COMPARISON OF THE REAL-TIME WRF AND MM5 FORECASTS FOR THE US ARMY TEST RANGES

Yubao Liu[1] and Thomas Warner

(National Center for Atmospheric Research /RAP, Boulder, Colorado)

1. INTRODUCTION

In the last few years, NCAR and the Army Test and Evaluation Command (ATEC) have jointly developed a real-time rapid-cycling FDDA and forecast (RTFDDA) system. This MM5-based system was deployed and has been running operationally at five Army test ranges, located in dramatically different geographic and climatological regions. As a part of the transition of the system from MM5 to WRF, semi-operational forecasts were conducted, twice-daily, for the Army ranges with both MM5 and WRF. The model forecasts are used to evaluate the WRF performance. Cases are presented here to illustrate the differences of the MM5 and WRF models for forecasting meso- (α, β and γ) scale weather systems, and for generating the model initial conditions as well. Statistical verifications are calculated by comparing the model outputs obtained in this spring season with the standard radiosonde and surface observations. The WRF model's performance in regions with very different climate and geographic features are compared.

2. MODELS AND EXPERIMENT DESIGN

To make fair comparisons between WRF and MM5, the two models are run in parallel for each case. The domain configurations of the models are set to be exactly the same, and the physics parameterizations were set as close as possible to each other, although it was impossible to make them identical because of differences in the physics schemes available in the two models at the time (MM5: version 3.6.2 and WRF: version 1.3). The basic model settings are listed in Table 1. There are two major differences in the physics settings. One is the land surface model: WRF makes use of the OSULSM, whereas MM5 uses the Noah LSM, which is a generalized and enhanced OSULSM scheme. The other is in the MRF PBL scheme. In MM5, the MRF PBL scheme that was recently modified to improve its computation of surface fluxes and mixing layer-depth (Liu et al. 2004, reported in this workshop also) is employed.

Both model forecasts are "cold-started" from the same NCEP operational model outputs -- the 40-km ETA-AWIPS analyses for the cases running within the CONUS, and the 1-degree GFS analyses for those outside of CONUS. The 3-hourly model forecast outputs of these NCEP models are used to derive the boundary conditions for the WRF and MM5 runs. The experiments were conducted for periods during spring 2004. By sharing the limited computing resources with other tasks, the models ran in episodes during the period. However, on each experiment day, two 24-hour WRF/MM5 forecasts were conducted, starting from 00 UTC and 12 UTC, respectively.

Table1: Comparison of Basic model settings

(Settings)	**WRF**	**MM5**
V-coord.	Mass core (36)	Sigma (36)
PBL	MRF	Modified MRF
LSM	OSULSM- 4 layers	Noah - 4 layers
Radiation	RRTM/Dudhia	RRTM/Dudhia
Microphysics	Simple ice (NCEP)	Mixing phase
CUP	KF-new	Grell

Three sets of parallel MM5 and WRF model forecasts were conducted to examine MM5 and WRF differences on different scenarios.

1). The first set (Set 1) is designed to compare the MM5 and WRF for meso-α scale weather forecasts over the mountainous western states. The domain is centered at Utah. The model grid spacing is set at 30 km. The model domains are 2900 km x 2500 km. A total of 28 forecasting days in April and May are collected for the statistical verification.

2). The second set (Set 2) was run with a smaller domain but with higher resolution. The model domain size is 600 km x 600 km, covering the state of Utah and some parts of its neighbor states. The model grid spacing is 3 km. This experiment started in late April, but only the recent two weeks of forecasts (between May 9 and 23, 2004) are available for statistic verification. Since high-resolution simulations take the best advantage of the WRF model, case studies from these model runs were carefully conducted to compare the WRF and MM5 models.

3). Finally, the third set (Set 3) is composed of a series of WRF model forecasts over each of the Army test ranges where the current MM5-based

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

NCAR/ATEC RTFDDA is running. This comparison allows us to investigate the WRF model performance variation for mesoscale forecasting in different climatological zones and geographical environments. Due to the limitation of computing resources, forecasts were conducted on the coarse-resolution domains, with a grid-spacing of 30 km and a domain size of 2900 km x 2500 km. Forecasts are performed for the Cold-Regions Test Center (CRTC) in Alaska, the Dugway Proving Ground (DPG) in Utah, the White Sand Missile Range (WSMR) in New Mexico and Yuma Proving Ground (YPG) in Arizona. The domain used for the forecasts for the Aberdeen Test Center is centered at Oklahoma. Note that these domains are fairly big, each of which covers about 1/2 of the CONUS region. Two weeks of model verification statistics (March 4 – 18, 2004) will be presented for this case.

