NOAA Technical Memorandum OAR FSL-31

COASTAL STORMS INITIATIVE: SUMMARY OF 1 JUNE-31 AUGUST 2003 EVALUATION

J.L. Mahoney
M. Kay
B. Shaw
J. McGinley
J. Smart
J. Savadel

Forecast Systems Laboratory
Boulder, Colorado
March 2004

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR FSL-31

COASTAL STORMS INITIATIVE: SUMMARY OF 1 JUNE-31 AUGUST 2003 EVALUATION

Jennifer L. Mahoney[1, 2]
Mike Kay[1, 3]
Brent Shaw[1, 4]
John McGinley[1]
John Smart [1]
Jeff Savadel[5]

[1]NOAA Research - Forecast Systems Laboratory
[2]Corresponding Author: Jennifer Mahoney, 325 Broadway FS5, Boulder, CO
[3]Joint collaboration with the Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, CO
[4]Joint collaboration with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO
[5]NOAA-National Weather Service

Forecast Systems Laboratory
Boulder, Colorado
March 2004

UNITED STATES DEPARTMENT OF COMMERCE

Donald L. Evans
Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

VADM Conrad C. Lautenbacher, Jr.
Under Secretary for Oceans and Atmosphere/Administrator

Office of Oceanic and Atmospheric Research

Richard D. Rosen
Assistant Administrator

NOTICE

Mention of a commercial company or product does not constitute an endorsement by the NOAA Oceanic and Atmospheric Research Laboratories. Use of information from this publication concerning proprietary products or the test of such products for publicity or advertising purposes is not authorized.

For sale by the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22061

CONTENTS

Coastal Storms Initiative: Summary of 1 June–31 August 2003 Evaluation

Jennifer L. Mahoney, Mike Kay, Brent Shaw, John McGinley, John Smart, and Jeff Savadel

ABSTRACT. As part of the Coastal Storms Initiative (CSI) project, an intensive statistical evaluation of the newly implemented Weather Research and Forecast (WRF) mesoscale numerical weather prediction model that was installed at the National Weather Service Weather Forecast Office in Jacksonville, Florida was conducted from 1 June–31 August 2003. The evaluation was subdivided into two parts: an objective real-time evaluation and a subjective meteorological evaluation. The results presented in this report summarize only the objective real-time evaluation. Forecasts were verified using the Forecast Systems Laboratory's Real-Time Verification System (RTVS; Mahoney et al. 2002). The variables chosen for evaluation were accumulated precipitation, surface winds, surface temperature, and surface relative humidity. Statistical results are available through the RTVS web site (*www-ad.fsl.noaa.gov/fvb/rtvs/*; link Coastal Storms Initiative). Only a selection of the results is summarized in this report. Two different configurations of the WRF (WRF-Hot and WRF-Cold) and the National Centers for Environmental Prediction (NCEP) Eta models were evaluated. The forecasts from the 0600 UTC run, 3-hourly forecasts from 3 to 24 hours, were compared.

Results of the evaluation indicated that:

• Improvement in local precipitation forecasts as a result of locally run numerical models was evident in the statistical results. The greatest improvement in the precipitation forecasts from the WRF-Hot was noted at short lead times and over lower thresholds. Improvement was also noted at larger thresholds where the WRF-Hot continued to produce precipitation while the Eta dramatically underpredicted the precipitation amounts. At 3 hours and at thresholds typically greater than 1.0 inch, little forecast skill was noted for all models.

• Improvement in local wind forecasts produced by the WRF-Hot as compared to the Eta was evident in the results. Overall, the WRF models had a statistically significant lower bias and statistical errors in the forecast wind speed than did the Eta at all forecast hours, although the greatest improvement occurred at 12 hours. The error in wind speed at 3 hours for the WRF-Hot was nearly 1 m s^{-1} less than for Eta. Errors were slightly smaller for the u-component of the wind for the WRF-Hot than for the other two models. Improved in the bias and MSE values for the WRF-Hot v-component of the wind. However, the bias-corrected RMSE errors for the v-component were slightly larger for the WRF-Hot than for the other two.

• No significant improvement was noted in temperature or relative humidity forecasts for the local WRF as compared to Eta. The positive and negative temperature bias exhibited by the WRF-Hot during the evaluation was the result of the WRF model generating too

much low-level cloudiness due to a bias in the LAPS-provided 3D temperature analysis. The bias in the LAPS temperature analysis led to rapid cloud development in the WRF model. This mechanism is only partially responsible for the surface temperature forecast bias in the WRF-Hot forecasts. This issue was corrected in the WRF, since the evaluation. An additional contributor to the temperature bias was that all of the models fail to accurately predict the full amplitude of the diurnal temperature curve. This was especially true in the WRF runs where the PBL schemes "flattened" the diurnal curve in the forecasts (Shaw et al. 2004).

1. Introduction

One of the goals of the NOAA Coastal Storms Initiative (CSI) Program is to advance short-term warning and forecast services by improving coastal wind, wave, and quantitative precipitation forecasts. To this end, the Weather Research and Forecasting (WRF) model was implemented at the Jacksonville, FL (JAX) National Weather Service Forecast Office (NWS WFO) to test the ability of a locally run, high-resolution mesoscale model to improve the local forecasts produced at the Jacksonville WFO.

A statistical evaluation of the JAX WRF was conducted for the period 1 June–31 August 2003 to determine whether forecast services provided by the WFO can be enhanced through the use of locally run mesoscale modeling systems. Objective statistics for precipitation, wind, and temperature forecasts are summarized in this report. Results for relative humidity forecasts were difficult to interpret because of the relationship to temperature so those results are excluded from this report. However, the interested reader can access the relative humidity results from the RTVS Website (*www-ad.fsl.noaa.gov/fvb/rtvs/csi/*). The statistics were generated through the Real-Time Verification System (RTVS; Mahoney et al 2002), which is a verification system that is being developed by NOAA's Forecast Systems Laboratory with funds provided by the Federal Aviation Administration's Aviation Weather Research Program. Ongoing statistics for the CSI project and statistics f are available from the RTVS Website. In addition, analyses of meteorological characteristics, such as the sea breeze and other forecast attributes, are summarized by Bogenschutz et al. (2004). Welsh et al. (2004) summarized the challenges and successes associated with the addition of the WRF forecasts for the CSI project to WFO operations at JAX and Shaw et al. (2004) summarized the model configurations used in the evaluation and successes and failures of running models at the local WFO.

This report is organized as follows. The evaluation overview is presented in Section 2. Section 3 briefly describes the numerical models that were included in evaluation, and the observations used to assess the quality of the forecasts are discussed in Section 4. The verification methods are described in Section 5. Results of the study are presented in Section 6. Section 7 includes a summary and future work, and additional statistics are presented in the Appendix.

2. Evaluation Overview

Forecasts produced by the high-resolution WRF numerical weather prediction model implemented at JAX (Shaw et al. 2004) and the National Centers for Environmental Prediction (NCEP) Eta (Black 1994) were evaluated using the RTVS (RTVS; Mahoney et al. 2002). The local runs of the WRF had forecast lengths of 24 hours, with data available hourly. The hourly precipitation forecasts were binned into 3-hour accumulations and were evaluated against accumulations of 3-hourly precipitation observations. Eta model forecasts were evaluated at 3-hourly increments to 24 hours to match the forecast lengths of the local WRF runs; hourly output data were not available. Owing to the constraints of the local forecast operations as well as the limitations of the computational resources at JAX, the 0600 UTC model initialization time was used as the basis for the intercomparison between the WRF and Eta forecasts. The statistical results were based on 69 events that were accumulated during the evaluation period. Two more WRF runs at 1500 and 2100 UTC were provided to meet the needs of the JAX WFO. The statistical results for these runs can be obtained from the RTVS web site. The model domain for the WRF runs was chosen to cover an area of the southeastern U.S. and was associated with the JAX forecast area and adjacent coastal waters (Fig. 1). Verification analyses were limited to this domain for the Eta model as well.

WRF forecasts produced at JAX were transferred to FSL for analysis while output from the Eta model arrived at FSL via NOAAPORT. During the evaluation period, several data outages occurred and are listed here for reference: June 7–9, 16, 19, 29; July 1–3, 11–12, 16, 20, 22; and August 17, 23–25, 29–31.

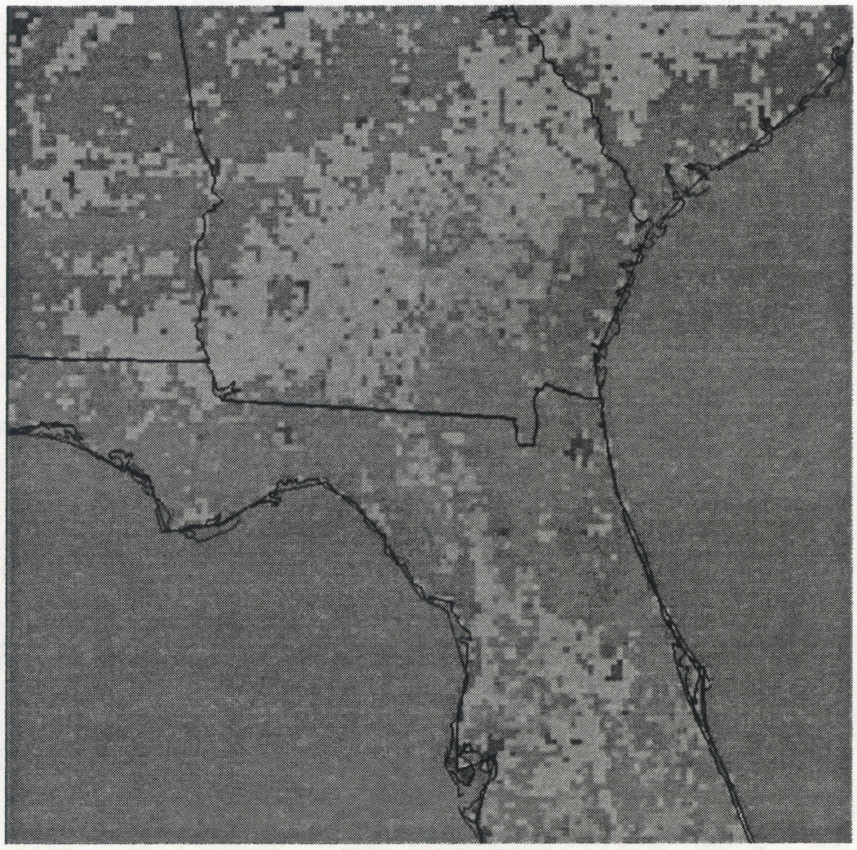


Figure 1. Model domain for the WRF runs. The map is a colorized version of the USGS vegetation category showing some small rivers and lakes at a 5-km resolution.

