Summer Season Verification of the First NWS Operational WRF Model Forecasts from the NOAA Coastal Storms Initiative Project in Northeast Florida

P. Bogenschutz and P. Ruscher, Florida State University, Tallahassee, FL
P. Welsh, J. Mahoney, J. A. McGinley, M. Kay, B. Shaw, J. Smart, J. Savadel, and J. McQueen, NOAA

I) Model

The WRF is a regional atmospheric model for operation numerical weather prediction and atmospheric research focusing on scales of 1-10 km (UCAR 2002). For this study the WRF model has a scale of 5km. The WRF's performance is verified using two different initial conditions. The first in which the locally run Jacksonville WRF uses only the Eta model for initial conditions (WRF-Eta or WRF "cold-start") and the other where the locally run WRF is initialized by NOAA FSL's Local Analysis and Prediction System (NOAA FSL) (WRF-LAPS or WRF "hot-start"). In the WRF-LAPS initialization LAPS is added to assimilate Doppler radar, satellite clouds, and local surface data. More elaborately, it is using several channels of GOES satellite imagery, mesonet data, radar data, ship and buoy observations, standard METAR reports, and ACARS, in addition to several other features.

II) Methodology

The statistical evaluation of the WRF versus the Eta model (12 km domain) performance focuses on the bias, mean absolute error, mean squared error, root mean squared error, and equitable skill score, using FSL's Real Time Verification System (NOAA 2003). The bias is simply the comparison of the average forecast with the average observation. The result tells us whether the forecast was over forecasted, under forecasted, or unbiased. An unbiased forecast will have a value of one. Over-forecasting, in which the event was forecasted more often than observed, holds a bias value greater than one. Under-forecasting has a bias value less than one and says that the event was forecast less often than observed.

The mean absolute error (MAE) is given by:

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|y_k - o_k\right|$$

and is simply the average of the absolute errors between the forecast and observation. Hence, if the MAE is zero that says the forecast is perfect and the value increases proportionally with the discrepancies in the forecast. The MAE describes the typical magnitude for the forecast error.

The mean squared error (MSE) is given by:

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2$$

and is the average squared difference between the forecast and observation. Unlike the MAE the MSE uses the squaring function instead of the absolute value function. In addition, the MSE will be more sensitive to larger errors than the MAE. The MSE indicates a perfect forecast when it exhibits a zero, with errors increasing with larger MSE values.

The root mean squared error (RMSE) is given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2}$$

and is simply the square root of the MSE. The RMSE has the same physical dimensions as the forecasts and observations and is useful since it is thought of as a typical magnitude for forecast errors.

The equitable skill score (ESS) is given by:

$$ESS = \frac{(H - CH)}{(F + O - H - CH)}$$

where F is the number of grid boxes in the model that forecast more than the threshold of precipitation, O is the number of grid boxes that observe more than the threshold, H is the number of grid boxes that correctly forecast more than the threshold, and CH is the expected number of forecasts due to chance. The ESS is a good estimate for the overall forecast skill for precipitation. Values typically range from small negative values to a perfect score of 1.0.

III) Statistical Evaluation

The statistical evaluation is performed using data from the RTVS. The RTVS is capable of producing statistics for the Operational Eta, WRF-Eta, and WRF-LAPS. All models produce results from the 06 UTC runs while the WRF-LAPS is also run at 15UTC & 21UTC. The three models run at 06 UTC; the Eta model, WRF-Eta, and WRF-LAPS 06 UTC are evaluated against each other while the WRF-LAPS 15 UTC and WRF-LAPS 21 UTC statistics are compared to the 06 UTC models to get a measure of their performance. The WRF-Eta and WRF-LAPS are capable of producing statistics for each forecast hour while the Eta produces output for each hour divisible by three up to twenty-four hours. Each model produces results for the bias, MAE, MSE, and RMSE for temperature, relative humidity, wind speed, and the u & v wind components. The RTVS for precipitation only calculates statistics for the ESS and bias. Each variable is analyzed for the average of all forecast hours in a 24 hour period as well as individually for the 6, 12, 18, and 24^{th} forecast hours. All statistics are for surface parameters only (2 m temperature and humidity, 10 m winds, and surface precipitation).

Two periods are used in the analysis of the statistics; the periods which will be referred to as the June period (June 1^{st}- July 8^{th}) and the July period (July 9^{th} – August 13^{th}). Each period contains roughly twenty-one days for which data are available for each statistic. The Eta model underwent a dramatic change to the initialization procedures and the surface physics packages on July 9^{th} 00 UTC (NCEP 2003). The two periods are deliberately defined to study the pre-Eta and post-Eta change, in case there are important differences in the derived statistics. The July period contains four days in which no radar data was fed into the WRF model (August 1^{st}-4^{th}), thus those days are omitted from the study.

a) Temperature

For both the June and July periods the Eta model outperforms both the WRF-LAPS 06 UTC and the WRF-Eta in every statistic for the average of all the forecast hours. The WRF-LAPS 06 UTC and WRF-Eta 06 UTC consistently under forecasts temperature with large errors. While the Eta model also has large errors they are not as large as those created by the WRF 06 UTC models. The typical magnitude of error for the models ranges from 2-2.8 degrees Celsius (Fig. 1 & 2).