3. MODEL RESULTS

3.1 Meso-α scale WRF-MM5 comparison

The WRF and MM5 models are first compared for forecasts of meso-α scale weather systems with the model output from the experiments in Set 1. The two models are run with exactly the same model domain, and the initial conditions and the lateral boundary conditions of both models are interpolated from the same 3-hourly ETA outputs on the 40-km AWIP grids. The location and terrain features of the model domains are shown in Fig.1. Obviously, the domain contains many interesting underlying features that affect mesoscale weather processes, including ocean to the west, the Rockies in the middle and flat plains to the east. As described above, a total of 28 days of model forecasts from the WRF and MM5 parallel runs during April and May 2004 are collected for the statistics calculation.

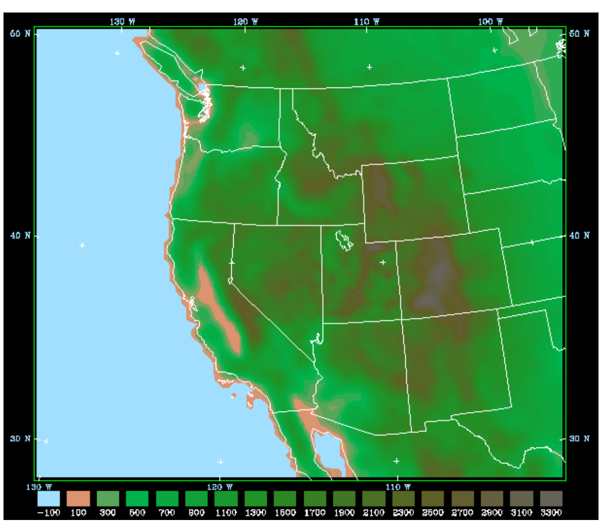

Fig. 1. Domain location and terrain distribution for the experiments with Set 1.

In the domain are 32 standard rawinsonde stations, where soundings are observed twice a day, at 00 UTC and 12 UTC, and 1500 - 2000 hourly surface observations with the number varying with the time of day. The model forecast errors are calculated by interpolating the model output to the observation location, and then comparing to the observation. Upper-air verification were calculated for the model analyses and forecasts valid at 00 and 12 UTC only, whereas the surface statistics are calculated hourly. The model errors are averaged over the domain for the 28 days of forecasts, according to the GMT hour of day and forecast leading time.

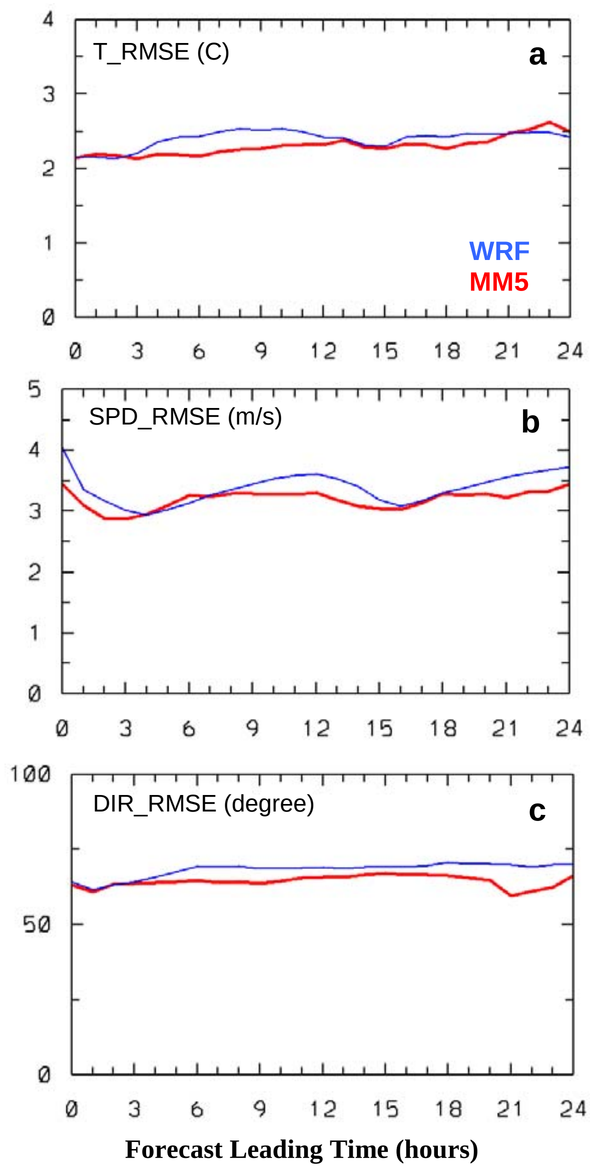

Fig. 2 Domain average of RMSE errors of the surface temperature (a), wind speed (b) and wind directions (c) forecasts of WRF (thin-blue) and MM5 (thick-red) with respect to forecast time, for the 28 forecast days in the April and May, 2004.

