and multiple ensemble (perturbation) schemes to address the uncertainties of the aforementioned initial conditions, boundary conditions, external forcing

[1]*Corresponding author address*: Yubao Liu, NCAR/RAL, P.O. Box 3000, Boulder, CO 80307-3000; email: yliu@ucar.edu

and physics properties associated with mesoscale weather processes. Another challenge with E-RTFDDA is the requirements for a multi-scale simulation based on the WRF and MM5 nested-grid technology. Inherently, uncertainties of most critical aspects vary from one scale to another. One needs to balance the uncertainty sampling between different scales to address this issue with limited computing resources. Finally, mesoscale ensemble prediction and ensemble data assimilation are currently exploratory. Thus a great flexibility is needed in order for E-RTFDDA to be able to continuously advance and incorporate new achievements.

3. TRI-TIER E-RTFDDA SYSTEM

Unlike most existing operational mesoscale ensemble systems, E-RTFDDA was designed to provide a generic framework that 1) integrates observation data-processing, data assimilation and ensemble forecasting together; 2) can readily incorporate new advancements by the mesoscale ensemble prediction research community; 3) is a "multi-tier" system (Fig. 1) where a backend tier is included that is able to generate a library of exhaustive ensemble perturbation schemes/ members, a second tier is included to automatically select a set of ensemble schemes/members that are most appropriate for given weather regimes and for the weather variables that users' specific applications are most concerned with, and a third tier for ensemble execution and model output post-processing, and 4) can be rapidly relocated over the globe. This flexible framework not only permits an easy adaptation of the system for new applications, but it also allows modelers to employ it in both research and operation modes, which thus facilitates quick transfer of new research result into operational applications.

Like the 4DWX RTFDDA (Liu et al. 2006), E-RTFDDA is a continuous data assimilation and forecasting system. Each ensemble member, except for the ETKF-perturbed ones, is run with a data assimilation (based on the RTFDDA "observation-nudging" approach) period from the last cycle hour to the current time, and then forecasts proceed from the current "spun-up" 4-D analyses. For 6-hour cycling intervals, the data assimilation period is from -6 to 0 h. Note that E-RTFDDA data assimilation is continuous between cycles, and thus it generates 4-D continuous ensemble analyses from one cycle to another. The spread of the ensemble analyses can be considered as a sub-optimum estimate of the uncertainty of the analyses. As discussed elsewhere in this paper, other data-assimilation methods (e.g. 3DVAR, EnKF) can and will be integrated into the E-RTFDDA system to further enhance its data assimilation capabilities.

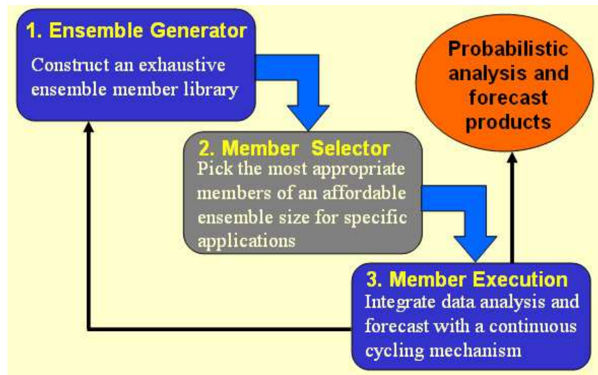

Figure 1 E-RTFDDA tri-tier diagram.

4. ENSEMBLE PERTURBATION SCHEMES

To accommodate the broad factors that control the mesoscale weather forecasts described in Section 1, multiple perturbation approaches that address the uncertainties in different aspects of mesoscale modeling systems are included in E-RTFDDA. This includes perturbations to model initial conditions (IC), lateral boundary conditions (LBC), model physical parameterizations (PP), and underlying land-surface (LS) characteristics. Previous research results on ensemble perturbations were adopted in E-RTFDDA, along with new approaches described herein. Note that E-RTFDDA is a rapidly-evolving system, and thus the perturbation schemes listed here are subject to further refinements and additions.

The multiple-model approach is addressed here by including both Penn State/NCAR MM5 and the Weather Research and Forecasting (WRF) models as E-RTFDDA core models. The perturbation schemes described below are applied for both models. Other mesoscale models may be added to E-RTFDDA in the future.

