WSI's Operational Implementation of the WRF model

Todd A. Hutchinson, Stephen Marshall, Peter Sousounis
Weather Services International, Andover, Massachusetts

1. Introduction

WSI Corporation has developed a real-time, national-scale modeling system that uses the Weather Research and Forecast modeling system (WRF) as the forecast engine. WSI began running WRF operationally in April 2004. The model runs every 3 hours out to 18 hours. WSI forecasters began using WRF output in their forecasts for clients shortly thereafter.

WRF was chosen as the forecast model for the operational system due to its superior performance in predicting convectively-driven precipitation, as compared to other available mesoscale models. A comprehensive study was undertaken to evaluate several models (WRF, MM5, ARPS and workstation Eta) and various configurations of each of those models. The configurations consisted of different combinations of parameterization options for convection, cloud microphysics, etc. Results of all available models and configurations were evaluated in terms of the accuracy with which heavy convective precipitation was simulated. Results from that study will be presented here.

Based on the results from the evaluation, a WRF (version 1.3.2) configuration was chosen for the real-time modeling system WRF is initialized with RUC analyses and run once every 3 hours. Selection of initialization and boundary condition data sources was driven primarily by frequency and timeliness. Because NCEP's rapid update cycle (RUC) analyses are available every hour approximately 40 minutes after the hour (except for 00 and 12 UTC analyses that are available approximately 1 hr 5 min after analysis time), the RUC was chosen as the initialization data source. In order to provide boundary condition data for model runs that will begin at time of receipt of the RUC analyses, forecast data from the most recently available NCEP Eta run are used for boundary condition data. Because the LSM requires soil data from specific levels not provided by the RUC model, but provided by the Eta model, soil moisture and temperature are also initialized using Eta forecast data. Further, Chen and Dudhia (2001) suggested that the LSM should be initialized with soil analyses produced by a similar LSM such as that used in the Eta model.

The operational domain covers the continental United States with a grid-spacing of 12 km. Forecasts extend out to 18 hours with output every 10 minutes. Forecasts are complete approximately 1 hour 40 minutes after the initialization time, except for the 00 and 12 UTC initialized forecasts that complete about 2 hours and 5 minutes after initialization time.

We have encountered a number of problems while running the WRF operationally. When using the NOAH LSM, surface temperatures can have significant errors. Further, improper initialization of low-level moisture can lead to unobserved convection and sometimes model crashes. These issues will be discussed further in section 5.

2. Model Evaluation

The active 2003 Spring severe weather season in the Great Plains of the United States provided a good opportunity to evaluate a series of nearly consecutive periods of very active convective weather. More than one hundred different model configurations (spanning four different models: ARPS, MM5, WRF and workstation Eta) combined with 39 different initialization times resulted in almost 4000 simulations. Model output was verified against NCEP stage 4 hourly precipitation amounts (Baldwin and Mitchell, 1997) using traditional techniques (e.g., threat scores) as well as a new technique called acuity-fidelity (Marshall et al. 2004). Because of added spatial and temporal fidelity when mesoscale models are run at high-resolution, traditional verification (i.e., equitable threat score) of precipitation can appear worse than verification of precipitation from lower fidelity forecasts (Mass et al. 2002). The acuity-fidelity method was designed to evaluate high-fidelity mesoscale model forecasts more accurately and fairly compared to traditional methods. The objective of this acuity-fidelity technique is to account for temporal and intensity errors as well as spatial errors and then to cast the result in terms of a unidimensional score.

