1M.3 **SURFACE DATA ASSIMILATION USING AN ENSEMBLE KALMAN FILTER APPROACH WITH INITIAL CONDITION AND MODEL PHYSICS UNCERTAINTIES**

Tadashi Fujita*(**,***,****), David J. Stensrud(**), and David C. Dowell(*****)

**NOAA, National Severe Storms Laboratory, Norman, Oklahoma
***Sasaki Institute, University of Oklahoma, Norman, Oklahoma
****Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan
*****Cooperative Institute for Mesoscale Meteorology Studies, Norman, Oklahoma

1. INTRODUCTION

The ensemble Kalman filter (EnKF) (Evensen 1994) is considered to be a promising four dimensional data assimilation scheme, as it naturally incorporates inhomogeneity and time dependence into the background covariance. It also provides an ensemble of analysis fields that can be directly used as initial conditions for ensemble forecasting. Its potential to serve as one of the major assimilation schemes in the next generation is being widely explored in various contexts.

In the EnKF, the ensemble members are regarded as Monte Carlo samples from the probability distribution function of the atmospheric state, and the modification of the fields is determined based on the statistical properties of the ensemble members. Thus, it is important that the ensemble adequately represents the possible uncertainties of the forecast fields. Besides one of the forecast uncertainties that is commonly taken into consideration within EnKF, which comes from the uncertainty in the initial analysis field, recent studies of ensemble forecasting indicates that there is another important uncertainty which comes from the imperfect representation of atmospheric processes in the model (e.g. Stensrud et al. 2000).

In the present study, the assimilation of the hourly surface observational data is performed using an EnKF scheme. Our focus is on investigating the role of the uncertainty in both the initial and boundary conditions and the model physical process schemes in the assimilation process. To this end, the ensemble Kalman filter using an ensemble with diversity in initial and boundary conditions, an ensemble with diversity in the model physical process schemes, and an ensemble with diversities in both the initial and boundary conditions and model physical process schemes are performed for two cases and the results compared.

2. Experiment

a. general configuration

The EnKF experiments are performed for two cases: 1200 UTC 1 July 2003 – 1200 UTC 2 July 2003 (hereafter, the July case) and 1200 UTC 8 May 2003 – 1200 UTC 9 May 2003 (hereafter, the May case). In each case, assimilation by the EnKF is performed during the first 6 hours of the period, followed by an 18-hour ensemble forecast.

The forecast model used in this study is the non-hydrostatic fifth-generation Pennsylvania State

* *Corresponding author address:* Tadashi Fujita, NOAA, National Severe Storms Laboratory, 1313 Halley Circle, Norman, OK, 73069; e-mail: Tadashi.Fujita@noaa.gov

University-National Center for Atmospheric Research (PSU-NCAR) Mesoscale Model (MM5) (Dudhia 1993; Grell et al. 1994). The domain configuration is listed in table1.

During the assimilation period, the observations are assimilated every hour following the formulation of the ensemble square root filter (Whitaker and Hamill 2002). The observation types assimilated in the analysis step are potential temperature, u and v components of the wind, and dewpoint temperature. The observations are available at approximately 1500 – 1600 points, which are distributed over the land area in the domain. The modification of the model field is performed on potential temperature, dewpoint temperature, u, v , w (vertical component of wind), and perturbation pressure. The half-radius of the localization range is taken to be approximately 150 km in the horizontal direction and from the surface to 700 hPa in the vertical direction.

Table 1. Domain setup used in the experiment.

Domain center	(37.0N, 97.0W)
Map projection	Lambert conformal projection
Grid spacing (Assimilation domain) (BGM domain)	 30 km 90 km
Dimensions (Assimilation domain) (BGM domain)	 120x180x24 130x180x24
Model top	100 hPa
Full-sigma levels	1.00, 0.99, 0.98, 0.96, 0.93, 0.89, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05, 0.00

b. generating the ensembles

i) initial condition ensemble

The first approach to producing the ensemble is to use an identical model configuration, but different initial conditions. The initial conditions for this ensemble (hereafter, the BGM ensemble), which has 25 members, are generated by adding different perturbations to the NCEP global tropospheric analysis field of 1.0 x 1.0 degree resolution. The perturbations are produced by the breeding of growing modes (BGM) approach (Toth and Kalnay 1993), performed over 84 hours on a domain approximately three times as large as the final forecast domain (see table. 1). After the breeding cycles, 24 hour forecasts are performed on the BGM domain from the 25 initial conditions obtained from the cycles, in order to obtain the boundary conditions for the assimilation experiment.