The LBC perturbation schemes of E-RTFDDA include three strategies. One is to derive LBCs from different global models, the second is to mimic potential phase errors of large-scale weather systems by translating the large-scale model output in symmetric directions by 30 km, and the last is to impose LBC perturbations based on model error statistic using WRF-3DVAR tools. The IC perturbation consists of perturbing observations and data analysis weighting with the WRF "observation-nudging" FDDA scheme (Liu et al. 2005), and an Ensemble Transform Kalman Filter (ETKF) approach (Wang and Bishop, 2003). The PP perturbation includes employing different parameterization schemes and perturbing the most sensitive and uncertain parameters in some physics schemes.

Table 1. Major physics schemes of WRF and MM5 included in E-RTFDDA physics perturbation

physics	**WRF**	**MM5**
Cumulus	Kain-Fritsch Betts-Miller-Janic Grell-Devenyi	Kain-Fritsch Grell Betts-Miller Fritsch-Chappell Kuo
Microphysics	Kessler Lin et al. WSM5 WSM6 Thompson et al. Ferrier	Hsie Dudhia ice Reisner I Reisner 2 Goddard Schwartz
Long-wave radiation	RRTM CAM GFDL	RRTM CCM2 Dudhia
Short-wave radiation	Dudhia Goddard CAM GFDL	Dudhia CCM2
PBL	YKU Meller-Yamada-Janic RUC GFS	MRF Blackadar Meller-Yamada-Janic Gyano-Seaman

Both MM5 and WRF models have several parameterization schemes for each major physical process, including land surface fluxes, boundary-layer mixing, long-wave and short-wave radiation transfer, sub-grid scale cumulus and grid-scale cloud microphysical processes. These physics schemes, among which some were adapted from other research or operational weather models, deal with the physics processes with varying degrees of complexity and different assumptions and algorithms. Sensitivity experiments using these schemes do not show any of them to be superior for all weather scenarios. Thus, running the models with varying physics schemes is considered a practical way for sampling the uncertainties in the model physics components. The PP perturbation approach is essentially a "multi-model" ensemble scheme. Table 1 lists the physics schemes available for MM5 and WRF that have been included in the E-RTFDDA ensemble perturbations.

Finally, the LS perturbations are constructed to take into account uncertainties in land-surface properties (e.g. albedo, vegetation, greenness factor, etc.). An off-line high-resolution land-surface data assimilation (HRLDAS) system (Chen et al. 2006) is used to generate an ensemble of land-soil thermal and moisture states. The roles of land-surface properties and thermal and moisture states in climate simulation are well recognized. For mesoscale weather forecasts, land-surface properties not only critically affect forecasts of surface weather and the diurnal evolution of boundary layer structure and processes, but also significantly impact model precipitation processes (Wu et al., 2007). Studies are conducted to identify, represent and simulate the uncertainties of sensitive physics processes and parameters in the HRLDAS model. The ensemble output of HRLDAS runs is then used to initialize a subset of E-RTFDDA forecasts.

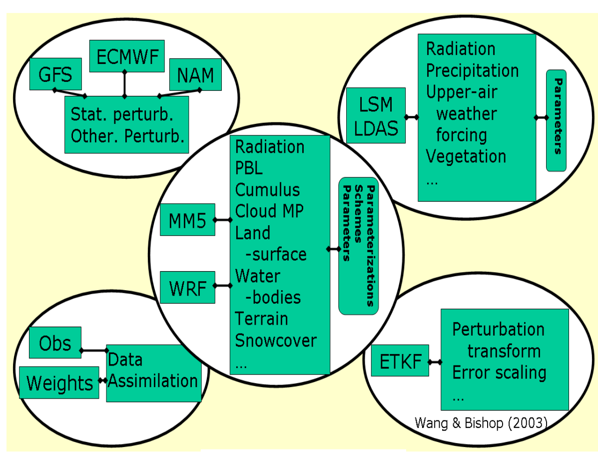

Figure 2 E-RTFDDA ensemble perturbation scheme.