Fig. 2 compares the errors of the surface temperature and wind forecasts of the MM5 and WRF models, and the error evolution with forecast time. Apparently, the two models possess very similar skills in the forecasts of the surface variables. MM5 performs a little bit better on both surface wind and temperature forecast, partially due to its more accurate Noah land surface model and partially due to its modified MRF PBL scheme that significantly improves the daytime surface flux computation and mixing-layer depth estimation (Liu et. al, 2004). The forecast error generally grows with increasing forecast time. The RMSE errors of the temperature grow by ~0.4 C in 24 hour in both models. Evident oscillations can be observed in the error growth of temperature and wind speed in response to the diurnal changes of surface heating/cooling and PBL mixing. It can be observed that significant wind speed errors exist in the initial conditions of both models, as a result of interpolation from the ETA analyses. These errors appear to spin down quickly in the first few hours of the forecasts, in both models. Finally, the two models perform pretty evenly in the surface moisture forecasts (not shown).

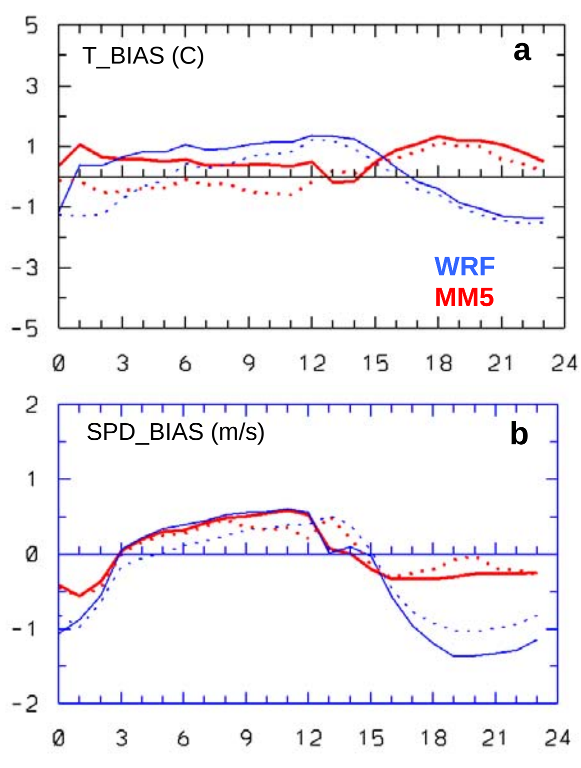

Fig. 3. Domain average of BIAS errors of surface temperature (a) and wind speed (b) of 0-12 (solid lines) and 12 -24 h (dotted lines) forecasts from WRF (thin-blue) and MM5 (thick-red) models, for the 28 forecast days in the April and May, 2004.

Averaging the errors of the twice-daily forecasts that are conducted each day at 12 hour intervals (00 and 12 UTC), according to the forecast time, masks some important diurnal components of the model forecast errors. Because the model surface forecasts are strongly affected by the diurnal evolution of surface heating and PBL development, it is necessary to know how well the WRF and MM5 models handle these processes. Fig. 3 plots the diurnal evolution of the BIAS errors of the temperature and wind speed forecasts at the surface. The 24 hour forecasts from each model are broken into two parts: (0 - 12h) nowcasts and short-range (12 – 24h) forecasts. The two models differ from each other greatly in the daytime period, with a warm bias in MM5 and a cold bias of a similar magnitude in WRF. The WRF model appears to significantly underestimate the intensity of the day-time surface wind, with a maximum bias of ~1.5 m/s occurred at noon time. In contrast, the biases are much smaller in MM5, thanks again to the improved MRF PBL scheme in the MM5 model. It is interesting to see that the short-range (12 – 24 h) forecasts have somewhat smaller bias than the shorter (0-12h) nowcasts for both temperature and wind speed. This may be related to the model dynamic and diabatic "spin-up" from the "cold-start" initialization. Nevertheless, short-range forecasts generally show larger RMSE errors than the nowcasts (not shown). Finally, during the night-time, the surface forecast errors of the two models are rather similar and both present a little bit of warm bias and an overestimate of wind speeds.

Upper-air verification statistics for the same period are calculated, and the RMSE errors of temperature, humidity and wind from the MM5 and WRF forecasts are compared in Fig. 4. First, some discrepancies can be seen between the initial conditions (0-h forecasts) of the two models, although both are interpolated from the same ETA analyses. These differences exist in most layers, and the magnitudes are comparable to the differences between the 12-h forecasts of the two models. Unlike those in the surface forecasts, the forecast errors in the upper-air grow with time in the two models are very similar, in terms of both the error sizes and the height distribution. Both models have the largest error growth of moisture and temperature in the lower troposphere and a relatively uniform wind error growth at different heights. However, the two models do differ in many details. The MM5 forecasts in the lower troposphere are more accurate than WRF, especially for the temperature and moisture fields, whereas WRF shows better performance in the upper-levels. The WRF model predicts better the winds in the layer between 700 and 200 hPa, with a RMSE error of the vector wind differences 0.5 m/s less than that of the MM5.