An identical model physical process scheme, which is listed in the first row of table 2, is used in all the members of this ensemble both in the breeding cycles and in the assimilation experiment.

ii) physics ensemble

The second ensemble (the physics ensemble) is generated by applying different model physical process schemes on its members, while the identical initial and boundary conditions are used for all of the members (Stensrud et al. 2000). The physical processes varied are the convection, planetary boundary layer (PBL), radiation, and land surface schemes. These different model configurations are chosen to set the focus of the ensemble primarily upon uncertainties in the physical processes that affect the environment near the surface and convection that develops in the environment. Each scheme is used without any modifications, and is considered to provide a reasonable description of the physical processes. Twenty five various combinations of these schemes are used to form an ensemble with a diversity corresponding to the uncertainty in the model physics (see table 2).

iii) initial condition and physics ensemble

The third ensemble (the BGM+physics ensemble) is generated by adding different perturbations to the initial conditions, and applying different model physical process schemes. In order to incorporate the uncertainties in the initial and boundary conditions, the conditions for this ensemble are generated by breeding cycles performed on the coarser domain with an identical physical scheme. Exactly the same initial and boundary conditions as those from the BGM ensemble are used in this ensemble. During the assimilation experiment, different model physical process schemes are used in order to take into account the model physics uncertainty. The configurations of the physical schemes are the same with those of the physics ensemble, listed in table 2.

Table 2. Model configurations used in the physics and the BGM+physics ensembles. Member number 1 is the model configuration used in all the BGM ensemble simulations and the breeding cycles.

	Cumulus	PBL	Radiation	Surface
1	Grell	MRF	Cloud	NOAH
2	Grell	Blackadar	Cloud	Five Layer
3	Grell	Burk-Thompson	Cloud	Force Restore
4	Grell	Eta	Cloud	NOAH
5	Grell	Blackadar	CCM2	Five Layer
6	Grell	Burk-Thompson	CCM2	Force Restore
7	Grell	Eta	CCM2	NOAH
8	Grell	MRF	CCM2	NOAH
9	Kain-Fritsch	Blackadar	CCM2	Five Layer
10	Kain-Fritsch	Burk-Thompson	CCM2	Force Restore
11	Kain-Fritsch	Eta	CCM2	NOAH
12	Kain-Fritsch	MRF	CCM2	NOAH
13	Betts-Miller	Blackadar	CCM2	Five Layer
14	Betts-Miller	Burk-Thompson	CCM2	Force Restore
15	Betts-Miller	Eta	CCM2	NOAH
16	Betts-Miller	MRF	CCM2	NOAH
17	Kain-Fritsch	Blackadar	RRTM	Five Layer
18	Kain-Fritsch	Burk-Thompson	RRTM	Force Restore
19	Kain-Fritsch	Eta	RRTM	NOAH
20	Kain-Fritsch	MRF	RRTM	NOAH
21	Betts-Miller	Blackadar	RRTM	Five Layer
22	Betts-Miller	Burk-Thompson	RRTM	Force Restore
23	Betts-Miller	Eta	RRTM	NOAH
24	Betts-Miller	MRF	RRTM	NOAH
25	Kain-Fritsch	MRF	Cloud	NOAH

iv) control ensemble

The fourth ensemble consists of the BGM ensemble initial and boundary conditions, but does not incorporate any data assimilation. This ensemble is used to compare with the other ensembles to quantify the benefits of the ensemble Kalman filter data assimilation process.

3. Results and discussion

a. the July case (1200 UTC 1 July 2003 – 1200 UTC 2

July 2003)

i) synoptic situation

A brief summary of the synoptic situation (Fig. 1) is provided before discussing the results. The low pressure system located in the southeast US is a tropical depression resulting from the weakening of tropical storm Bill, which formed on 29 June near Yucatan and made landfall on the south coast of Louisiana around 1900 UTC 30 June. It is located near the border of Mississippi and Alabama at 1200 UTC 1 July, and moves toward the northeast during the experimental period, weakening in intensity.

Lee cyclogenesis is occurring over the region from the northern plains to southwestern Canada. This developing low pressure system slowly moves toward the east as a whole, shifting its center to the south. A trough stretches southward from the center of the low, and reaches to the eastern border of Colorado, and stays nearly at the same location during the experimental period. In the upper levels (not shown), a ridge is present over the Rocky Mountains with a closed low at 500 hPa over far southwestern Canada. The ridge shifts slowly eastward and begins to break down during the experimental period.