3. Numerical Model Descriptions

A summary of the numerical weather models used in the evaluation is described in this section.

Weather Research and Forecast (WRF) Mesoscale Model – The WRF model is the next-generation numerical model being developed by the atmospheric science community to be used for both research and operational forecasting purposes. WRF is available to the general community for download at *www.wrf-model.org*. Specific dynamic and physics configurations used for CSI are described in Shaw et al. (2004). The grid configuration of WRF that was used in this study was 5-km horizontal resolution with 42 vertical levels. For the 0600 UTC model initialization time, two different simulations were produced from the WRF utilizing different initial conditions but identical lateral boundary conditions based on the 6-hour forecast from the 0000 UTC NCEP Eta. The "operational" run (hereafter referred to as the WRF-Hot) was initialized with Local Analysis and Prediction System (LAPS; Albers et al. 1996; Shaw et al. 2001) and incorporates the diabatic effects of micophysical species through the analysis of satellite and radar data (Shaw et al. 2004) into the initial conditions. The second comparison run (hereafter referred to as the WRF-Cold) does not include any diabatic effects and represents a routine model initialization. The WRF-Cold is initialized by simply interpolating the 6-hour forecast from the 0000 UTC Eta run to the WRF grid. The initial conditions from the 0600 UTC Eta run are not used in this process. The 0600 UTC run from the WRF-Cold is essentially the same as the first-guess field being input into the 0600 UTC run of LAPS used for the WRF-Hot runs. Therefore, the difference between the WRF-Cold and the WRF-Hot is that the WRF-Hot adds current data to a first-guess (i.e., the Eta 0600 UTC forecast) and runs the LAPS analysis and cloud-balance procedure. An advantage of using LAPS initial conditions is that LAPS is able to ingest a wide variety of observational data that may not have been included in the NCEP initial conditions for the Eta. Two other model simulations with initialization times of 1500 and 2100 UTC were also produced at JAX for the WRF-Hot but will not be discussed further in this document. The interested reader may find statistics for these forecasts on the CSI verification web site (above).

Eta Mesoscale Model – The Eta model is NCEP's primary forecast model used for short-range operational forecasting. The description of the model formulation and background information can be obtained from Black (1994). During the CSI evaluation period the Eta model's horizontal resolution was 12 km. On 8 July 2003, upgrades to the Eta were implemented at NCEP, which included modifications of the cloud microphysics and radiation, improved precipitation and cloud predictions through assimilation of cloud information from satellites, assimilation of radar-derived wind fields, and an expanded suite of forecast products (Ferrier et al. 2003).

4. Observations

The observations used to evaluate the forecasts were provided by the Meteorological Assimilation Data Ingest System (MADIS; *www-sdd.fsl.noaa.gov/MADIS/network_info.html*; Miller and Barth 2003; Barth et al. 2002). MADIS is a distribution system that accumulates observations from various data sources, applies limited quality control procedures to the observations, and makes those observations available to users. The observations from MADIS used to evaluate the WRF forecasts were provided by AWS Convergence Technology, Florida Mesonet, Multiagency Weather Stations, Weather for You, Florida Automated Weather Network (FAWN), University of Southern Florida Coastal Ocean Monitoring and Prediction System, and the National Weather Service hourly surface observations (METAR reports). For the precipitation verification, the Hydrometeorological Automated Data System (HADS) data were also utilized to increase the sample size. Although the number of stations available for verification varied each day, it is important to note that all models were verified against the same set of observations so this effect does not influence the results.

5. Verification Methods

At each valid time, the forecasts of precipitation, winds, temperature, and relative humidity from the WRF and Eta models are bilinearly interpolated to the observation location. Errors between the interpolated forecasts and the corresponding observations are computed and the results are stored. The precipitation provided by the models represents a precipitation forecast that accumulates over a 3-hour period. Therefore, before verifying the precipitation forecasts, the observations of precipitation are also accumulated over the 3-hourly periods to match the forecasts.

In order to produce a fair comparison, forecast data are subjected to an equalization process. This process operates in the following manner: when more than one model simulation is being analyzed, only those dates for which data are available for all of the models are used in the generation of the statistics. This allows for an equitable evaluation between two model forecasts where the statistical results are not affected by missing data for one or more model simulations.

A variety of statistics computed for each meteorological variable are summarized below and follow the descriptions set forth in Wilks (1995).

Bias for continuous variables such as temperature, winds, and relative humidity is simply the difference between the average forecast and the average observation. An unbiased forecast has a value of 0.

$$\text{BIAS}_{\text{(continuous)}} = \bar{y} - \bar{o},$$

where $\bar{y}$ represents the mean value of the model forecasts, and $\bar{o}$ represents the mean value of the observations.

Bias for dichotomous (binary) forecasts for variables such as precipitation is simply the ratio of the number of "yes" forecasts to the number of "yes" observations. Unbiased forecasts exhibit a Bias=1, indicating that the event was forecast the same number of times as was observed.

$$\text{BIAS}_{\text{(dichotomous)}} = \frac{a+b}{a+c},$$

where **a, b**, and **c** represent the entries in the 2x2 contingency table for the number of forecasts hitting the criteria (Yes-Yes; a), forecasts missing (Yes-No; b) and underforecasts of the criteria (No-Yes; c).

Equitable Threat Score: This score is known by several other names including Gilbert Skill Score and Equitable Skill Score (Schaefer 1990) and is often used to verify precipitation forecasts where nonevents (no rain) occur much more frequently than events (rain) occur. The score ranges from negative one for poor forecasts to unity for the best possible forecasts.

ETS = a - C / (a+b+c) - C,

where **C = (a+b)(a+c)/n** and n is the total number of events. **C** represents the portion of forecasts that can be expected to be correct purely by chance.

Mean Absolute Error (MAE) is the arithmetic average of the absolute values of the differences between the members of each pair. MAE is 0 if the forecast is perfect and increases as discrepancies between the forecasts and observations become larger. It can be interpreted as the typical magnitude for the forecast error in a given verification dataset.

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

where, (y_k, o_k) are the kth of n pairs of forecast and observation values.

Mean Squared Error (MSE) is the average squared difference between the forecast and observation pairs. This measure is similar to the MAE except that the squaring function is used rather than the absolute value function. Since the MAE is computed by squaring forecast errors, it will be more sensitive to larger errors than will the MAE. The MSE increases from zero for perfect forecasts through larger positive values as the discrepancies between forecasts and observations become increasingly large.

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

Root Mean Squared Error (RMSE): Sometimes the MSE is expressed as its square root, which has the same physical dimension as the forecast and observations, and can also be thought of as a typical magnitude for forecast errors. Since the RMSE is a function of bias, the values of RMSE presented in this report are corrected for the bias.

$$\textbf{RMSE} = \sqrt{MSE}$$

$$\textbf{Bias-Corrected RMSE} = \sqrt{MSE - (Bias)^2}$$

6. Results

Statistical results for precipitation, wind, and temperature forecasts from the 0600 UTC run of the WRF-Hot, WRF-Cold, and the Eta are summarized in this section. Time series and box plots are also provided for forecast hours 3 and 12. Tables listing overall statistics for the 3-, 6-, 9-, 12-, 15-, 18-, 21- and 24-hour forecasts are presented in the Appendix. Statistical results for the relative humidity forecasts were difficult to interpret because of its dependency upon the temperature and are not summarized here. If the reader is interested, the results for relative humidity can be obtained through the RTVS Website (above).

6.1 Precipitation

ETS and bias scores for 3 and 12-hour forecasts are depicted graphically in Figs. 2–5. Overall, the ETS scores for WRF and Eta models at all thresholds are typically less than 0.1 (Figs 2, 3, and Table A1 in the Appendix). This may be indicative of the prevalent air mass convective regime over northern Florida in the summer. At 3 hours, the distributions of ETS scores for the three models range from less than 0.0 to nearly 0.3 (Fig. 4), but are much less at 12 hours (Fig. 5).

At 3 hours, the WRF-Hot maintains a larger ETS score for all thresholds (Fig. 2a) than does the WRF-Cold or the Eta, suggesting a slight improvement in precipitation forecast accuracy for the WRF-Hot. At thresholds greater than 0.1 inches, the WRF-Hot maintains a bias close to 1.0 (Fig. 2b) than do the other models, although the bias drops below 1.0 at thresholds greater than 0.5 inches indicating underforecasting by all the models. In addition, the WRF-Hot maintains a larger bias at all thresholds than does the WRF-Cold, indicating more rapid development of the precipitation field in the WRF-Hot runs.

At 12 hours (Fig. 3a), the Eta model has a higher ETS score for thresholds between 0.1 and 1.0 inches than the WRF-Hot or WRF-Cold. More interesting is the behavior of the bias statistic shown in Fig. 3b. The Eta model tends to overpredict precipitation for thresholds less than 0.5 inches and underpredict precipitation for thresholds less than 0.5 inches with a bias that goes from 5.0 to nearly 0.0. The WRF-Cold has a bias close to 1.0 with overprediction of amounts less than 0.5 inches and underprediction of higher amounts. The WRF-Hot overpredicts at all precipitation thresholds less than 2.0 inches with a bias over 2 for amounts less than 0.75 inches and a bias in the range of 2–1 for larger amounts. The bias at 12 hours (Fig. 3b) for the WRF-Hot is nearly double the bias for the WRF-Hot produced at 3 hours (Fig. 2b).