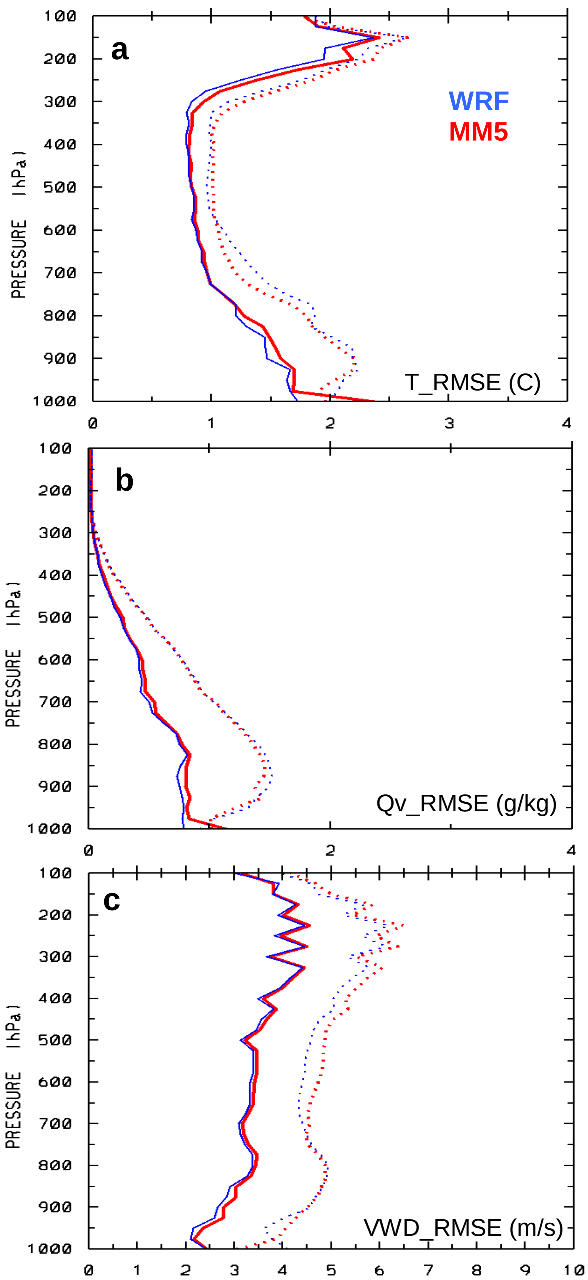

Fig.4 Domain-average RMSE errors of upper-air temperature (a), specific humidity (b), and vector wind differences (c) of WRF (thin-blue) and MM5 (thick-red) 0-h (solid) and 12-h forecasts (dotted) for the 28 forecast days in the April and May, 2004.

3.2 Fine-scale WRF-MM5 comparison

The forecast experiments in Set 2 are used to illustrate the difference of the MM5 and WRF models for simulating meso- β and γ scale weather systems. The forecasts are conducted on a single high-resolution domain with a grid-spacing of 3 km. The model forecast of a snowstorm event, occurring on April 19, 2004, is selected here for comparing the two models. After that verification statistics calculated for 14 day MM5 and WRF forecasts will be shown.