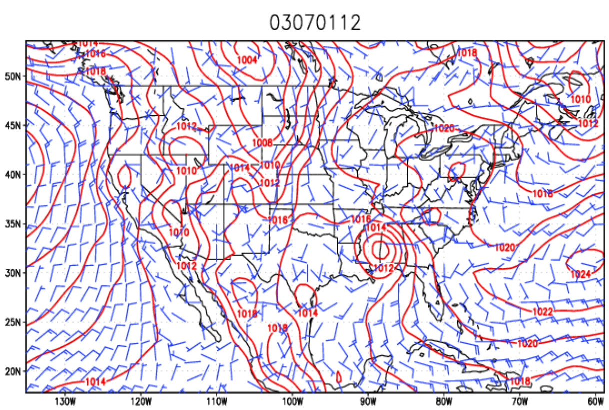

Fig. 1 : The NCEP global tropospheric analysis of mean sea level pressure (hPa) and wind at 10m above the ground (blue: full barb is 5 m/s) from 1200 UTC 1 July 2003. Mean sea level pressure (red lines) contoured every 2 hPa.

ii) spread distribution

The distribution of the spread of potential temperature at 2 m above the ground level at 1300 UTC 1 July before the assimilation of the observations shows large uncertainty in the background field (Fig. 2). After only one hour of forecast integration from the initial conditions, the BGM and the physics ensemble show different features from each other, corresponding to their different sources of diversity. In the BGM ensemble, large spread occurs near the tropical depression Bill in the southeast US and from southwest Canada to Montana just west of the low pressure system. Since the spread in this ensemble often comes from displacements of atmospheric systems, it is reasonable that the large spread is distributed around these low pressure systems. On the other hand, in the physics ensemble (Fig. 2b), large spread is distributed over the western US. This is believed to be due to sunrise not occurring yet over the western US. The delicate balance between outgoing and incoming long wave radiation, along with cloud effects, leads to large diversity in the values of potential temperature. We also can see large spread over the mountains in southern Idaho and Wyoming, and over the Great Lakes. This suggests that terrain effects and land surface model processes also play important roles in generating spread in this ensemble. Thus, the spread in the

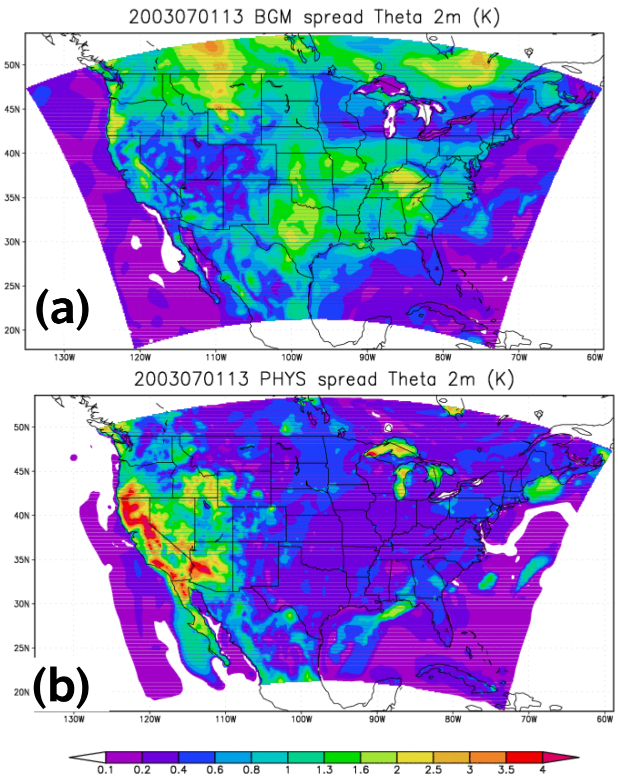

Fig. 2 : Ensemble spread of potential temperature (K) at 2m above the ground at 1300 UTC 1 July 2003. The results by (a) the BGM ensemble and (b) the physics ensemble are displayed.

physics ensemble is tied to the different responses of the physical process schemes to various conditions in terrain, radiation, and stability. In the BGM+Physics ensemble (not shown), the distribution of the spread is in a sense the sum of those of the BGM and the physics ensemble. The spread of this ensemble is large near the locations of the low pressure systems, and also over western US, and Great Lakes. This is reasonable because the diversity of this ensemble has origins both in uncertainties in initial conditions and in model physical process schemes.