A very useful way to examine the distributions of the verification statistics is through box and whisker plots as shown in Figs. 4 and 5. The boxes enclose the middle 50% of the distribution with the middle line showing the median value. The ends of the whiskers (the vertical lines) present the 5^{th} and 95^{th} percentile values of the distributions. Points at the top and bottom represent outliers with values in the upper and lower 5% of the distributions. One way to determine if the distributions of statistical results from two different forecast models are significantly different from one another is to evaluate the notches on the box plots. If the notches do not overlap from one distribution to the next, then the distributions are significantly different.

The box and whisker plots showing the distribution of the ETS and bias scores for the 3- and 12-hour forecasts from the WRF-Hot, WRF-Cold, and Eta for thresholds of 0.1 and 1.0 inch are presented in Figs. 4a and b and 5a and b, respectively.

Overall, for a threshold of 0.1 inch (Figs. 4a and b), the distribution of ETS scores increase for the WRF and Eta runs at 12 hours as compared to the 3 hour forecasts indicating a larger range of forecast skill at 12 hours. However, little improvement in the median from 3 to 12 hours is noted. At both 3 and 12 hours, the overlapping notches in the box plots indicate that the distributions for the WRF and Eta runs are similar, suggesting little difference in ETS skill between the three models. However, there is a better chance of obtaining "good" ETS score at 12 hours than at 3 hours. At a threshold of 1.0 inch (Figs 5a and b), the ETS scores continue to be extremely small for all models indicating little forecast skill. The dramatic increase in bias for the WRF-Hot from 3 to 12 hours is noted in Fig. 5b. As indicated by the nonoverlapping notches in the bias box plots at 12 hours, the distribution of bias values from the WRF-Hot is distinctly different from those produced for the Eta or the WRF-Cold. Overall, at a threshold of 1.0 inch, there is a better chance of obtaining an unbiased forecast with the WRF-Hot than with the other two models.

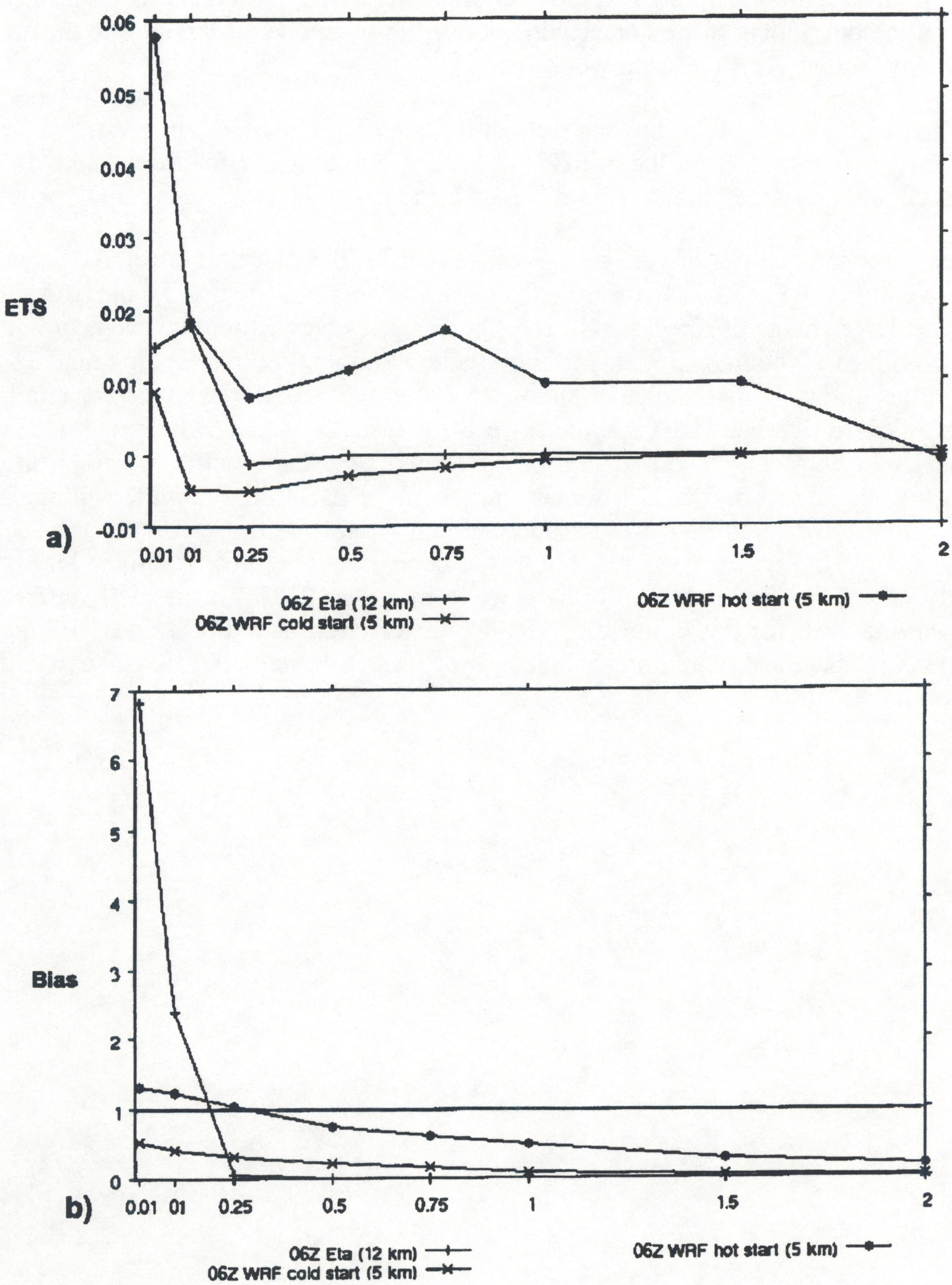


Figure 2. *ETS (a) and bias (b) for precipitation from WRF-Hot (*), WRF-Cold (x), and Eta (+) for 3-hour forecasts from the 0600 UTC run, computed from 1 June–31 August 2003. Horizontal line indicates unbiased forecast.*