Fig. 2 summarizes the ensemble perturbation approaches available in the current E-RTFDDA system. Note that, although it is not clear whether some specific perturbation approaches, e.g. observation perturbations and EtKF perturbation, can be run together to form a new perturbation member, perturbation permutations of different aspects of model systems can generally be combined to construct a new perturbation ensemble member. This results in more than 200 perturbation possible members using the perturbation schemes shown in Fig. 1.

An ensemble of 200+ mesoscale model members is prohibitively expensive at present. Fortunately, it is known that certain response functions (i.e., forecast variables of interest) are more sensitive to some factors than others, and these sensitivities may vary with the weather regimes. As a result, a member-selection tool is needed for choosing 1) the most important (relevant) members from the ensemble perturbation library according to the application goals and weather scenarios and 2) the ensemble size appropriate for the available computing resource. This task is very challenging due to the inherent complexity of mesoscale weather and the intelligence needed for representing users' application needs. Development of such tool is one of our on-going research topics. For now, a subjective evaluation is conducted to identify a subset of members from the ensemble perturbation scheme library. An algorithm of neural-net-based Self-Organized-Maps (SOMs) is under consideration, for automatically determining the most relevant members by running the system for a training-period.

5. IMPLEMENTATION FOR ATEC APPLICATIONS

E-RTFDDA is being implemented on a 69-node, 272-processor Linux cluster (HPC) to support routine testing at the ATEC test ranges. The ATEC E-RTFDDA system is designed to produce very high resolution model forecasts at the test ranges. For computational efficiency, multiple nested grids are configured, with a coarse mesh (Domain 1) of 30 km grid size covering about one third size of the CONUS, an intermediate grid (Domain 2, at 10 km grid size) spanning a few states, and a fine mesh of 230 by 230 km and a 3.3 km grid increment (Domain 3) covering the range and surrounding area. Similar grid configurations are set for the seven Army test ranges across the US, and the E-RTFDDA system can be switched to any range with a button-click. Figure 3 shows an example of the domain configuration for the Aberdeen Test Center (ATC). E-RTFDDA can also be relocated to other regions on the globe, and it is planned to build a user-friendly graphic user interface (GUI) for controlling E-RTFDDA jobs and monitoring ensemble-execution status.

The HPC is capable of running 50 - 55 members of the nested-grid ensemble, with 4 forecast cycles a day at 6-hr intervals, producing 36–48 h forecasts in each cycle when it is in full production mode. The system is planned to begin operational production in July 2007. A full suite of post-processing procedures, including ensemble calibration, verification, and graphic products, are now under development.

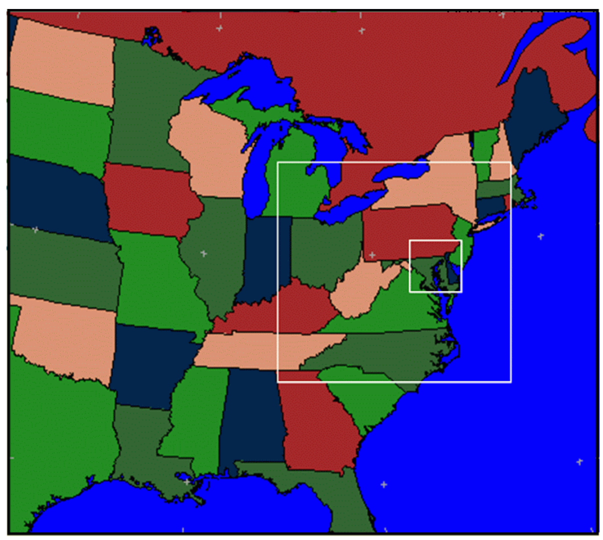

Figure 3 An example of ATEC E-RTFDDA model domain configuration: Aberdeen Test Center (ATC), Maryland.

6. CASE EXAMPLES

Three case studies were carried out using the E-RTFDDA system. The objective is to evaluate and compare the performance of the ensemble perturbation schemes. The cases are a strong wind event in a complex-terrain area in New Mexico, a blizzard event in Colorado and a weak synoptic event in the northeastern states. The model results presented here are preliminary and they are shown to illustrate very basic features of the E-RTFDDA forecasts.