iii) rms difference time sequence

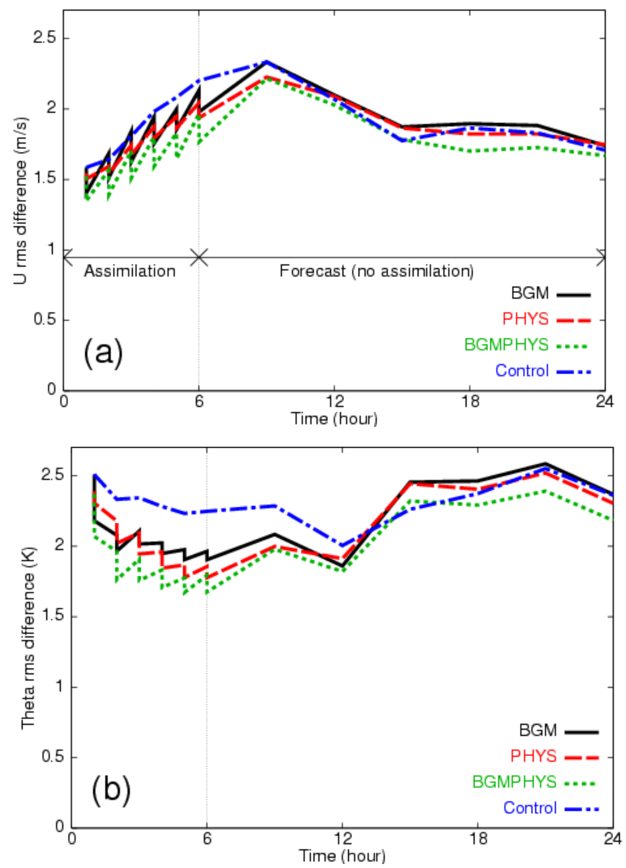

Fig. 3 : Time sequence of root-mean-square difference between the ensemble mean and the surface observation, averaged over the observation points of (a) u at 10 m above the ground and (b) θ at 2 m above the ground for the July case. The results by the BGM ensemble (black solid line), the physics ensemble (red dashed line), the BGM+physics ensemble (green dotted line), and the control ensemble (blue dash-dotted line) are displayed.

Over most of the assimilation period, the EnKF runs yield smaller rms difference between the ensemble mean and the observations, averaged over the observation points, than the control ensemble (Fig. 3). The BGM+physics ensemble gives the smallest rms difference of the three EnKF runs. In u and v, the reductions of the difference by the analysis steps in the

BGM ensemble is larger than those in the physics ensemble. Since the diversities are not directly tied to the dynamics of the model in the physics ensemble, the spread of dynamical variables tends to be small in this ensemble, resulting in relatively poorer fit to the observations. On the other hand, in thermodynamical variables θ and T_d, which often are sensitive to the model physical process schemes, the difference reduction in the physics ensemble is similar to that of the BGM ensemble.

During the forecast period after the assimilation, we find smaller rms difference in the EnKF runs than in the control ensemble for approximately 3 to 6 hours after the assimilation. This suggests that the assimilation by the EnKF has a potential to provide an improved surface environment for several hours after the assimilation period ends.