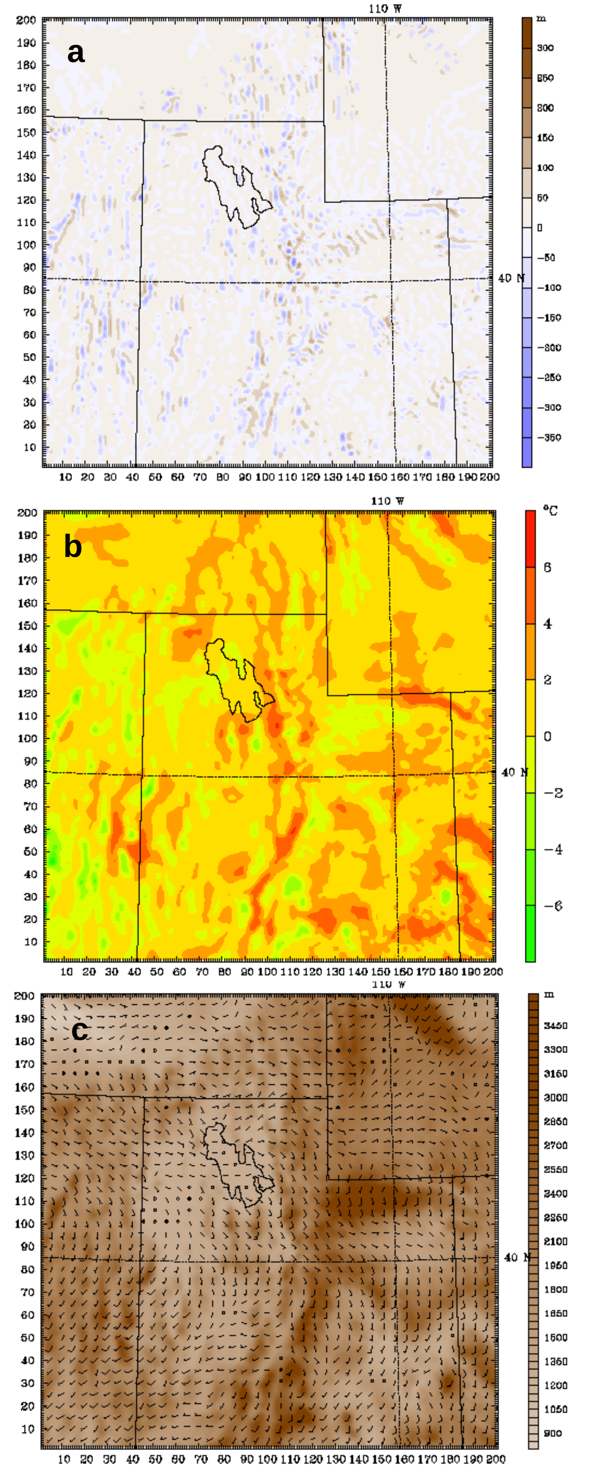

Fig.5 Differences of terrain height (a) and initial surface temperature (b), and wind vectors (c) between WRF and MM5 for the April 19,2004, 12Z. The WRF terrain heights are shown in the bottom panel also. A full bar is 5 m/s

First, model terrain fields and initial conditions are compared, and some interesting and surprising differences are observed between WRF and MM5. For example, Fig. 5a shows that the terrain heights in the WRF model differ from those of MM5 by more that 100 meters over all mountain ranges. A maximum of 200+ meters terrain differences can be seen in the mountain regions to the southeast of the Great Salt Lake. The two models make use of the same USGS 30” terrain datasets, but WRF SI (static initialization) tends to more strongly smooth the terrain field, which cut down the mountain peaks and fill up the valleys.

Besides the differences in the static terrain field, the initial conditions generated by the WRF SI are also quite different from those of MM5. It can be seen in Fig. 5 (b) that the surface temperature from the initial conditions of the two models differs by 4 – 6 C in many areas. Large differences are preferably located in the complex terrain areas. However, it is obvious that they are not simply caused by the artifact of terrain differences in the two models. The size of the patches varies from a few to tens of kilometers. Like the temperature, the initial surface winds of the two models are also different (Fig. 5c), with large difference vectors located over high terrain. The magnitude of the vector wind difference can reach 5 – 7 m/s over a few mountain ranges in the domain. Vertical cross-sections (not show) of the difference fields indicated that the differences between the two models decrease with height and become negligible above 600 hPa.