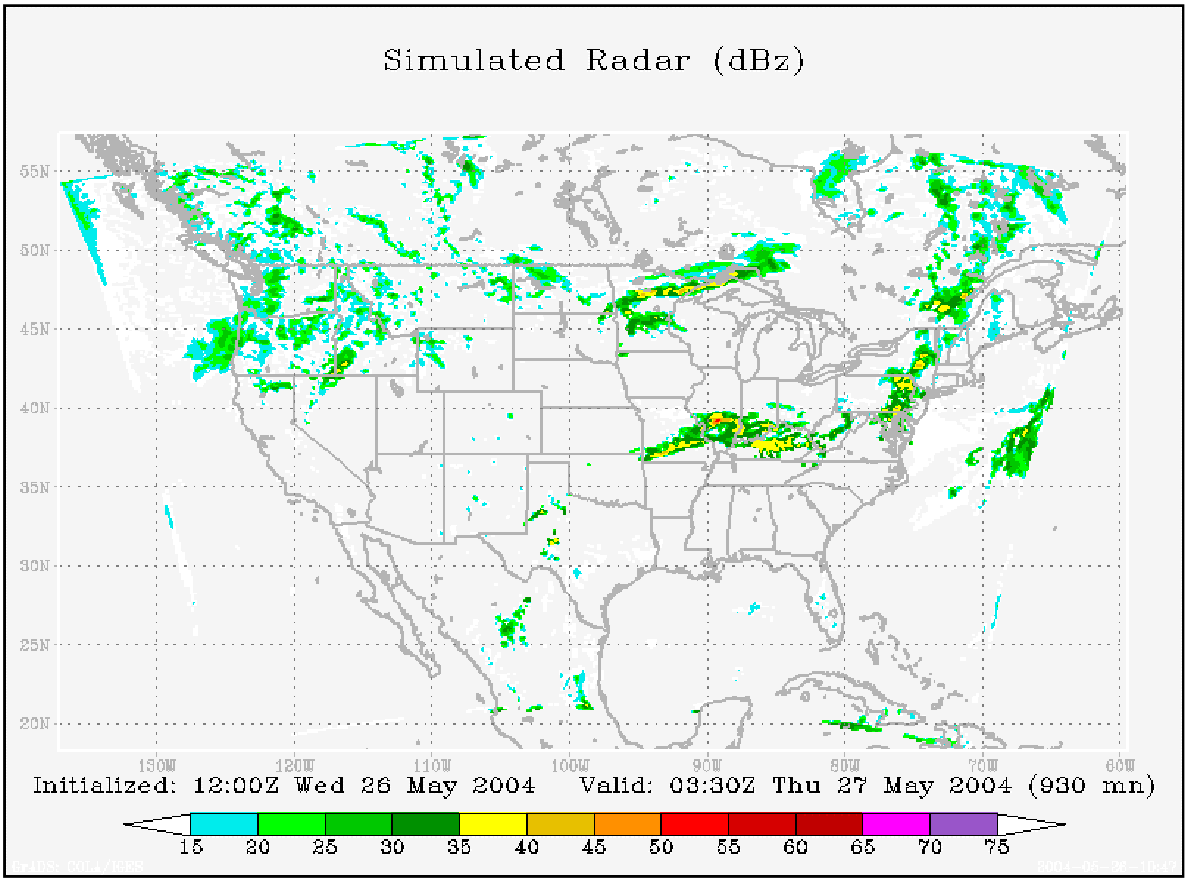

FIG. 1 Simulated radar forecast valid at 03:30 UTC 27 May 2004, from WSI's operational WRF simulation initialized 12 UTC 26 May 2004.

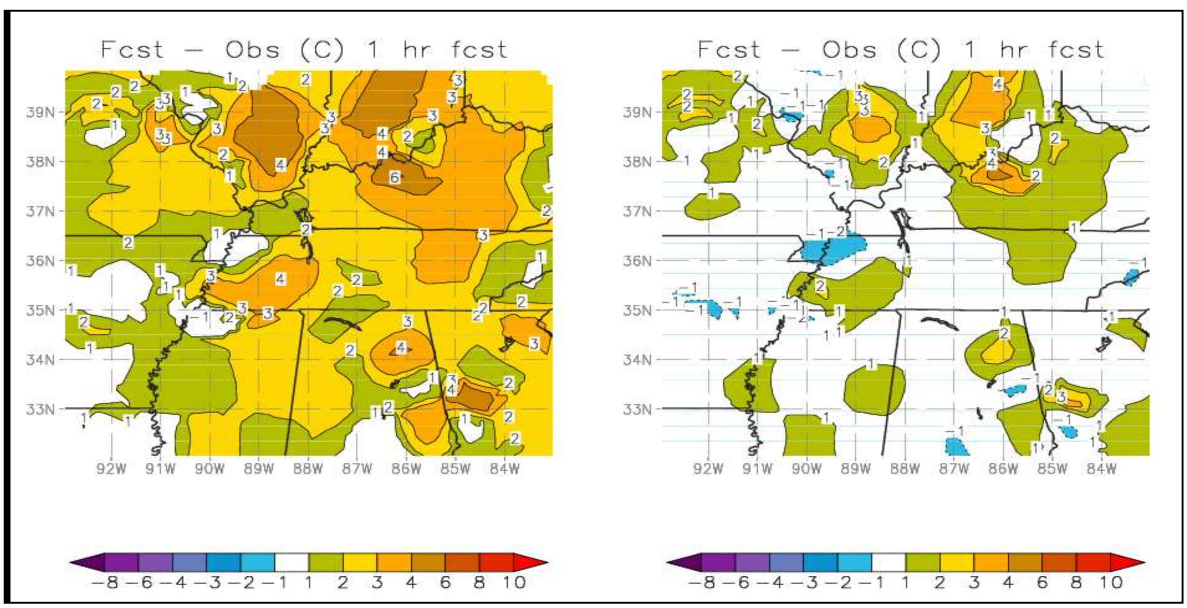

FIG. 2 Two-meter temperature difference between model forecast and observations (^{o}C) for a 1 hour forecast initialized 06 UTC 04 May 2004 for a) a WRF simulation using the NOAH LSM and b) a WRF simulation using the thermal diffusion scheme.