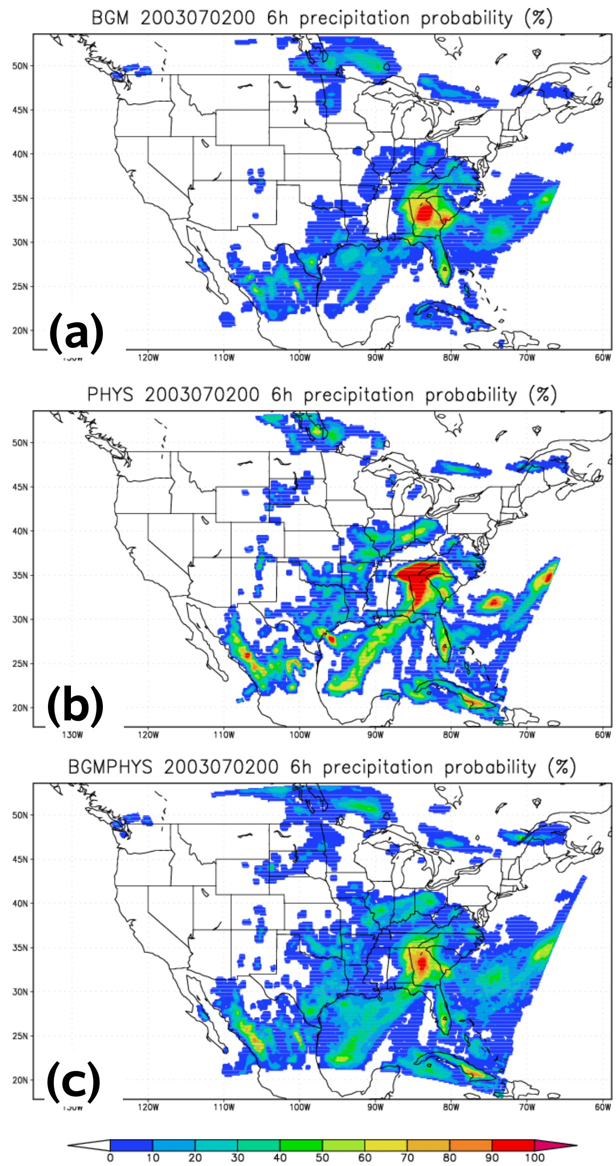

Fig. 4 : Probability (expressed in %) of accumulated rainfall from 1800 UTC 1 July 2003 – 0000 UTC 2 July 2003 with threshold 6 mm from (a) the BGM ensemble, (b) the physics ensemble, and (c) the BGM+physics ensemble.

iv) rain probability distribution

From the ensemble-forecast probability of accumulated rainfall over a 6-h period exceeding 6 mm, we infer that at 0000 UTC 2 July, the BGM ensemble (Fig. 4a) has a broader accumulated precipitation distribution than the physics ensemble (Fig. 4b). The physics ensemble produces higher probabilities in the southeastern US along the surface pressure trough that extends southwestward from the tropical depression than the BGM ensemble. However, also note the differences in the probabilities over Oklahoma and west Texas, New Mexico, western Mexico and South Dakota with the physics ensemble showing higher probabilities in these regions. While several of these regions have relatively small spread in 2 m θ at 1800 UTC in the BGM ensemble (not shown), suggesting that they are not being strongly perturbed by the bred modes, others are near local maxima of spread and yet no convection

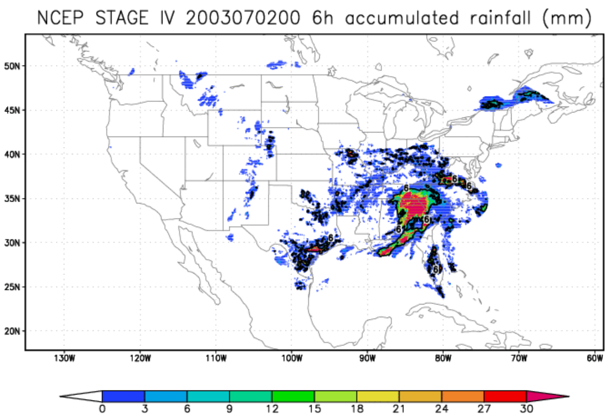

Fig. 5: Observed accumulated rainfall (mm) from 1800 UTC 1 July 2003 – 0000 UTC 2 July 2003 (NCEP stage IV at 4 km grid spacing). Black solid line indicates the contour of 6 mm accumulated rainfall.

develops. This comparison highlights one of the benefits of varying the physical process schemes in an ensemble. Different physical process schemes can yield vastly different responses for the same environmental conditions, and developing enough spread in the ensemble to overcome these physical scheme biases is not necessarily wise. Using a variety of quality physics schemes makes more sense.

The BGM+physics ensemble (Fig. 4c) gives a probability distribution over the widest area. It has a broader distribution of the precipitation probability than the physics ensemble, which comes from using the different initial conditions, but at the same time, its distribution also shows more fine-scale structures than that of the BGM ensemble, because of the diversity of the model physics process schemes.

Comparisons with the NCEP stage IV 6 hour accumulated precipitation distribution (Fig. 5) for 1800 UTC 1 July – 0000 UTC 2 July suggests that the physics ensemble captures many of the observed areas of precipitation. In particular, note the regions of precipitation in northern Missouri, Oklahoma, New Mexico, Colorado, and Wyoming that are suggested in the physics ensemble, but missing in the BGM ensemble. In the physics ensemble, we can also see that some of its members forecast the effect of Mexican monsoon, distributing rain probability along the northwest coast of Mexico. This is consistent with satellite imagery (not shown), which indicates deep moist convection over this region.

b. the May case (1200 UTC 8 May 2003 – 1200 UTC 9 May 2003)

i) synoptic situation

A surface low is located in western Kansas at 1800 UTC 8 May 2003 (Fig. 6). The low is moving toward the northeast, and reaches central to northern Kansas by 0000 UTC 9 May. A trough stretches out to the south from the center of the low, and a dryline is found along it. At 1200 UTC 8 May, the dryline runs from border of Colorado and Kansas southward to northeast Mexico. The northern end of the dryline moves east along with the movement of the low, reaching into central Kansas by 0000 UTC 9 May. In the northern part of the domain, a trough also extends northward from the low in Kansas to a shallower surface low in southern Canada.