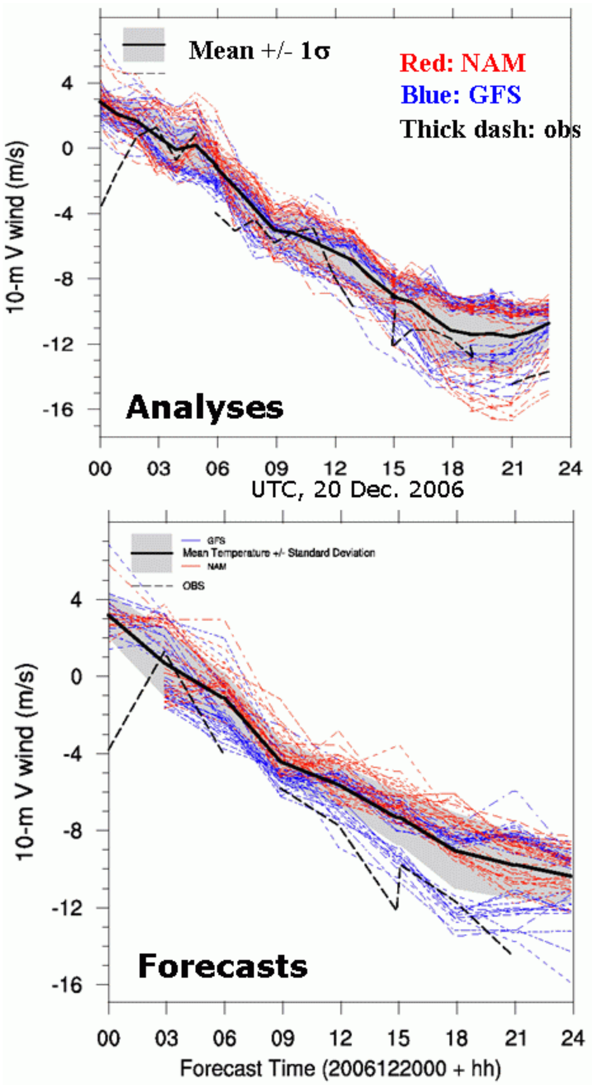

Figure 4 Ensemble (69 members) analyses (top) and forecast (bottom) of north wind components at DIA. Solid black line represents the ensemble mean. Thick dashed lines are observations. Grey-shaded region mark the value within two standard deviations from the ensemble mean. Forecasts using GFS output for lateral boundary condition is colored in blue, while those using NAM model output is colored in red.

Severe blizzard and exceptional cold-weather conditions occurred in Colorado during 20 – 22 December 2006. More than 80 cm of snow were observed in the Front Range. The E-RTFDDA system was run for this case with 84 members, among which 15 members were based on ETKF using lateral boundary conditions from the NAM model forecasts. The other members included PP, LBC and other types of IC perturbations. LS perturbations were not used.