The Eta model performs strongest at the 12 & 18^{th} forecast hours for both the June and July periods where the 18^{th} hour forecast is nearly unbiased. However, at the 24^{th} forecast hour the WRF-LAPS 06 UTC and WRF-Eta outperforms the Eta model for both periods. The Eta model starts off weak at the 6^{th} forecast hour, strengthens for the 12-18 hour forecast and begins to weaken once again at 24 hour forecast. The performance for the WRF-LAPS 06 UTC is inversely related to the Eta model pattern as it starts out strong, becomes weak at the 12-18 hour forecasts but strengthens dramatically for the 24 hour forecast. The WRF-LAPS 21 UTC starts out very strong for each period yet the forecast quality quickly deteriorates, especially at the 18^{th} forecast hour. The WRF-LAPS 15 UTC start has less than desirable statistics at the 6^{th} forecast hour, gets slightly better during the 12 & 18 hour forecasts, and becomes inferior again at the 24^{th} forecast hour. These patterns hold true for both periods.

b) Relative Humidity

Examining the output for the average of all forecast hours for relative humidity it can easily be seen that the Eta model exhibits smaller errors compared to the WRF-Eta and WRF-LAPS 06 UTC once again. Each model exhibits substantially large errors however the WRF-LAPS 06 UTC shows a bias that severely over-forecasts the relative humidity and exhibits very high MSE's. For example the WRF-LAPS 21Z has biases as high as 12 and MSE scores commonly reaching 368. These significant errors parallel the results for the WRF-LAPS 15 UTC and WRF-LAPS 21 UTC runs (Fig. 3& 4).

The examination of the individual forecast hours for each period shows the Eta model clearly outperforming the WRF-Eta and WRF-LAPS 06 UTC significantly for the 6 and 12 hour forecasts. At the 18^{th} forecast hour the WRF-Eta begins to show more favorable statistics for both periods. The 24 hour forecast has the Eta model outperforming both of the WRF 06 UTC models once again. However, the magnitude of errors is relatively small at the 24 hour forecast compared to the other forecast hours. The largest errors associated between all three models occurred during the daylight hours and the errors become less dramatic in the evening and night forecast hours.

c) Precipitation

The RTVS produces results for several different precipitation thresholds, ranging from 0.01 to 2.0 inches. For this variable the only statistics to test are the ESS and bias. The ESS score is a ratio of the correct forecast area to the total area of the forecast and observed precipitation. Thus the model will get penalized for forecasting rain in the wrong place as well as not forecasting rain in the right place. So while the ESS is a good measure of the model with the best forecast skill in terms of placement of precipitation the bias gives an indication if a model is consistently over-or under-forecasting areas of precipitation.

The bias for the average of all forecast hours shows it is evident that the WRF-LAPS 06 UTC and WRF-Eta outperforms the Eta model at all periods. For June the WRF-Eta 06 UTC is the dominant model for the lower thresholds while the WRF-LAPS 06 UTC clearly did better at the higher thresholds of precipitation. When looking at July it is interesting to note the prolific deterioration of the WRF-LAPS 06 UTC as well as the WRF-LAPS 15 UTC and 21 UTC runs. All three models exhibit biases as high as 4.5. While the Eta model has biases as high 5.0 it does improve dramatically as the threshold gets larger. The most consistent model during the July period was the WRF-Eta. It is well known that models tend to over-forecast precipitation at small thresholds and under-forecast precipitation at larger thresholds. The occurs here, as well, although the tendency is less with the WRF-LAPS 06 UTC model (Fig. 5).

The ESS scores for the average of all forecast hours shows the WRF-LAPS 06 UTC having the better forecast for the lower thresholds of precipitation while the Eta model performs the best at the higher thresholds (Fig. 6). This pattern is relevant for both periods. Examining the 6th forecast hour the ETA-LAPS 06 UTC holds the advantage for nearly every threshold for both periods. For the 12 hour forecast the Eta model is the best scoring forecast for June and the WRF-Eta for July. The 18 & 24th forecast hours have the Eta model exhibiting the best ESS scores. This suggests that the Eta model is most accurate when it comes to forecasting precipitation location, although it does suffer from very poor bias scores. It is important to note that none of the ESS scores are particularly high, the highest score being found is only 0.159 from the Eta during the June period for the 18th hour threshold of 0.25.