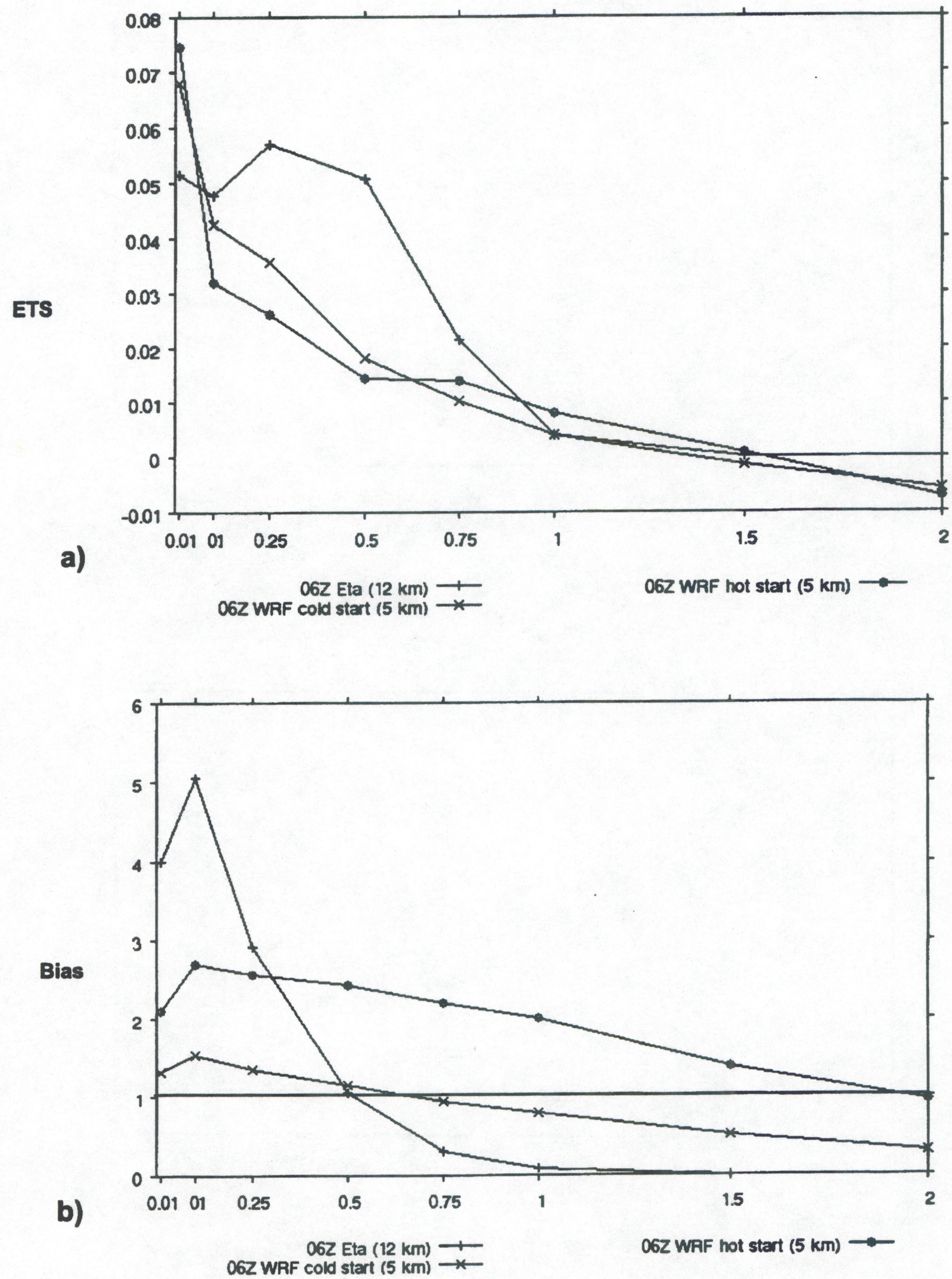


Figure 3. *Same as Fig. 2, except for 12-hour forecasts.*

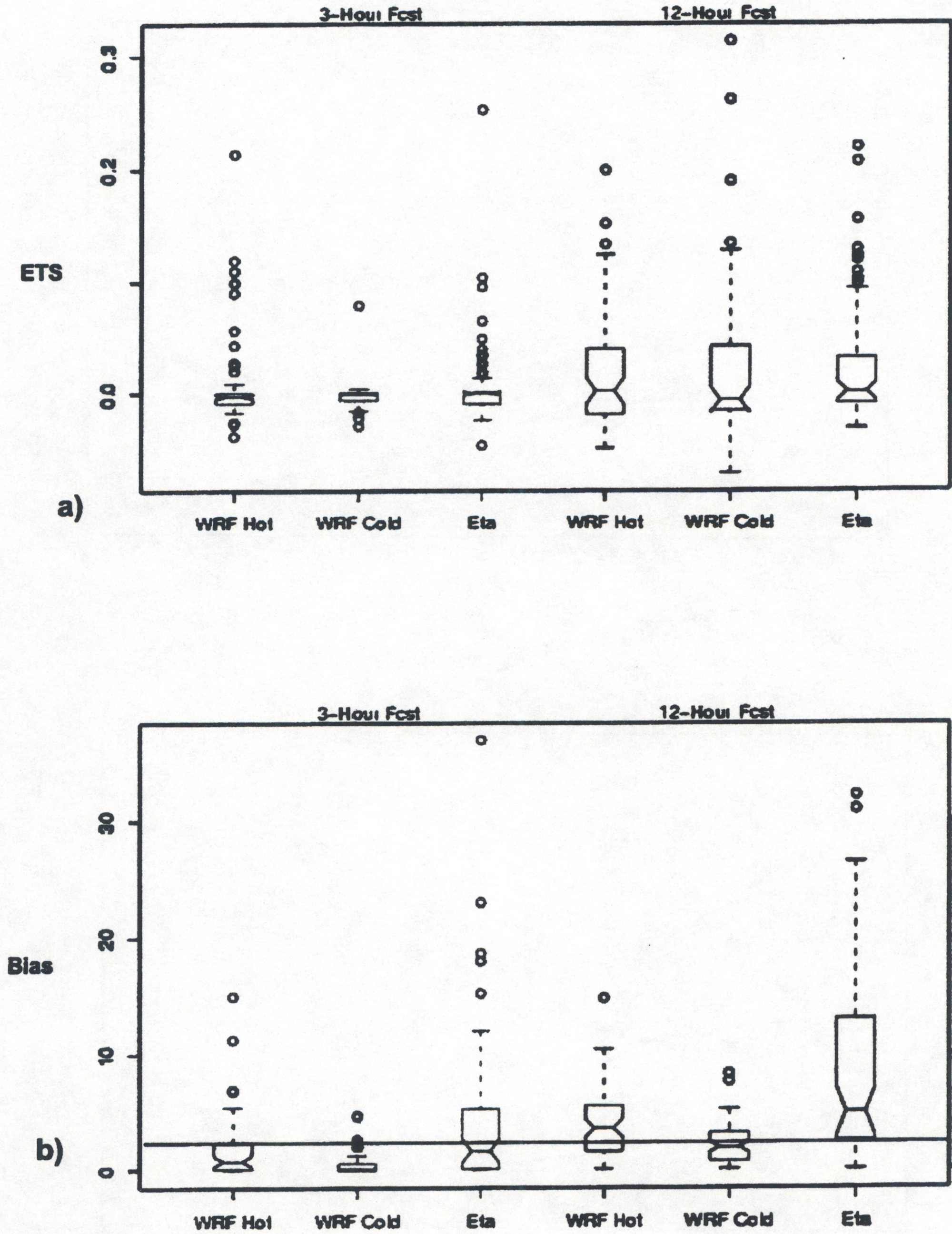


Figure 4. *Box plots of precipitation forecasts, threshold 0.1 inch, 3-hour (left) and 12-hour (right) forecasts from 1 June–31 August 2003; ETS (a) and Bias (b).*

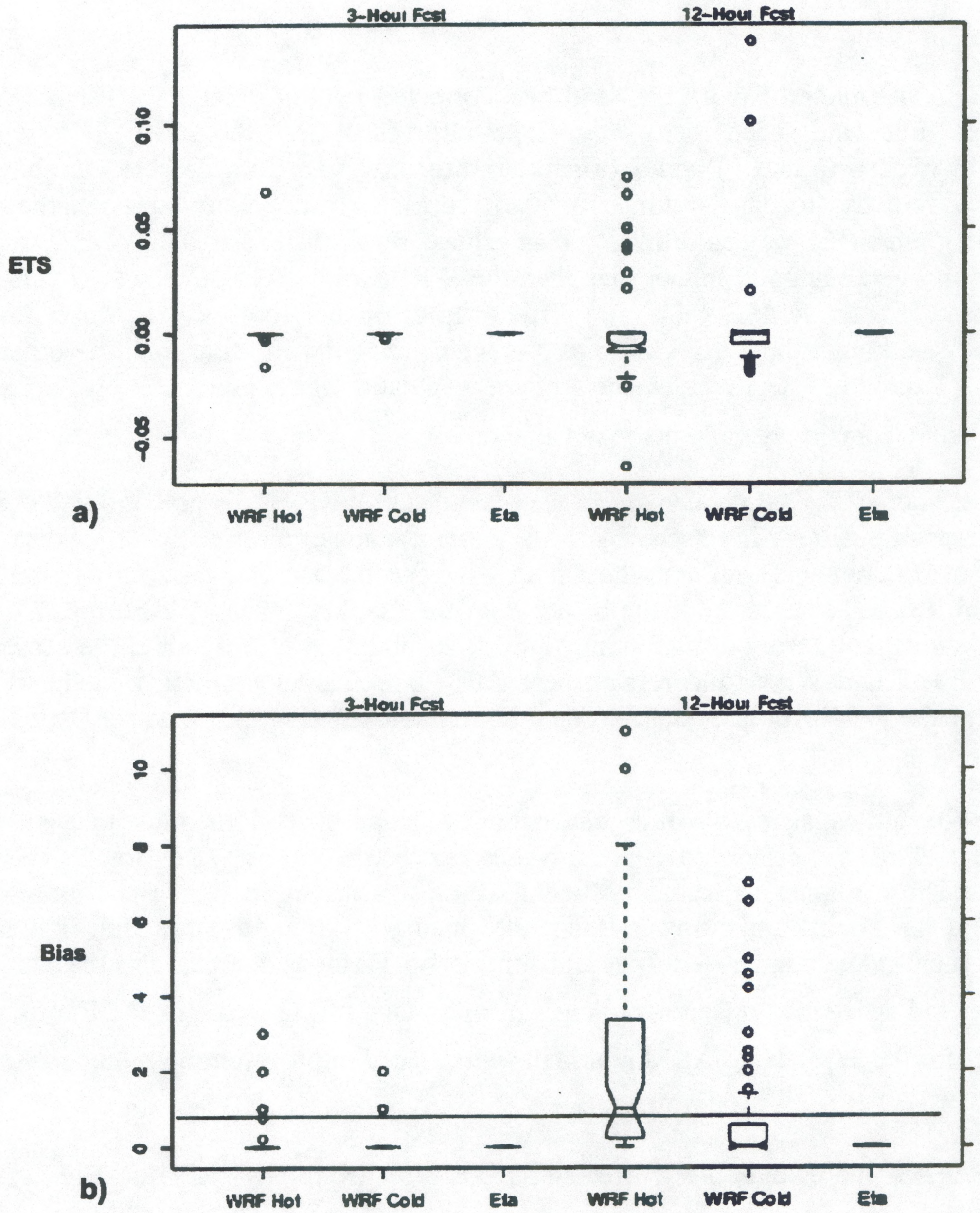


Figure 5. Same as Fig. 4, except threshold 1.0 inch; ETC (a) and Bias (b).

6.2. Winds

Time series plots of daily bias and bias-corrected RMSE for the WRF-Hot, WRF-Cold, and Eta wind speed for 3- and 12-hour forecasts from the 0600 UTC runs are shown in Figs. 6 (a and b) and 7 (a and b), respectively. Considerable variability in accuracy from day to day is evident in the figures. All models overpredict the wind speed at 3 hours (Fig. 6a), as shown by bias values greater than zero. However, Eta tends to have an overall slightly higher bias than the WRF-Hot or WRF-Cold at 3 hours. The daily bias-corrected RMSE values are difficult to distinguish between the three models. It appears as though the WRF-Cold has a greater variability in score than the other two models. Excluding the first few days of the evaluation, at 3 hours the bias-corrected RMSE values remain typically below a value of 2 m s^{-1}.

At 12 hours (Fig. 7a), a decrease in the bias for the WRF-Hot is noted indicating less overforecasting of the wind speed by the WRF-Hot compared to results at 3 hours. The bias for the Eta remains similar to the 3-hour wind speed bias. The variability in the bias-corrected RMSE values at 12 hours is larger for the Eta than for the WRF runs. The Eta RMSE values range from a maximum of 4.0 to a minimum of 1.0, while the scores for the WRF-Hot and WRF-Cold remain near 2.0 $\pm$ 0.5. Larger variability in the RMSE scores for the WRF-Hot is evident in the 3 hours forecasts (Fig. 6b) than for the 12-hour forecasts (Fig. 7b).

The box and whisker plots in Fig. 8a and b show the distributions of the bias and bias-corrected RMSE scores for the 3- and 12-hour forecasts from the WRF-Hot, WRF-Cold, and Eta. The median bias from the 3- and 12-hour wind speed forecasts (Fig. 8a) for WRF-Hot is significantly lower than the median value for the Eta, indicating improvement in the wind speed forecast from the WRF-Hot as compared to the Eta. The median wind speed bias difference ranged from a value of 1.3 m s^{-1} for the Eta to about 1.0 m s^{-1} for the WRF-Hot. At 12 hours, the wind speed improved in both models: 1.2 m s^{-1} for the Eta , 0.5 m s^{-1} for the WRF-Hot.

At 3 hours, the median bias-corrected RMSE value from the WRF-Hot (Fig. 8b) was roughly 1.7, nearly 0.5 m s^{-1} less than those generated for the Eta. At 12 hours, the difference in the median bias-corrected RMSE value between the WRF-Hot and the Eta is somewhat less. Both of these differences are statistically significant, suggesting that the WRF wind speed forecasts are an improvement over the Eta forecasts.