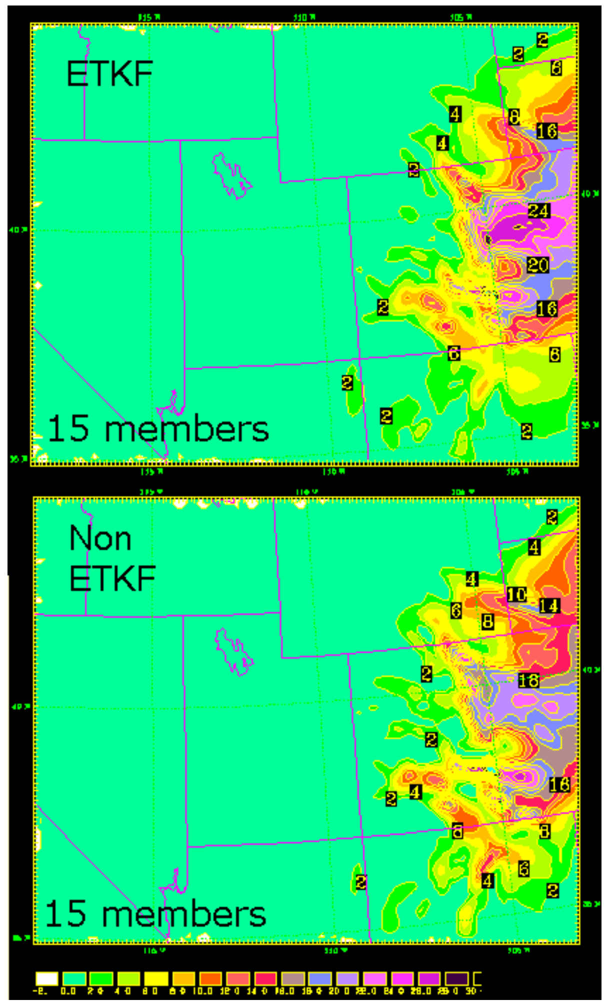

Figure 5 Ensemble mean of 12 hour accumulated precipitation (mm) from a blizzard in Colorado between 12Z Dec 20 and 00Z Dec. 21, 2006. Top panel shows 15-ETKF member mean and bottom mean of 15 non-ETKF members.

Fig. 4 shows the analyses from E-RTFDDA continuous 4-D data assimilation and 0 – 24 h forecasts (initialized from analyses at 00Z Dec 20) of surface north-south wind (V) at the Denver International Airport (DIA). E-RTFDDA well forecasted (bottom panel, Fig.4) intensification of the north winds up to 12Z Dec. 20 with a relatively small ensemble spread (width of the grey shade). In contrast, ensemble forecasts started to diverge from 15Z, and only a subset of ensemble members forecasted the strong north winds. A larger ensemble spread appears between 15 and 24Z, suggesting a smaller predictability for the strong north winds. Comparing to the forecast, ensemble FDDA analyses (the top panel, Fig.4) reduced the ensemble mean errors and the corresponding ensemble spread also.

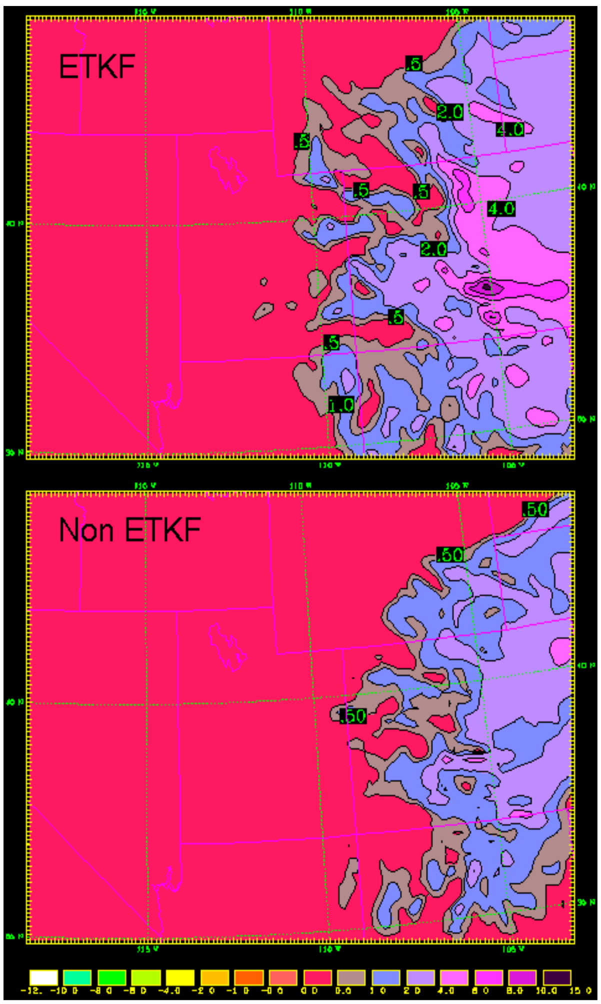

Figure 6 Same as Figure 4, but for ensemble spread (Standard deviation; mm).

Fig. 5 compares the ensemble mean of 12h precipitation accumulations from 00 to 12Z Dec. 20 of 15 ETKF members and 15 other (non-ETKF) members. In spite of a general similarity in the spatial distributions of the surface precipitation, the means of the ETKF members and the non-ETKF members differ significantly in terms of the precipitation core structures and precipitation amounts. Furthermore, the ensemble spreads of the precipitation forecasts (Fig. 6) of the two sets of ensemble members differ more dramatically. It is interesting to see that ETKF members produce stronger precipitation and larger spread than the non-ETKF members.