The analysis of the hourly statistics for the bias shows the WRF-LAPS 06 UTC as the consistent performing model at the 6 hour forecast during the June period. During the forecast hours of 12, 18, and 24 it is clear that both of the WRF 06 UTC (as well as the WRF 21 UTC & 15 UTC) models outperformed the Eta, although the WRF-LAPS 06 UTC was the strongest at the 6th forecast hour. While the Eta's performance is acceptable during the later forecast hours, at the 6th forecast hour for both June and July periods it suffers from very high bias scores. This could be attributed to the Eta model's characteristic of having very slow spin up of the moist processes which makes it likely to under perform at the 6th forecast hour. The July hourly forecast shows the WRF-Eta 06 UTC overall performing much stronger than the WRF-LAPS 06 UTC and Eta. However, the strongest hour for the WRF-LAPS 06 UTC once again was the 6th hour forecast. While the quality of the WRF-LAPS 06 UTC model deteriorates with increasing forecast hours the Eta progressively gets better with time.

d) Wind Speed

The WRF-LAPS 06 UTC and WRF-Eta are substantially more accurate than the Eta model in the evaluation of all forecast hours. While the WRF-LAPS 06 UTC is the best performer during the June period (and WRF-Eta during July) the two models are very close in their performance and exhibit exceptional jobs. The WRF-LAPS 15 UTC also does a superior job during both months while the WRF-LAPS 21 UTC lags behind slightly, although with better statistics than the Eta model which consistently over forecasts the wind speed (Fig. 7 & 8). This substantial change in performance between the Eta and the WRF models is most likely in part of the WRF's finer grid spacing than the Eta model which allows it to capture more local effects.

The hourly forecast shows no unexpected results during the June period with the WRF-LAPS 06 UTC performing the best at each forecast hour and statistic, with the exception of the 12th forecast hour where the WRF-Eta shows an advantage. At this hour the WRF-LAPS 06 UTC statistics seem to be uncharacteristically unreliable, especially the low bias score. The July period is where the Eta shows a surprisingly strong performance during the 6th forecast hour and a near perfect bias at the 12th forecast hour, although the model quickly deteriorates in later forecast hours with very high MSE and MAE scores, as well as severely over forecasting during the 18th forecast hour. During the later forecast hours the WRF-Eta and WRF-LAPS 06 UTC are the two dominate models with fairly equal statistics. While the WRF-LAPS 15 UTC model has statistics comparable to the two WRF 06 UTC runs the WRF-LAPS 21 UTC shows unreliable statistics at the 6th forecast hour for both periods then becomes more reliable than the WRF-LAPS 15 UTC until the 24th forecast hour.

e) Summary

The results from the Eta, WRF-Eta, and WRF-LAPS clearly states that each model has its strength and areas of weakness. While the Eta model appears to be the most accurate for temperature and precipitation, the statistical results suffers in comparison to the WFR-Eta and WRF-LAPS models for precipitation and wind speed. Both the June and July periods hold very similar statistics, with the exception of the WRF-LAPS 06 UTC performance in precipitation. A third period was statistically evaluated from August 14th-Septempter 21st and the results match those from the June and July periods, suggesting very little fluctuation in model performance over the summer months. In addition, the statistics from the post-Eta change did not show any significant differences from the pre-Eta statistics.

Future work plans consists of an evaluation of a land falling tropical storm, several sea breeze cases, and other interesting events. A statistical evaluation for the winter season may also be conducted.

REFERENCES:

UCAR, 2002: Weather Research and Forecast (WRF) Model, http://www.esmf.ucar.edu

NOAA, 2003: Real-Time Verification System, http://www-ad.fsl.noaa.gov/fvb/rtvs/index.html.

NOAA, 1999: Local Analysis and Prediction System, http://laps.fsl.noaa.gov.

NCEP, 2003: Changes to the NCEP Meso Eta Analysis and Forecast System, http://wwwt.emc.ncep.noaa.gov.

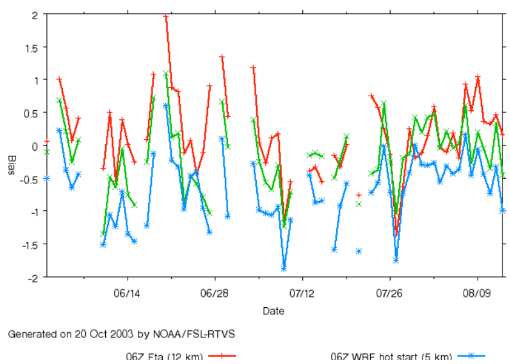

Fig 1. Bias for the June & July period Temperature 06 UTC model runs.

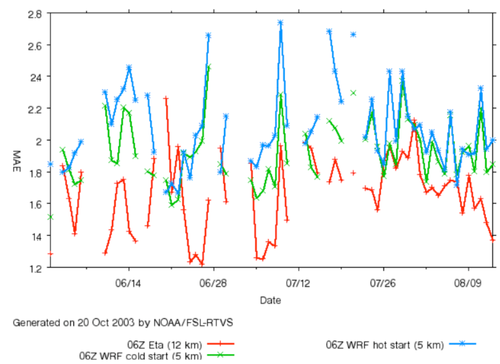

Fig 2. MAE for the June & July period Temperature 06 UTC model runs.