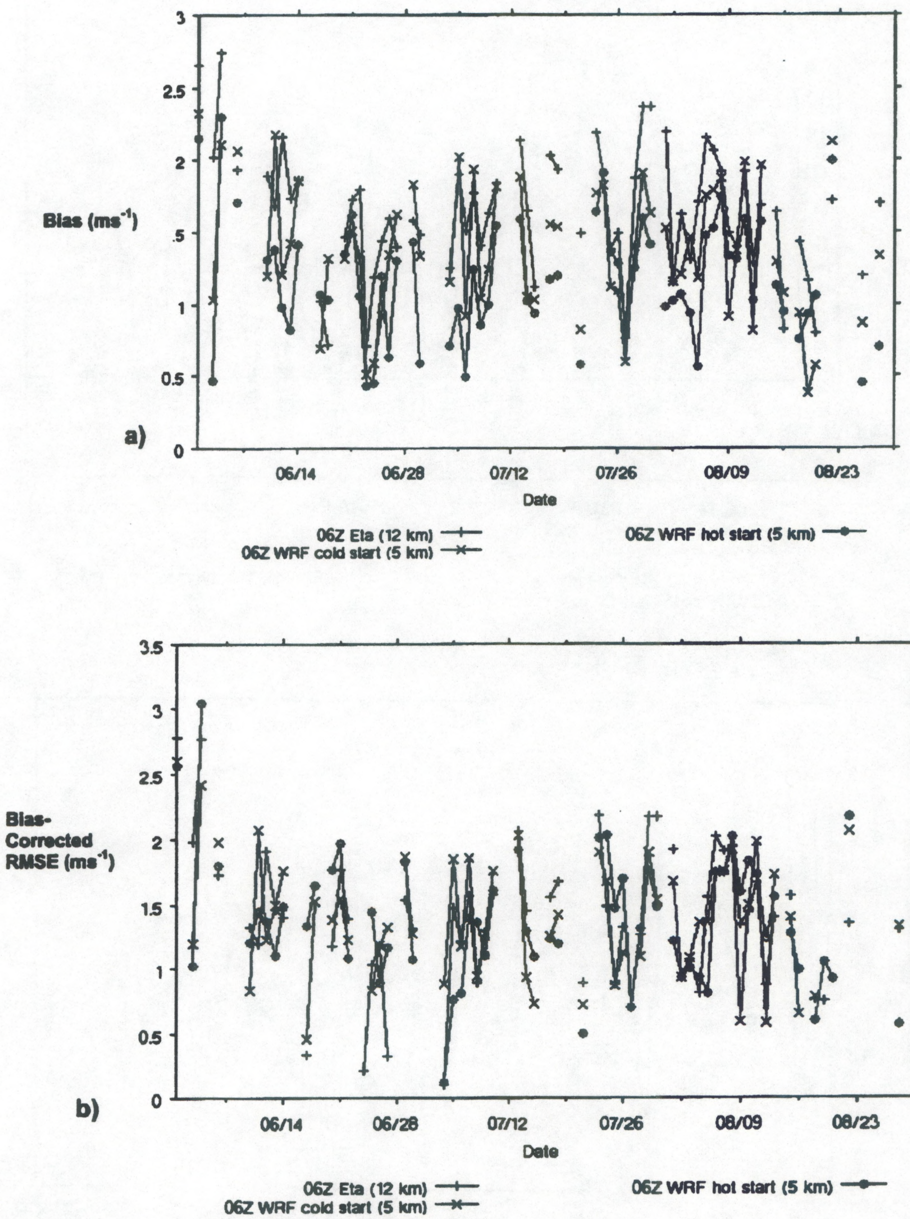


Figure 6. *Time series plots of daily bias (m s^{-1}) (a) and RMSE (m s^{-1}) (b) for 3-hour wind speed forecasts from WRF-Hot (*), WRF-Cold (x), and Eta (+) for 3-hour forecasts from the 0600 UTC run, computed from 1 June–31 August 2003.*