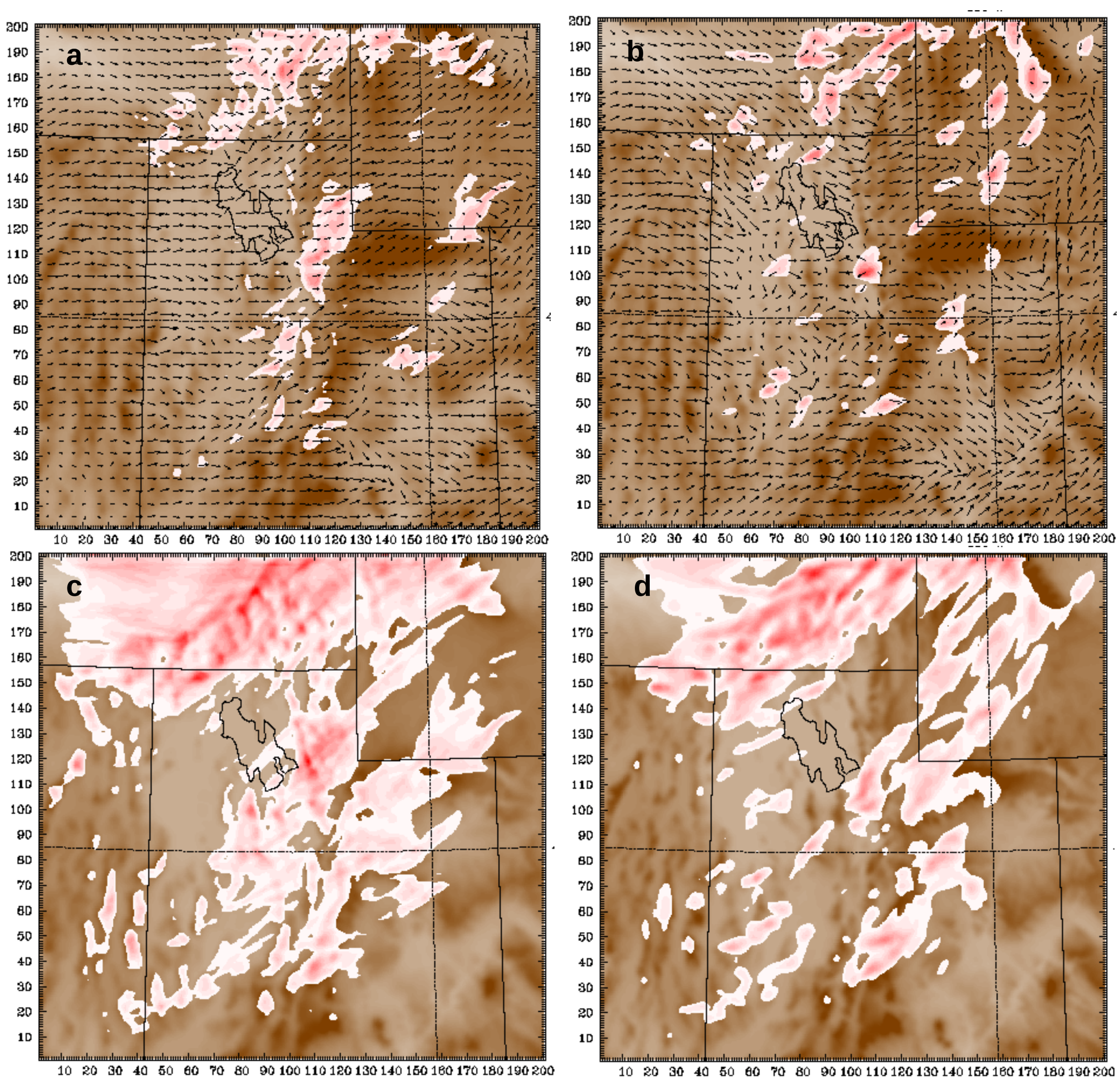

Fig. 6 1-h (a,b) and 12-h (c,d) accumulated precipitation of WRF (a,c) and MM5 (b,d) forecast, ending at 00 UTC, April 20, 2004. The precipitation amount is shaded from 1 – 14 mm with color changing from white to red. The maximum wind vector is 10 m/s. The terrain height is plotted the same as Fig.5c.

Because of the existence of the discrepancies in both terrain and initial fields between the two models, the forecasts of the models should differ from each other even if the two models were the same. The more accurate model dynamic and physical processes, the high-order advection computation and small numerical diffusion in the WRF model are known to better resolve the details of weather systems, compared with other mesoscale models with a similar model grid spacing. A complete analysis of how all these factors impact the model forecasts for various weather regimes will be an important and significant task in the WRF community. Here, an interesting orographic effect on forecasting local precipitation distribution with WRF and MM5 will be discussed.

The snow storm on April 19, 2004 produced large areas of snowfall over the northern and eastern part of Utah. Both WRF and MM5 forecast the event correctly. Fig, 6 compares the WRF and MM5 1-hour and 12-hour accumulated precipitation, ending at 00 UTC April 20. Although both models appear to forecast a similar overall distribution and amount of 12-hour precipitation accumulation, the detailed structures differs from each other greatly. Two significant precipitation features should be pointed out. Firstly, the snowbands forecasted by MM5 are more scattered, but those in WRF tend to be better organized and concentrated. Secondly, the precipitation bands and centers of the WRF forecasts are more clearly associated with orographic forcing than those in the MM5. Most rain centers in the WRF model, isolated or embedded in larger snowbands, are associated with a particular mountain range or peak (Fig.6a and c). In response to the synoptic scale forcing, both WRF and MM5 produced large scale precipitation bands oriented in a southwest to northeast direction. However, the intensive precipitation cores in the WRF model (Fig.6c) are preferably aligned along the orientation of the mountain ranges in the regions, mostly in the direction from south-southeast to north-northwest. In contrast, these precipitation cores in the MM5 forecast (Fig.6d) are aligned in the same direction as the larger scale snowbands. It should be pointed out that the WRF model precipitation forecast verifies better against the local radar observation (not shown).