In the upper levels, a large-scale trough is located over the western United States with southwestern flow across much of the southern plains states (not shown). A shortwave trough is moving across eastern New Mexico at 1200 UTC 8 May and

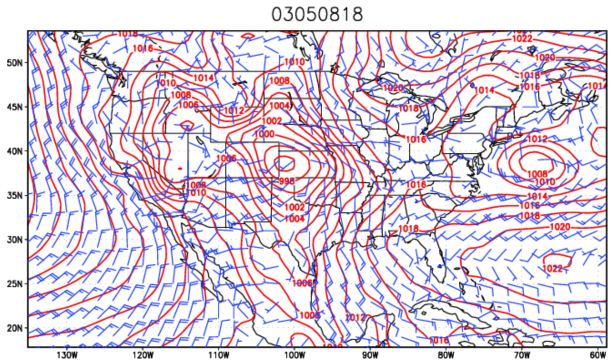

Fig. 6 : The NCEP global tropospheric analysis of mean sea level pressure (hPa) and wind at 10m above the ground (blue: full barb is 5 m/s) from 1800 UTC 8 May 2003. Mean sea level pressure (red lines) contoured every 2 hPa.

this trough reaches Kansas and Oklahoma by 0000 UTC 9 May, helping to initiate deep convection from Oklahoma northward to Iowa. Several tornadic supercell thunderstorms developed in Oklahoma and Kansas, tracking northeastward. There are a total of 48 reports of tornadoes, 64 reports of severe wind damage, and 175 reports of hail during the 24-h period beginning 1200 UTC 8 May. This day is much more active than the July case previously examined.

ii) rms difference time sequence

The rms difference between the ensemble mean and the observations, averaged over the observation points, of the BGM+physics ensemble is the smallest among the three ensembles over most of the assimilation and forecast period, reflecting the fact that the uncertainties both in the initial and boundary conditions, and in the model physical process schemes, are incorporated in this ensemble (Fig. 7). Improvement in the forecasts after the assimilation is seen in all of the three ensembles, especially for θ and T_d, and

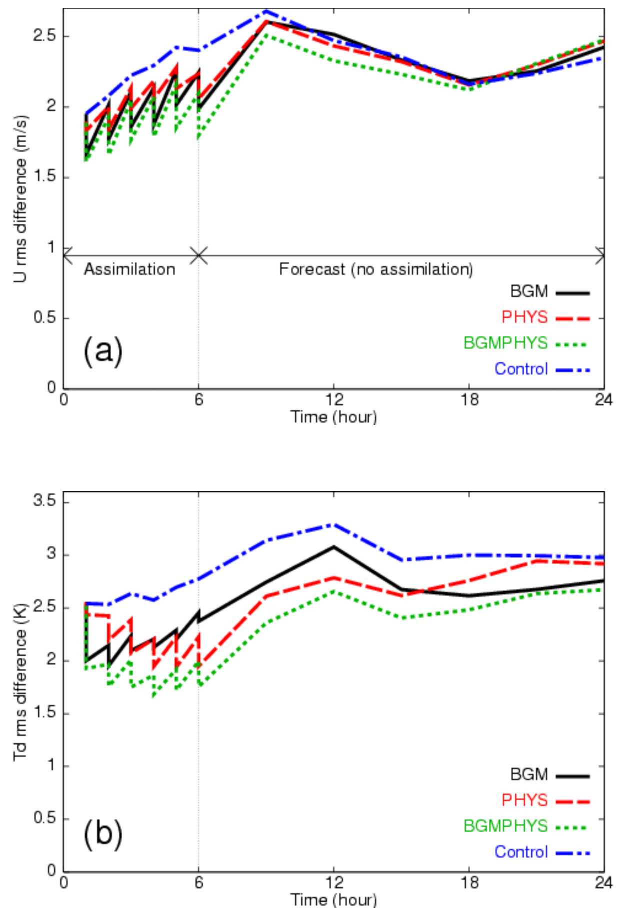

Fig. 7 : Time sequence of root-mean-square difference between the ensemble mean and the surface observation, averaged over the observation points of (a) u at 10 m above the ground and (b) T_d at 2 m above the ground for the May case. The results by the BGM ensemble (black solid line), the physics ensemble (red dashed line), the BGM+physics ensemble (green dotted line), and the control ensemble (blue dash-dotted line) are displayed.