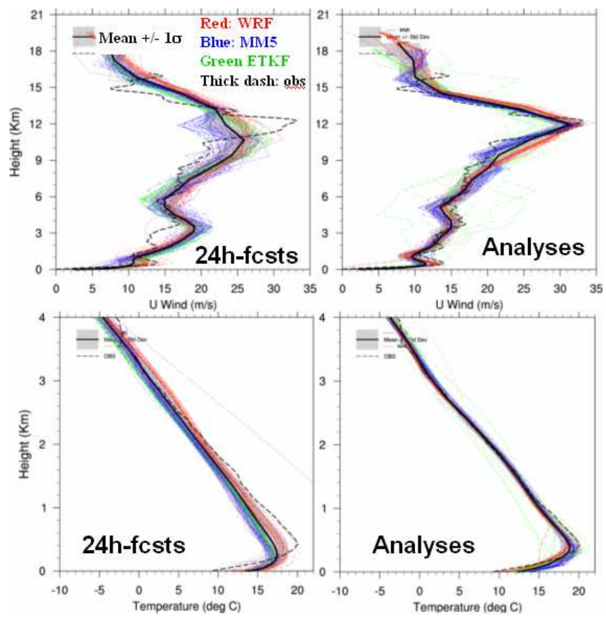

Figure 7 Same as Figure 4, but for vertical profile of 24h forecasts and analyses of U wind component (top panels) and temperature (bottom panels) valid at 12Z March 27, 2007 at the US Army Aberdeen Test Center. The observation (thick-dashed line) is valid at 10Z (2 hour ago). WRF members are plotted in red and MM5 members in blue. Note also that U winds are plotted from the ground to the model top, whereas the temperature is up to 4-km height only.

Forecasts of vertical wind profiles and associated uncertainty are valuable for many ATEC applications. Fig. 7 compares vertical profiles of the west-east wind (U) components and temperature (T) at an ATC site generated by WRF and MM5 ensemble members for a weak synoptic weather situation. The ensemble has 59 members. Like in the Colorado blizzard case, we can see a regime separation among the ensemble forecasts for the low-level jet below 1 km above ground level (bottom, Fig. 7). Ensemble spreads are larger in the layer where the ensemble mean presents larger error, which generally suggests that E-RTFDDA performs positively. Ensemble analyses valid at the same time show reduced errors of ensemble mean and smaller spread from the ensemble forecasts.

Another interesting feature in Fig.7 is that although MM5 and WRF have some common and/or very similar model physics schemes, the forecasts of ensemble members based on each model tend to cluster together. It is somehow odd that the ensemble mean is often located in the middle between the MM5/WRF ensemble clusters and the mean is closer to the observation. Very similar results are obtained from the ensemble runs of the Colorado blizzard case discussed earlier (i.e. when plot Fig. 4 coloring WRF and MM5 separately. Not shown). This indicates an advantage of multi-model ensembles based on varying dynamics formulation. The physics diversity and other perturbation approaches appear to be less dispersive, though important also.

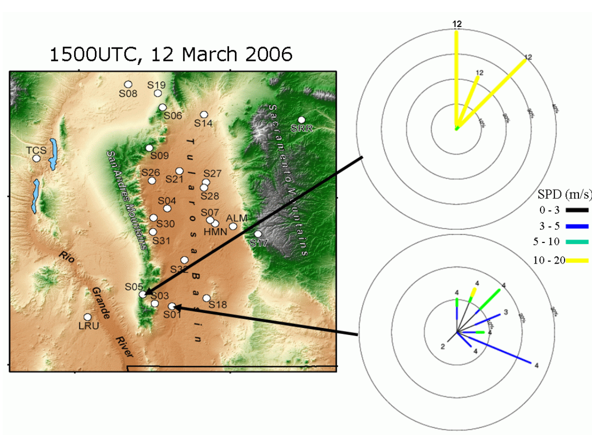

Figure 8 Comparison of ensemble forecasts of surface winds at a mountain pass station (SAMS 5) and a valley station (SAMS 1) for a strong wind case that occurred at 15Z 12 March 2006.