Statistical verification of the model forecasts of the WRF and MM5 models over the 14-day period from May 9 to May 23, 2004 shows qualitatively similar relative skills to those associated with the coarser-resolution tests (Set 1), described in 3.1. Overall, WRF and MM5 show similar accuracy in terms of the RMSE errors of the model forecast fields at the surface, with MM5 again showing slightly better results (possibly for the same reason aforementioned). However, as shown in the above case studies, many fine-scale features of the two models are different. The dominance of the complex mountain in this high-resolution domain tends to generate larger error for surface wind forecasts, in both models, than those in Set I. In addition, the surface temperature analyses in the initial conditions of WRF and MM5 are also much larger, which can be caused by the inconsistency between the coarse ETA data and the fine-scale steep terrain. This error tends to spin-down in the first 3 – 5 hour forecasts in both models.

Mesoscale weather processes over complex terrain can be very complicated and very difficult to simulate and verify, even with high-resolution models. This makes the inter-model comparison challenging. To bring about some basic ideas about small-scale forecast error properties, verifications at individual stations at selected times are computed and averaged over the two-week forecast period. Fig.7 shows an example of the results. For clarity, only a sub-domain over the Salt Late valley and nearby regions is plotted.

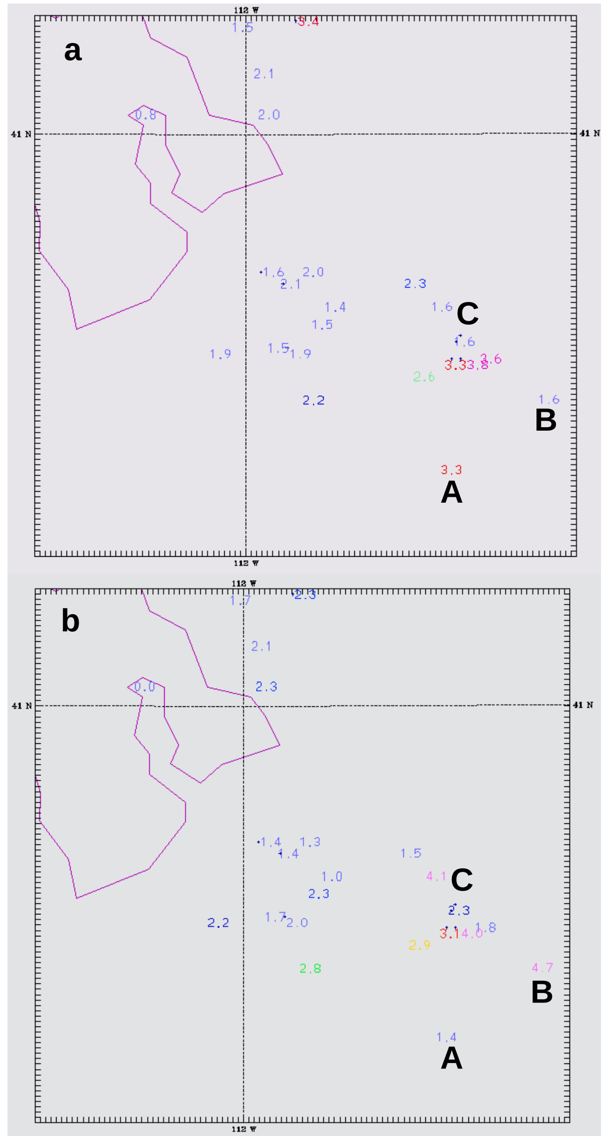

Fig. 7 Surface temperature RMSE errors (C) at available observations stations of the 18-h MM5 (a) and WRF (b) forecasts, valid at 18Z and averaged for a 14 forecast days between May 9 and 23, 2004.

It can be seen from the Fig.7 that, again, WRF and MM5 possess similar skills in most of areas. Both models forecast accurately the temperature in the Salt Lake Valley but are less skillful over the Rockies in the eastern domain areas. In some regions, for example, at the station A, B and C labeled in the Fig.7, the WRF and MM5 forecast accuracy can differ from each other dramatically. Similar phenomena are seen the surface wind and moisture forecasts (not shown).

Unfortunately, the upper-air data is not sufficient to verify the overall properties of the model forecasts over the domain.

3.3 Inter-range comparison of WRF forecasts

In this section, verification of the WRF model forecasts with Set 3 during March, 2004 is presented to assess and compare the general performance of the WRF model forecasts in the regions with different climatological and geographic environments. The domains of these forecasts are set to be the same as the coarse meshes of the operation RTFDDA systems running at the Army test ranges, with a size of about the half CONUS. Therefore, the acronyms of the Army ranges used here are purely for identification purposes and the verification statistics shown in this section represent large-scale weather property over the full domains.