the BGM+physics ensemble gives lower rms difference than the control ensemble over almost the whole of the experimental period. Even in this environment with stronger large-scale forcing, the physics ensemble often produces smaller values of rms difference than the BGM ensemble. This result is unexpected and further

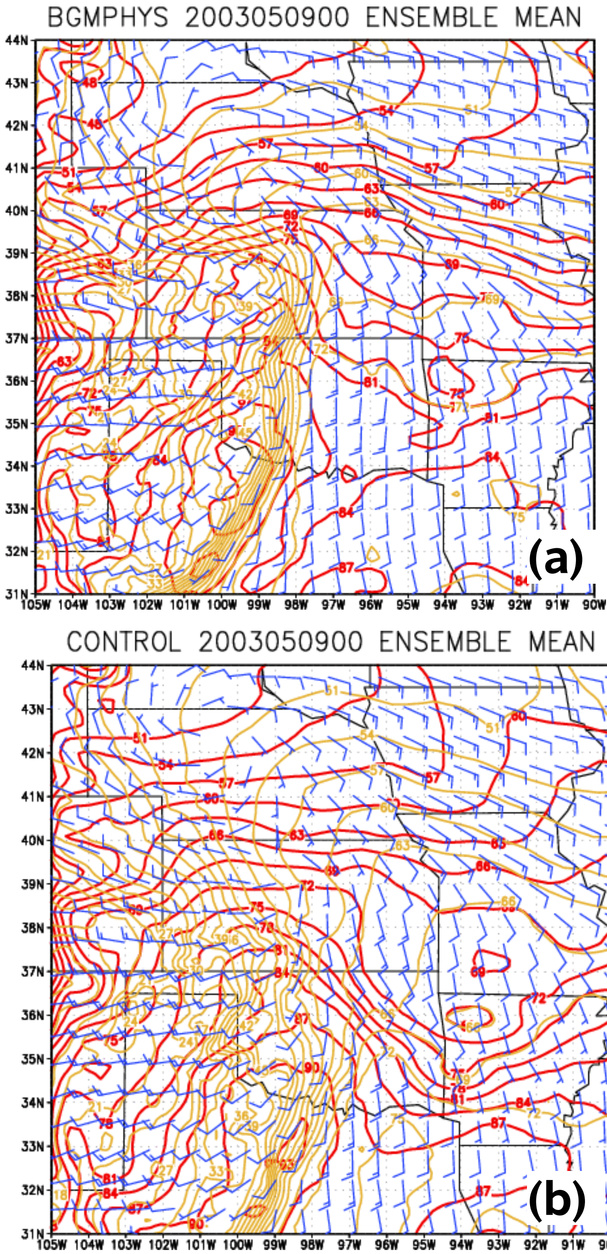

Fig. 8: (a) and (b) : Temperature at 2 m above the ground (°F), dewpoint temperature at 2 m above the ground (°F), and wind at 10 m above the ground (blue: full barb is 5 m/s) at 0000 UTC 9 May 2003 in central plains from (a) the BGM+physics ensemble and (b) the control ensemble. Temperature (red lines) contoured every 3 °F (1.7 °C), and dewpoint temperature (dark yellow lines) contoured every 3 °F (1.7 °C).