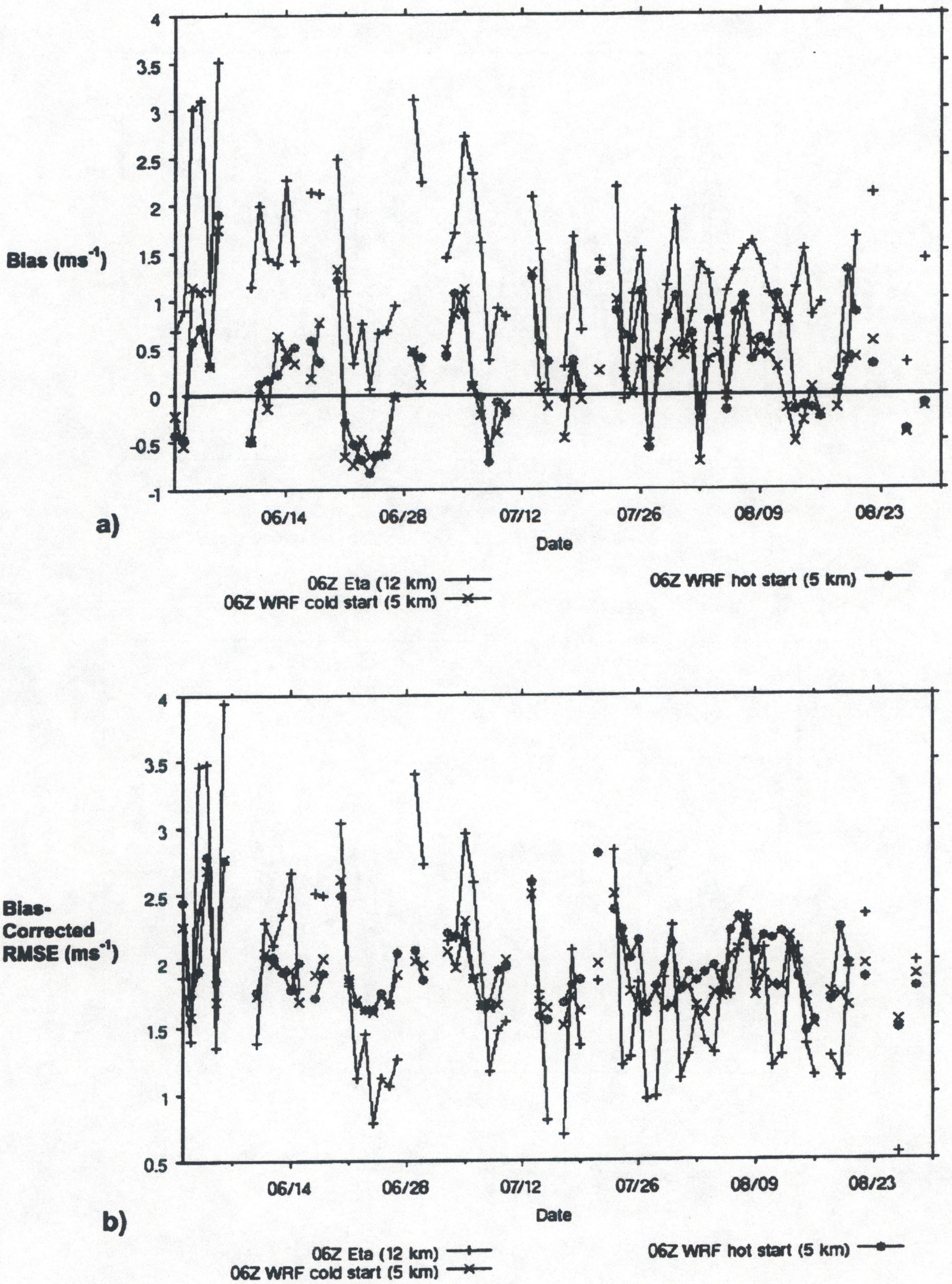


Figure 7. *Same as Fig. 6, except for 12-hour forecasts.*

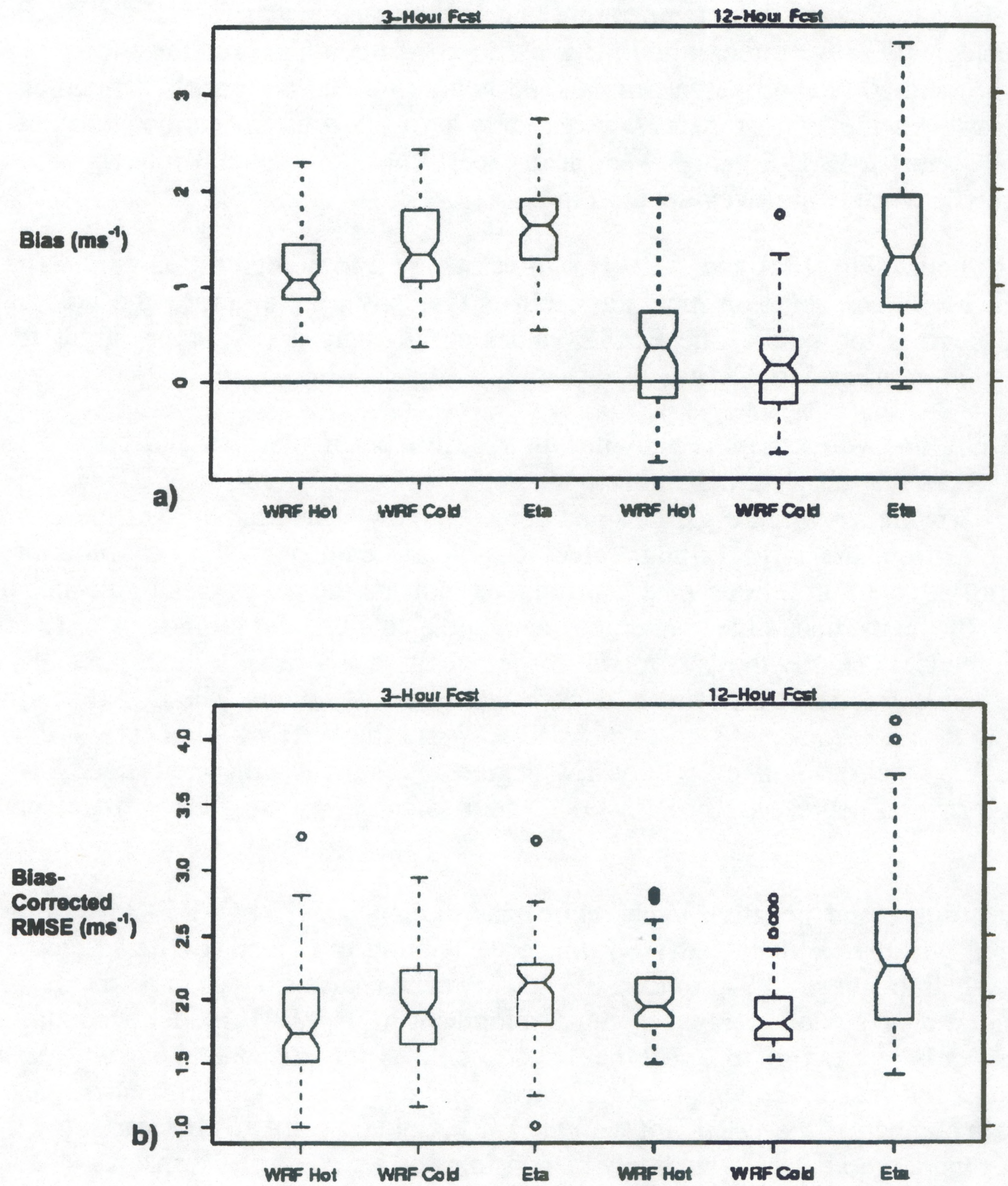


Figure 8. *Box plots of 3-hour (left) and 12-hour (right) wind speed forecasts (m s^{-1}), from 1 June–31 August 2003; Bias (a) and bias-corrected RMSE (b).*

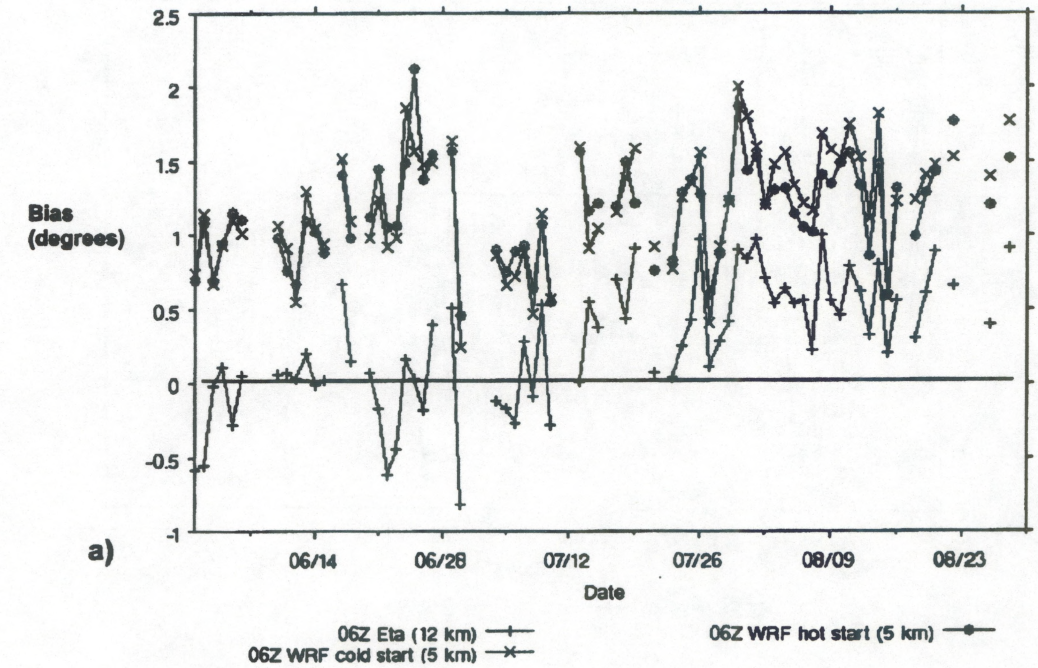
6.3 Temperature

Time series plots of daily temperature bias and bias-corrected RMSE for WRF-Hot, WRF-Cold, and Eta for forecast hours 3 and 12 for the 0600 UTC run are shown in Figs. 9 (a and b) and 10 (a and b). In general, at 3 hours (Fig. 9a) all models contain a warm bias. However, the WRF forecasts were nearly 1 to 1.5 degrees warmer than the Eta. The bias-corrected RMSE values were nearly identical between the 3 models with little separation between the lines evident in Fig. 9b.

At 12 hours (Fig. 10a), the WRF-Hot forecasts are 2 to 3 degrees too cold. The Eta also indicates a cold bias on most days, but is typically only a degree too cold and is rarely 3 degrees too cold. The RMSE values at 12 hours are typically larger for the WRF-Hot as compared to the Eta.

The box and whisker plots showing the distribution of the bias and bias-corrected RMSE scores for the 3- and 12-hour temperature forecasts from the WRF-Hot, WRF-Cold, and Eta are shown in Figs. 11 a and b, respectively. The dramatic difference in the temperature from the 3- to 12-hour forecast is clearly evident in the box plots of bias (Fig. 11a). Overall, at 3 hours, the distributions indicate the warm bias in all models as indicated by the median value greater than zero and a cold bias for all models at 12 hours. The Eta model is clearly the best model for temperature forecasts with the smallest bias and the bias-corrected RMSE errors at both the 3 and 12 hours. The difference in the median bias value from the 3- to 12-hour forecasts in the WRF-Hot is nearly 4 degrees. For the Eta, the difference is only 1.2 degrees. The bias-corrected RMSE values increased several degrees from the 3- to 12-hour forecast with smaller errors evident in the Eta.

The positive and negative temperature bias exhibited by the WRF-Hot during the evaluation was the result of the WRF model generating too much low-level cloudiness due to a bias in the LAPS-provided 3D temperature analysis. The bias in the LAPS temperature analysis led to rapid cloud development in the WRF model accounting for 12-hour (early-afternoon) forecasts being too cool. This mechanism is only partially responsible for the surface temperature forecast bias in the WRF-Hot forecasts. One can see that in comparing the WRF-Hot with the WRF-Cold (non-LAPS) forecasts, the LAPS initial temperature analysis was likely responsible for a degree of warm bias at 3 hours and a degree of cold bias at 12 hours. This analysis bias problem was diagnosed and has been corrected. An additional contributor to the temperature bias is that all of the models fail to accurately predict the full amplitude of the diurnal temperature curve, and this was especially true in the WRF runs where the PBL schemes "flattened" the diurnal curve in the forecasts. The implementation of the PBL schemes needs further investigation (Shaw et al. 2004).



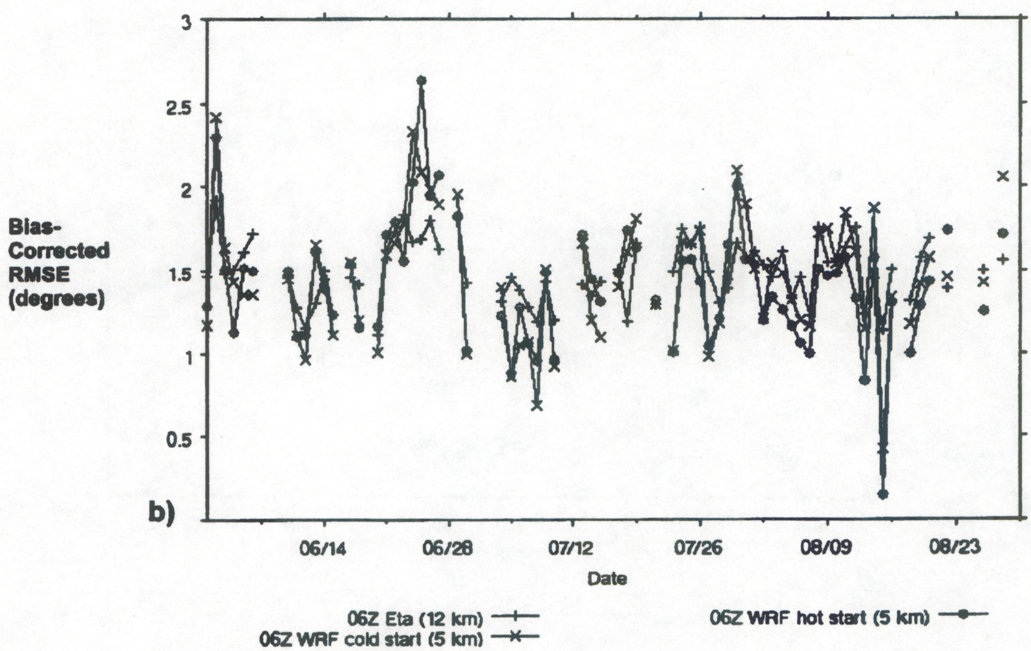
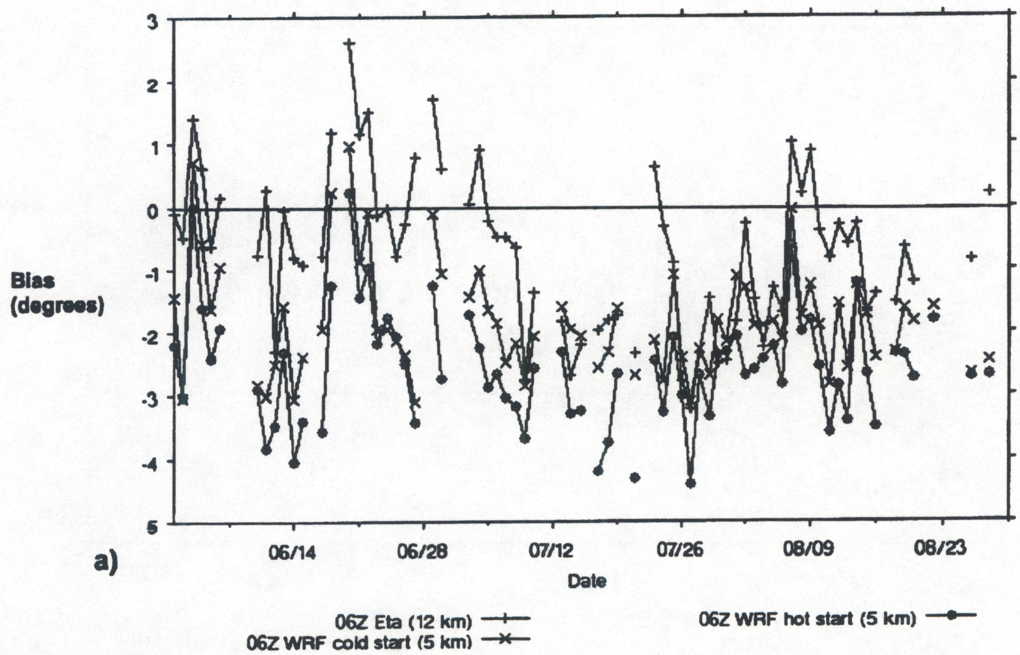