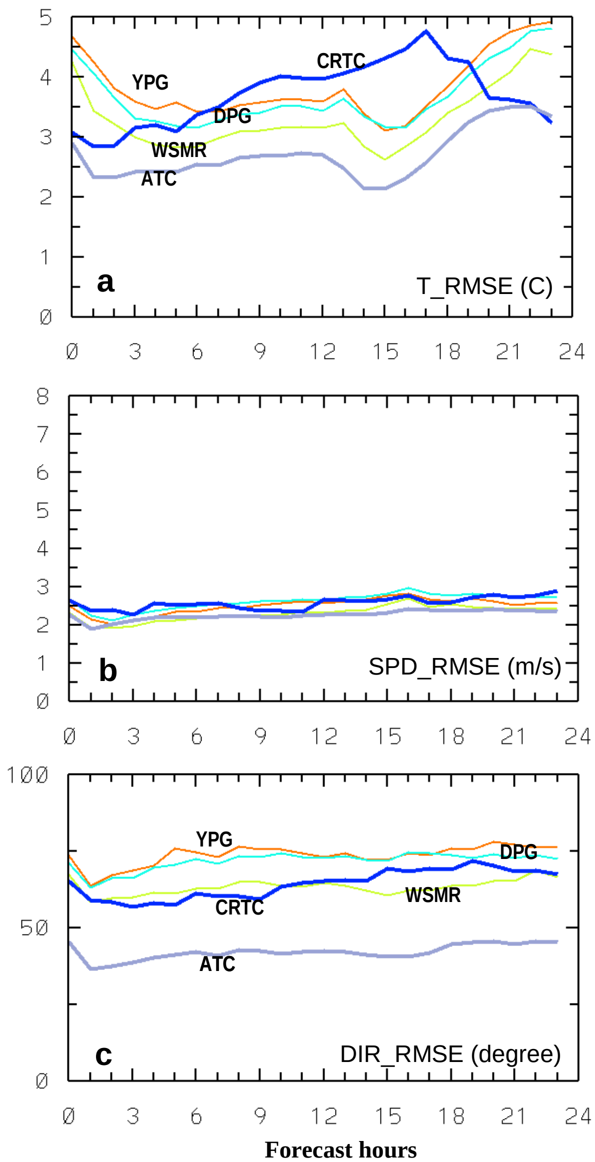

Fig. 8 Domain-average of RMSE errors of WRF surface temperature (a), wind speed (b) and wind direction (c) forecasts on the domains centered at 5 Army test ranges, between March 04 and 18, 2004.

Fig. 8 compares the domain-average the RMSE errors of surface temperature and wind forecasts in the five regions for the two-week period of March 4 - 18, 2004. It should be mentioned that at this time it is still winter over Alaska (CRTC), early spring for DPG and ATC, but fairly warm summer weather over WSMR and YPG. In spite of the huge differences of weather regimes, the WRF model, in general, performs reasonably well in all regions. The model forecasts over the flat eastern states (ATC) are most accurate for all surface variables, with an RMSE error 1 - 3 C lower than other ranges for the surface temperature and 40% less for the surface wind directions. The wind speed errors from the forecasts in the regions are closer to each other. The forecasts over CRTC display a quicker error growth with the forecast time than those at other ranges over lower latitudes. The DPG forecasts are very close to those of YPG, partially because their domains contain similar weather systems and geographic properties, even though YPG domain is shifted further south. Again, as shown in the Fig. 8a, initial conditions of the model over the complex terrain, including DPG, WSMR, and YPG are less accurate than those over the flatter terrain. It normally takes 3 -5 hour model integration to "spin-down" these errors. Finally, the conclusions drawn here are generally applicable to the upper-air situation, except that the ATC forecasts appear to have a larger wind error in upper-troposphere and the CRTC forecasts have a smaller error in most layers, especially for the temperature in the lower layer below 700 hPa.

4. SUMMARY

A number of semi-operational parallel experiments were conducted with the WRF and MM5 models for periods of two weeks to a month, with the exactly same domains and resolutions, and similar model physics schemes. The model forecasts are verified against observations to assess the performance of the models. It is found that the WRF model generally possesses similar skills to those of MM5 for forecasting different mesoscale weather processes and in different regions. WRF appears to better forecast the upper-troposphere circulations, whereas the MM5 produces more accurate forecasts at surface and in the lower troposphere, possibly due to the use of a Noah land surface model and a modified MRF scheme in MM5. The WRF model forecasts at a high-resolution could more properly simulate mountain-flow interaction and produce topographic precipitation, although the terrain used in the model is more heavily smoothed (by WRF SI) than that in MM5. The WRF model forecasts in the regions with dramatically different geographic and climatological environments exhibit reasonably-good performance in these regions. Finally, new physics schemes introduced in the recently-released WRF2.0, in particular, the surface layer and PBL schemes, will be tested, and adjusted for surface flux calculation if needed in order to improve the WRF surface forecasts.

5. REFERENCES

Liu, Y., F. Chen, T. Warner and S. Swerdlin, 2004: Improvements to surface flux computations in the MRF PBL scheme and refinements on Noah LSM urban processes with the NCAR/ATEC real-time FDDA and forecast system (in this workshop).