The last case to be shown is a strong wind event in the complex-terrain areas in south-central New Mexico. The observed wind speed at one surface station (Fig. 8) reached 28 m/s at 15Z 12 March 2006. E-RTFDDA forecasts were conducted for the event with 32 MM5 members. Winds roses of 15-h ensemble forecasts at a mountain pass (SAMS 5) and a valley station (SAMS 1) was plotted. At the mountain pass, winds are strong and less variable in direction, whereas in the valley, circulations are more transitional due to the mountain/lee waves propagation and channeling of winds. The E-RTFDDA ensemble appears to capture the wind regimes at the two stations reasonably well.

7. SUMMARY AND ON-GOING WORK

A WRF and MM5 based mesoscale ensemble analysis and prediction system (E-RTFDDA) has been developed at NCAR, and a version of the system has been deployed to produce real-time short-term (0 – 36 hour) probabilistic forecasts to support test operations at Army test ranges. Compared with other mesoscale ensemble systems reported previously, E-RTFDDA possesses the following unique features.

1) It contains multiple ensemble perturbation schemes that sample the uncertainties of broad aspects of mesoscale modeling systems.

2) It is a multiple-tiered system that separates the procedures of ensemble perturbation generation, ensemble scheme/member selection, ensemble execution and post-processing. It possesses the flexibility to allow upgrades and is adaptable to new applications. The system design allows the user to dynamically configure ensemble perturbation members according the the weather regimes and user's application needs and allows to conveniently incorporate new advances by ensemble forecast community.

3) It integrates data processing, data assimilation and ensemble forecasting into a general framework which allows model developers to test different data assimilation approaches, such as 3DVAR, "Newtonian relaxation"-based nudging FDDA and ensemble Kalman Filter (EnKF) schemes (e.g. Anderson 2003, Whitaker and Hamill, 2002).

4). It is a continuously cycling system which produces "spun-up" forecasts and 4-D analyses, and provide uncertainty information for both analyses and forecasts.

The performances of E-RTFDDA ensemble schemes were studied with case studies. E-RTFDDA output from ATEC operational runs over the eastern states, which is planned to begin in early July 2007, will be archived for systematic validation and verification. A suite of ensemble model products, including ensemble probability calibration and interfaces for driving secondary user application models (e.g., transport and diffusion) will be developed.

Finally, a hybrid data assimilation and ensemble forecast scheme, that combines "observation-nudging" with EnKF technology has been designed. The plan is to enhance the spatial weighting function of "observation-nudging" with the formulation of Kalman gain estimated using ensemble forecasts. This approach can be an effective way to extend the Kalman Filter into 4-d space. Because the accuracy of Kalman gain depends on the ability of an ensemble to forecast the probability distribution functions (PDFs) of the future atmospheric states, optimization of ensemble forecast performance can immediately leverage the accuracy of the hybrid data assimilation scheme and thus will continue to be a major research goal.

Acknowledgements: We highly appreciate assistance from and fruitful discussion with Scott Halvorson, Elford Astling, James Bowers from the US Army Dugway Proving Ground; Terri Betancourt, Wanli Wu, Laurie Carson, Wei Yu, Andrea Hahmann, Francois Vandenberghe and others in NCAR. Dr. Xuguang Wang provided us the ETKF software package.

8. REFERENCES

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev*., **131**, 634-642.

Chen, F., K. Manning, D. Yates, M. A. LeMone, S. B. Trier, R. Cuenca, and D. Niyogi, 2004: Development of high resolution land data assimilation system and its application to WRF. *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*. 11-15 January, Seattle, WA, American Meteorological Society, Boston.

Liu, Y., and co-authors, 2005: Implementation of observation-nudging based FDDA into WRF for supporting ATEC test operations. 2005 *WRF Users Workshop*, Boulder, Colorado, June, 2005.

Liu, Y., F. Chen, T. Warner, and J. Basara, 2006: Verification of a mesoscale data-assimilation and forecasting system for the Oklahoma City area during the Joint Urban 2003 Field Project. *J. Appl. Meteor. and Climatol.*, **45**, 912-929.

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

Whitaker, J and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.

Wu, W., Y. Liu, A. Hahmann, F. Chen and T. Warner, 2007: Estimating Parametric Uncertainties in a High-resolution Land Surface Data Assimilation System and Impact on Short-term Weather Prediction in WRF Model. *18th Conference on Numerical Weather Prediction*, AMS, June 25 -29, 2007. Park City, Utah.