***Figure 9**. Time series plots of daily bias (K) (a) and RMSE (b) for temperature forecasts (K) from WRF-Hot (*), WRF-Cold (x), and Eta (+) at 3 hours from the 0600 UTC run, computed from 1 June–31 August 2003.*



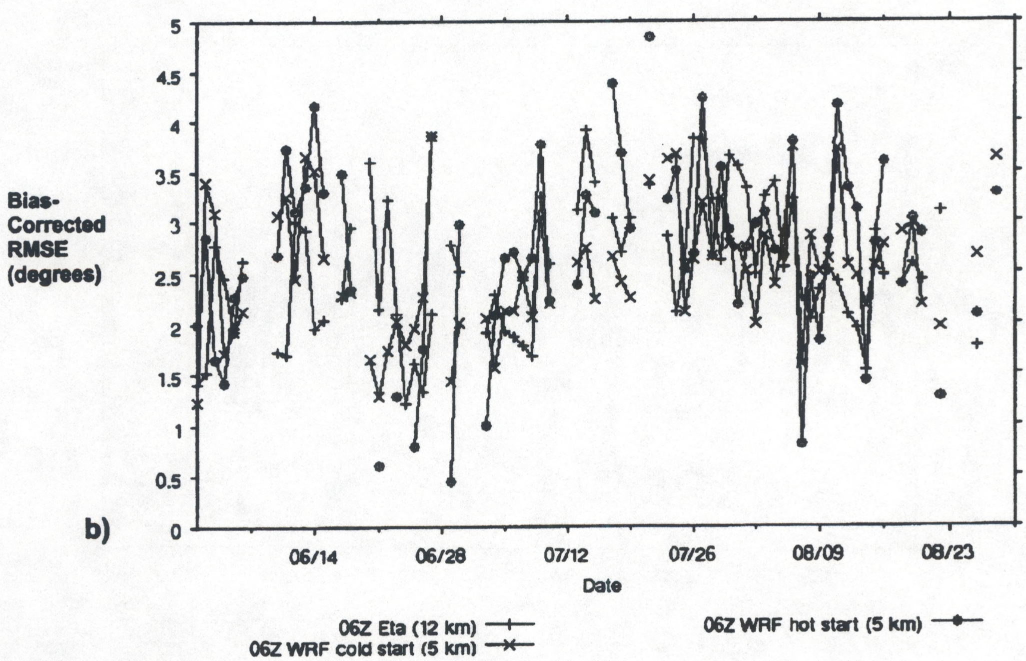


Figure 10. Same as Fig. 9, except for 12-hour forecasts.

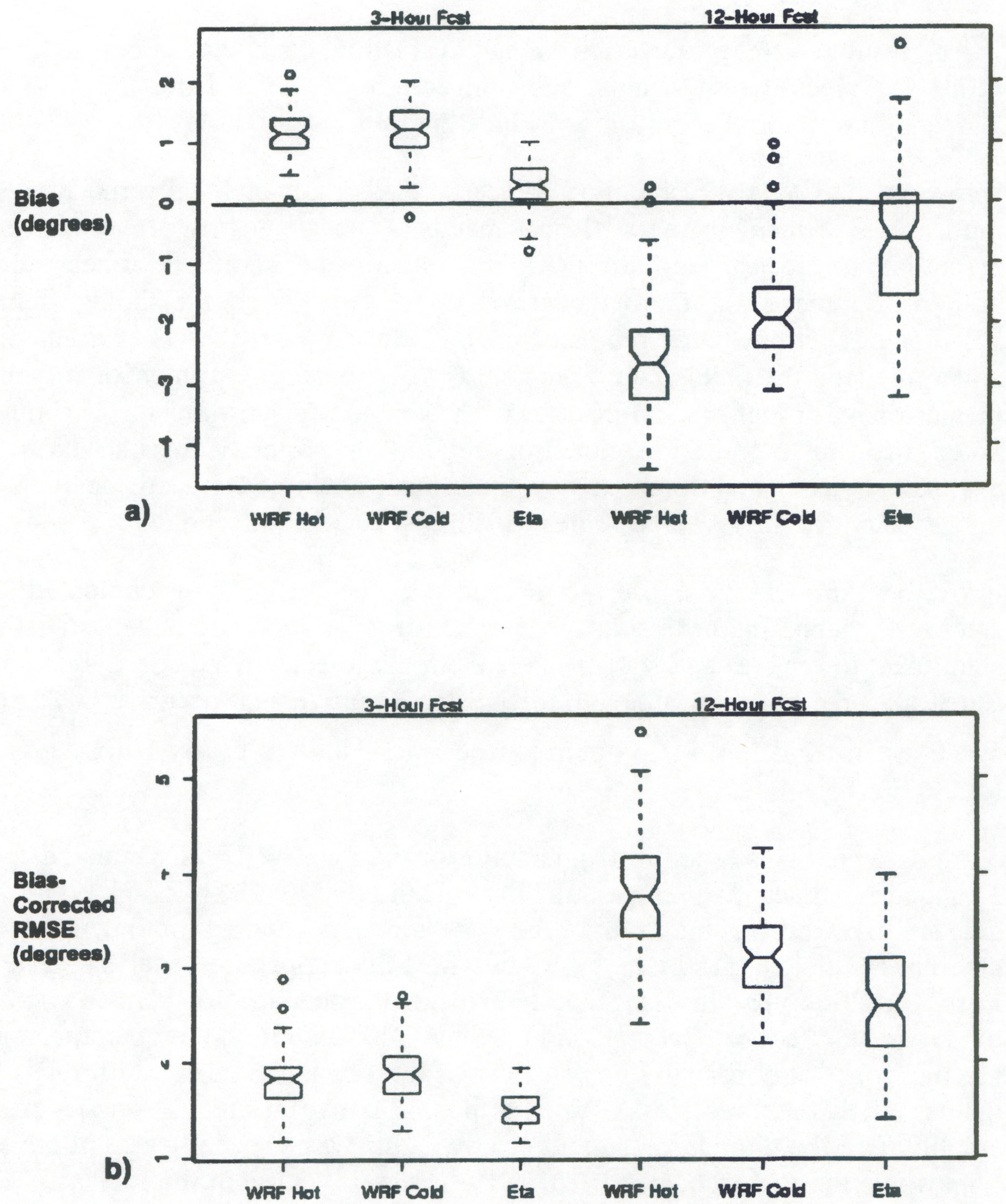


Figure 11. *Box plots of 3-hour (left) and 12-hour (right) temperature forecasts, from 1 June–31 August 2003; Bias (a) and bias-corrected RMSE (b).*

7. Summary and Future Work

Statistical results were presented from an evaluation conducted from 1 June–31 August 2003 for precipitation, winds, and temperature forecasts from the WRF-Hot, WRF-Cold, and Eta. Overall conclusions include the following:

- Improvement in local precipitation forecasts as a result of locally run numerical models was evident in the statistical results. The greatest improvement in the precipitation forecasts from the WRF-Hot was noted at short lead times and over 0.25 to 1.5 thresholds. Improvement was also noted at larger thresholds where the WRF-Hot continued to produce precipitation while the Eta dramatically underpredicted the precipitation amounts. At 3 hours and at thresholds typically greater than 1.0 inch, little forecast skill was noted for all models. It should be noted that the summer precipitation regime in northern Florida during the evaluation was mainly dominated by an air mass of single convective cells, which may account for the low ETS scores for all models at all thresholds.

- Improvement in local wind forecasts produced by the WRF-Hot as compared to the Eta was evident in the results. Overall, the WRF models had a statistically significant lower bias and statistical errors in the forecast wind speed than did the Eta at all forecast hours, although the greatest improvement occurred at 12 hours. The error in wind speed at 3 hours for the WRF-Hot was nearly 1 m s^{-1} less than for the Eta.

- No improvement was noted in temperature for the local WRF as compared to Eta. The larger and negative temperature biases exhibited by the WRF-Hot during the evaluation period was the result of the WRF model generating too much low-level cloudiness due to a bias in the LAPS-provided 3D temperature analysis. The bias in the LAPS temperature analysis led to rapid cloud development in the WRF model and consequent shading, and thus cooling of afternoon temperatures. This mechanism is only partially responsible for the surface temperature forecast bias in the WRF-Hot forecasts. All of the models fail to accurately predict the full amplitude of the diurnal temperature curve. This was especially true in the WRF runs where the PBL schemes "flattened" the diurnal curve in the forecasts (Shaw et al. 2004).

Even though the WRF is in the early stages of development, the results presented in this report are very encouraging, not only for WRF development, but also for improving local forecasts. To improve the temperature and moisture forecasts, the implementation of the radiation schemes should be further investigated. In addition, a second evaluation exercise is planned for winter 2003–2004 to assess the WRF performance in different weather regimes and to take advantage of upgrades to the JAX system.