model	2-3 hr	5-6 hr	8-9 hr	11-12 hr	avg
MM5-36	163	176	166	175	170
WRF-36	135	146	148	160	147
WRF-12	101	99	114	136	113
WSEta-36	172	196	188	169	181
NCEPEta	185	196	193	177	188
NCEPRUC	202	213	206	205	206

Table 1. Values of acuity+fidelity for a precipitation intensity threshold of 0.25 in/hr at four different forecast times for the models evaluated as well as for two NCEP models. All values are averaged over the 39 cases.

Acuity-fidelity verification was developed and implemented to quantify the skill of a forecast using the three metrics of space, time and intensity. Acuity represents the model's skill at detecting the features of the observed data. The acuity of a forecast is calculated for each observed data point by finding the best matching forecast for that observation. Instead of automatically associating an observation with the forecast that shares its location and time, the best match is obtained by minimizing a cost function calculated between the target observation and many candidate forecast data. The candidate forecast datum that produces the smallest penalty is deemed the best match, and is therefore associated with the observation. Fidelity represents the faithfulness of the model's predictions to the observed data. The fidelity of a forecast is calculated much like the acuity, except roles of the observations and forecasts are reversed. Thus for each target forecast datum, the best matching observation is found within a multidimensional field of candidate observations.

The combined acuity-fidelity scores for several models are shown in Table 1 for a threshold of 0.25 in/hr (6.35 mm/hr). For MM5, WRF and workstation Eta, the results from the configuration with the best scores are shown. A comparison of MM5 36 km and WRF 36 km results demonstrates the slightly better (i.e., lower scores) performance of WRF over MM5. Further, results from the WRF 12 km indicate that the WRF 12 km output has better acuity-fidelity scores than the 36 km output. Based on these results, WRF was chosen as the operational model. More detailed results from this study can be found in Sousounis et al. (2004).

3. Results

Verification from operations during Spring 2004 (not shown) suggest that WRF is overpredicting areal coverage of precipitation. Verification results are preliminary at the time of this writing, but, we will present additional results at the workshop.

The high-fidelity of WRF is proving to be extremely useful. WRF has been shown to resolve smaller features than other models that are run at similar resolutions (i.e., the Eta model; Baldwin and Wandishin, 2002). With frequent output, WRF portrays physically realistic features, such as those seen in radar animations. Forecasters can often infer convective storm type by analyzing simulated radar loops. Fig. 1 shows WSI's 15 h 30 min forecast initialized at 12 UTC 26 May 2004.

4. Challenges

In developing this system, several problems were discovered. Initially, the NOAH land-surface model (initialized with Eta forecast soil data) was used. However, surface (2m) temperature errors were often large even early in the model run (as early as 10 minutes). To determine the cause of the errors, we ran test simulations over the Ohio River Valley. Configurations for these simulations were identical to the operational configuration, except that the domain was smaller and soil data was initialized with Eta analysis data rather than Eta forecast data. Fig. 2 shows 2 m temperature error (computed by interpolating forecast data to station locations and subtracting observed values from the interpolated forecast values, then objectively analyzing the differences) for a 1 hour forecast from simulations using (a) the NOAH LSM and (b) the thermal diffusion scheme. It is evident that at this early time in the model run, the simulation using the NOAH LSM suffers from a warm bias throughout most of the domain. The thermal diffusion scheme (b) did not suffer nearly as large a bias. To verify (and double check) that the soil was initialized correctly in the simulation using the NOAH LSM, the WRF 0 hr forecasts was examined to assure that the soil temperature and moisture were identical to that in the Eta model analysis.