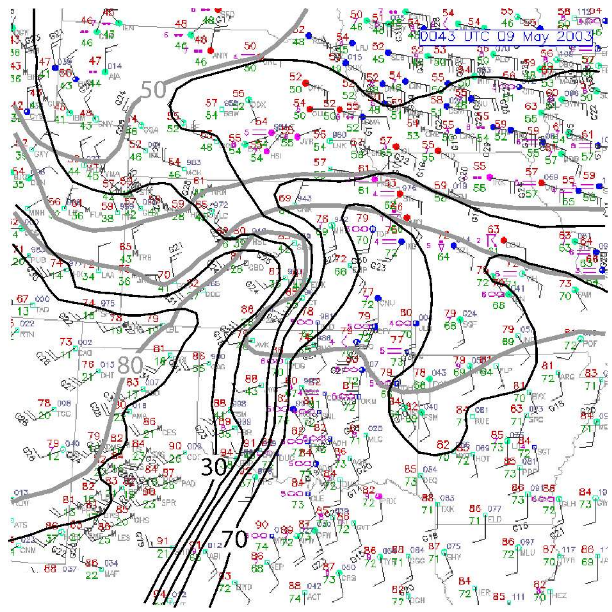

Fig. 9 : Surface observations corresponding to Fig. 8. Temperature contoured every 10 °F (5.6 °C) from 50 to 80 °F (gray lines) and dewpoint temperature contoured every 10 °F (5.6 °C) from 20 to 70 °F (black lines).

emphasizes the importance of physics diversity in ensembles.

iii) regional forecasts

Most of the severe weather reports for this day are in Oklahoma and Kansas, so an exploration of how the ensembles performed in this region is instructive. Improvement by the EnKF in the description of the dryline that runs from central Kansas southward to northern Texas is obvious at 0000 UTC 9 May, 6 hours after the end of the assimilation period (Fig. 8). Although a relatively large gradient in the dewpoint temperature is seen across western Kansas, Oklahoma, and northern Texas in the control ensemble, it is broad and the northern part in Kansas is hardly recognizable as a dryline. On the other hand, the tight, steep gradient of dewpoint temperature is clear in the BGM+physics ensemble even over the northern part of the dryline in Kansas. The dewpoint temperature drops from over 70 F on the east side to 20s F on the west side of the dryline, which is very consistent with the observations (Fig. 9). The occluded and warm front of the low also are more clear in the BGM+physics ensemble than in the control ensemble. A stronger gradient of temperature is seen in the BGM+physics ensemble extending eastward from northern Kansas with temperatures in the 70s F to the south of the front and in the 50s F to the north, while in the control ensemble the temperature gradient is weaker especially as you move eastward of the low center. Although the observations indicate an even stronger temperature gradient with the warm front than provided by the BGM+physics ensemble, the assimilation procedure clearly modifies the field toward the observations.

4.Summary

The assimilation of surface observations using an ensemble Kalman filter approach is evaluated. We produce ensembles in three different ways, by using different initial and boundary conditions (the BGM ensemble), by using different model physical process schemes (the physics ensemble), and by using both different initial and boundary conditions and different model physical process schemes (the BGM+physics ensemble). The three ensembles are compared in order to investigate the role of uncertainties in the initial and boundary conditions and physical process schemes in ensemble data assimilation.

Several characteristic features of the uncertainties incorporated in each ensemble are reflected in distribution of the spread. The property of the spread from the BGM ensemble, where diversity is introduced by the different perturbations in the initial and boundary conditions, often is associated with the location of baroclinic systems. Large spread is found in the locations that are sensitive to displacement of these baroclinic systems, such as near low pressure systems and along fronts that separate different air masses. On the other hand, in the physics ensemble, the origin of the spread is from the different model physical process schemes. Thus, large spread is expected to be seen in the locations where the atmosphere is sensitive to different responses of physical process schemes. As a result, different responses to terrain effects, the sensitive balance of radiation processes, and convective processes around precipitation regions are found to contribute to the spread. The distribution of the spread of these two ensembles shows different features as described above, which suggests that these ensembles cover different portions of the probability distribution function of the atmospheric state.

The spread of the BGM+physics ensemble has its origins in both the initial and boundary conditions and the model physical process schemes. As a result, the spread distribution of this ensemble reflects properties of the spread in both the BGM ensemble and the physics ensemble. The improvement from the control ensemble, in terms of the rms difference, is found to be the largest among the three ensembles. And the improvement lasts throughout the forecast period after the assimilation. Looking into the changes brought into the fields near the surface by the assimilation in detail, improvement from the control ensemble is clearly seen in the descriptions of characteristic features of atmospheric systems. The BGM+physics ensemble showed considerable improvement in the placement and intensity of the dryline and front that formed during the experimental period. These results indicate that the ensemble Kalman filter can successfully produce a reasonable environment within the lower troposphere by assimilating surface observations, especially when both initial condition and model physics uncertainty are included in the ensemble formulation.

references

Dudhia, J. 1993: A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev*., **121**, 1493-1513.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10 143-10 162.

Grell, G. A., J. Dudhia and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). *NCAR Technical Note*, NCAR/TN-398+STR, 117 pp.

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

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

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