In order to determine the diurnal variability of this bias, a series of one-hour forecasts of 2 m temperature at the grid point nearest to Paducah, KY (KPAH), both for simulations using the NOAH LSM and simulations using the thermal diffusion scheme, are compared to observations at KPAH in Fig. 3. The initialization time for the forecast is indicated along the abscissa. For WRF forecasts initialized at 18, 00, 06, and 12 UTC soil data from Eta analyses were used to initialize the LSM. For all other WRF forecasts, Eta 3 hr forecasts initialized 3 hours earlier were used to initialize soil moisture and temperature. During this period, clear skies were observed and all simulations forecast clear skies (not shown). For simulations initialized between 00 and 12 UTC, one hour forecasts are too warm. Similar biases seem to be prevalent over land east of the Rocky Mountains

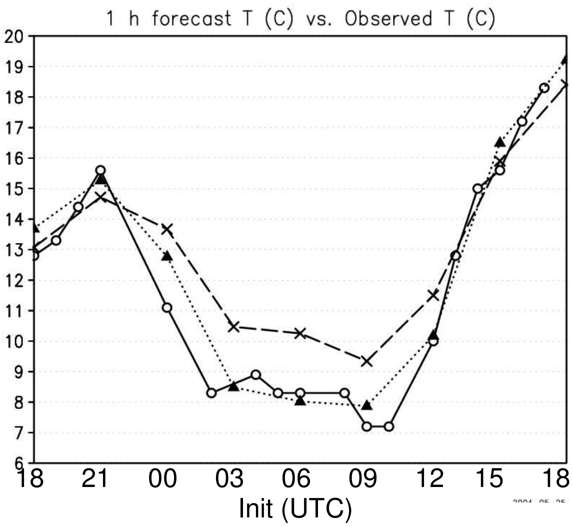

Fig. 3. One hour forecasts of 2 m temperature (°C) at the grid point nearest to Paducah, KY (36.98 N, 88.82 W) are compared to observations (solid line). Two simulations are shown, one using the NOAH LSM (long-dashed line) and one using the thermal diffusion scheme (dotted line). Simulations were initialized between 18 UTC 3 May 2004 and 18 UTC 4 May 2004 as indicated along the x-axis.

in most of our simulations that use the NOAH LSM (not shown). Analyses of Eta soil temperatures suggest that they are often too warm, as compared to RUC soil analyses. Thus, we hypothesize that warm Eta soil temperatures are resulting in the 2 m temperature bias early in the forecast.

initialized between 00 and 12 UTC, one too warm.

The 20-km RUC analysis on isentropic coordinates is the initialization data source for our real-time simulations. We have occasionally found problems with this data source. Specifically, excessive moisture over northwest Mexico and the western mountains of Mexico can lead to intense convection that is not observed, and occasionally, this convection causes the simulation to abort within the first 2 hours of the forecast. Our tests show that a time step reduction reduces the number of simulations that abort, however we are reluctant to reduce the time-step much, since this results in longer simulation run-times. The Forecast Systems Laboratory (FSL) develops and maintains the RUC model. They are aware of problems with analysis of low-level moisture and are currently working to improve the low-level moisture analyses (Stan Benjamin, personal communication). However, we are continuing to look into other options for reducing the frequency of aborted simulations.

5. Summary

WSI Corporation has developed a real-time, national modeling system that uses the Weather Research and Forecast modeling system as the forecast engine. Forecasters at WSI are using output from WRF as guidance in producing forecasts for clients. WRF is proving that it can generate physically realistic fine-scale structure not seen in the standard output resolutions of other operational forecast models. Challenges have been encountered while implementing this system. We have found that when the NOAH LSM is used, WRF often exhibits a warm surface temperature bias in the first few hours of the simulation. Further, while WRF has been very robust, we have occasionally observed model crashes due to improperly initialized moisture fields that lead to unobserved convection.

REFERENCES

Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. On Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54-55.

Baldwin, M. E., and M. S. Wandishin, 2002: Determining the resolved spatial scales of Eta model precipitation forecasts. Preprints *15th Conf. On Numerical Weather Prediction*, Fort Lauderdale, FL, Amer. Meteor. Soc. 3.2.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: model implementation and sensitivity. *Mon. Wea. Rev.* **129,** 569-585.

Marshall, S. F., P. J. Sousounis, and T. A. Hutchinson, 2004: Verifying mesoscale model precipitation forecasts using an acuity-fidelity approach. *Preprints 20th Conference on Weather Analysis and Forecasting*, Seattle, Amer. Meteor. Soc., J13.3.

Mass, C. F., D. Ovens, K. J. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce better forecasts? The results of two years of real-time numerical weather prediction in the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **83**, 407-430.

Sousounis, P. J., T. A. Hutchinson, Stephen F. Marshall, 2004: A comparison of MM5, WRF, RUC, and ETA performance for great plains heavy precipitation events during the spring of 2003. *Preprints 20th Conference on Weather Analysis and Forecasting,* Seattle, Amer. Meteor. Soc., J24.6.