Acknowledgments

The authors would like to thank the Coastal Storms Initiative for funding this project. We would also like to thank Pat Welsh at WFO JAX for his perspective and insight into this project. The authors would also like to thank Barbara Brown and Nita Fullerton for their helpful reviews of this paper.

Appendix A

Overall Statistics for Precipitation, Winds, and Temperature Forecasts from the WRF and Eta Models

Tables summarizing the overall statistics for the precipitation, wind, and temperature forecasts from the WRF-Hot, WRF-Cold, and the Eta are presented. The MSE and RMSE scores for the temperature and wind forecasts have been corrected for the bias.

A.1 Precipitation

Overall scores for the bias and ETS for precipitation from the 0600 UTC run, 3-, 6-, 9- 12-, 15-, 18-, 21-, and 24-hour forecasts of the WRF-Hot, WRF-Cold, and Eta for thresholds 0.01, 0.10, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0 inches from 1 June–31 August 2003 are presented in Table 1. In general, the results in Table 1 suggest that for all forecast hours, the WRF-Hot and WRF-Cold tend to produce more precipitation at thresholds that are greater than 1.0 inch than the Eta as indicated by the larger bias values for the WRF than for the Eta. At very small precipitation thresholds (e.g., 0.01 and 0.1 inches), the Eta overpredicts the precipitation nearly twice as much as the WRF forecasts with bias values that are more than double for the Eta than for the WRF. For all models, the ETS scores are typically largest at small precipitation thresholds.

Table 1. ETS and Bias generated for precipitation forecasts from the 0600 UTC run of the WRF-Hot, WRF-Cold, and Eta models from 1 June–31 August 2003.

	Precipitation Threshold (inches)							
Model	**0.01**	**0.10**	**0.25**	**0.50**	**0.75**	**1.00**	**1.50**	**2.00**
3-hour Forecasts ***ETS***								
WRF-Hot	0.1	0.02	0.01	0.01	0.02	0.01	0.01	0.0
WRF-Cold	0.01	-0.01	0.0	-0.01	0.0	0.0	0.0	0.0
Eta	0.02	0.02	0.0	0.0	0.0	0.0	0.0	0.0

Bias								
WRF-Hot	1.3	1.2	1.0	0.7	0.6	0.5	0.3	0.2
WRF-Cold	0.5	0.0	0.3	0.2	0.2	0.1	0.02	0.0
Eta	6.8	2.4	0.1	0.0	0.0	0.0	0.0	0.0
6-hour Forecasts ***ETS***								
WRF-Hot	0.1	0.02	0.02	0.01	0.01	0.01	0.01	0.01
WRF-Cold	0.04	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Precipitation Threshold (inches)							
Model	**0.01**	**0.10**	**0.25**	**0.50**	**0.75**	**1.00**	**1.50**	**2.00**
6-hour Forecasts ***ETS***								
Eta	0.03	0.02	0.0	0.0	0.0	0.0	0.0	0.0
Bias								
WRF-Hot	1.9	2.2	1.8	1.5	1.2	1.0	0.7	0.5
WRF-Cold	0.9	1.0	0.7	0.6	0.4	0.3	0.1	0.1
Eta	5.4	4.4	0.9	0.1	0.0	0.0	0.0	0.0
9-hour Forecasts ***ETS***								
WRF-Hot	0.1	0.03	0.02	0.01	0.01	0.01	0.01	0.0
WRF-Cold	0.1	0.02	0.01	0.0	0.0	0.0	0.0	0.0
Eta	0.04	0.03	0.03	0.01	0.0	0.0	0.0	0.0
Bias								
WRF-Hot	1.9	2.5	2.2	2.1	1.8	1.5	1.0	0.6
WRF-Cold	1.0	1.2	1.1	0.9	0.7	0.5	0.3	0.1
Eta	4.8	5.7	2.2	0.4	0.1	0.0	0.0	0.0

12-hour Forecasts ***ETS***							
WRF-Hot	0.1	0.03	0.02	0.01	0.01	0.01	0.0	0.0
WRF-Cold	0.1	0.04	0.03	0.02	0.01	0.0	0.0	0.0
Eta	0.1	0.04	0.1	0.1	0.02	0.0	0.0	0.0
Bias								
WRF-Hot	2.1	2.7	2.6	2.4	2.2	2.0	1.4	0.9
WRF-Cold	1.3	1.5	1.4	1.2	0.9	0.8	0.5	0.3
Eta	4.0	5.0	2.9	1.0	0.3	0.1	0.0	0.0
15-hour Forecasts ***ETS***								
WRF-Hot	0.1	0.03	0.02	0.02	0.02	0.01	0.01	0.0
	Precipitation Threshold (inches)							
Model	**0.01**	**0.10**	**0.25**	**0.50**	**0.75**	**1.00**	**1.50**	**2.00**
15-hour Forecasts ***ETS***								
WRF-Cold	0.1	0.03	0.04	0.02	0.02	0.01	0.0	0.0
Eta	0.1	0.1	0.1	0.1	0.04	0.02	0.0	0
Bias								
WRF-Hot	2.1	2.8	2.7	2.7	2.6	2.3	1.7	1.2
WRF-Cold	1.8	2	1.8	1.6	1.3	1.1	0.7	0.4
Eta	3.1	4.1	2.7	1.4	0.1	0.3	0.0	0.0
18-hour Forecasts ***ETS***								
WRF-Hot	0.1	0.04	0.03	0.03	0.02	0.01	0.01	0.0
WRF-Cold	0.1	0.04	0.04	0.03	0.02	0.01	0.01	0.01
Eta	0.1	0.1	0.1	0.1	0.1	0.04	0.02	0.0

				Bias				
WRF-Hot	1.9	2.4	2.4	2.4	2.4	2.1	1.7	1.3
WRF-Cold	1.6	1.9	1.7	1.6	1.3	1.1	0.8	0.5
Eta	2.5	3.4	2.4	1.4	0.9	0.4	0.1	0
				21-hour Forecasts ***ETS***				
WRF-Hot	0.1	0.04	0.04	0.03	0.02	0.01	0.01	0.0
WRF-Cold	0.1	0.04	0.04	0.03	0.03	0.02	0.01	0.01
Eta	0.1	0.1	0.1	0.1	0.1	0.1	0.03	0.0
				Bias				
WRF-Hot	1.8	2.3	2.4	2.4	2.4	2.3	1.8	1.4
WRF-Cold	1.6	1.9	1.8	1.6	1.5	1.3	0.9	0.6
Eta	2.4	3.2	2.4	1.5	1.1	0.6	0.1	0.009
	Precipitation Threshold (inches)							
Model	**0.01**	**0.10**	**0.25**	**0.50**	**0.75**	**1.00**	**1.50**	**2.00**
				24-hour Forecasts ***ETS***				
WRF-Hot	0.1	0.04	0.04	0.03	0.02	0.02	0.01	0.0
WRF-Cold	0.1	0.04	0.04	0.04	0.03	0.03	0.02	0.01
Eta	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0
				Bias				
WRF-Hot	1.8	2.3	2.4	2.4	2.5	2.3	1.9	1.5
WRF-Cold	1.6	1.9	1.8	1.7	1.5	1.4	1	0.7
Eta	2.3	3.3	2.5	1.7	1.3	0.8	0.2	0.04

A.2 Winds

Overall error statistics for wind (speed, and u and v component) forecasts from the 0600 UTC run for the 3-, 6-, 9-, 12-, 15-, 18-, 21-, and 24-hour forecasts from the WRF-Hot, WRF-Cold, and Eta are presented in Table 2. Positive bias is noted in Table 2 for all models suggesting that all models overpredict the wind speed. For all lead times, the bias in wind speed for the WRF-Hot is smaller than the bias for the Eta. More specifically, the bias improves for the WRF-Hot over the Eta at forecast hours 9, 12, 15, and 18. The error statistics for wind speed (MAE, MSE, and bias-corrected RMSE) are generally smaller for the WRF-Hot than for the Eta with the exception of the 6-hour forecasts where the errors are nearly identical. The bias and error statistics for the u-component of the wind are roughly similar among all the models. The bias for the v-component of the wind is slightly larger for the Eta than for the WRF-Hot, particularly at forecast hours greater than 12. Interestingly, the largest bias-corrected RMSE errors in the v-component for all models occur at 12, 15 and 18 hours with values for all models ranging between 1.9 and 2.2, respectively. However, all models only slightly overpredict the v-component of the wind, with the largest errors produced for the Eta at forecast lead times greater than 12 hours.

Table 2. Verification scores generated from 1 June–31 August 2003 for wind forecasts (speed, and *u* and *v* components; m s^{-1}) from the 0600 UTC run of the WRF-Hot, WRF-Cold, and Eta models.

Wind Speed								
Forecast Length (hours)								
Model	**3-h**	**6-h**	**9-h**	**12-h**	**15-h**	**18-h**	**21-h**	**24-h**
Bias (ms^{-1})								
WRF-Hot	1.2	1.3	0.6	0.3	0.6	1.1	1.5	1.5
WRF-Cold	1.4	1.5	0.4	0.2	0.6	1.2	1.6	1.6
Eta	1.6	1.3	1.3	1.4	1.7	2.0	2.0	1.9

MAE (ms^{-1})								
WRF-Hot	1.5	1.6	1.5	1.6	1.7	1.7	1.8	1.8
WRF-Cold	1.6	1.7	1.4	1.5	1.7	1.7	1.8	1.8
Eta	1.8	1.6	1.7	1.9	2.2	2.3	2.2	2.1
Bias Corrected RMSE (ms^{-1})								
WRF-Hot	1.4	1.5	1.8	2.0	2.0	1.9	1.5	1.4
WRF-Cold	1.4	1.4	1.7	1.9	2.0	1.7	1.4	1.4
Eta	1.4	1.4	1.7	1.9	2.1	1.9	1.5	1.5

Wind u-component Forecast Length (hours)								
Model	**3-h**	**6-h**	**9-h**	**12-h**	**15-h**	**18-h**	**21-h**	**24-h**
Bias (ms^{-1})								
WRF-Hot	0.2	0.0	0.1	0.3	0.3	-0.2	-0.1	0.2
WRF-Cold	0.2	0.2	0.2	0.3	0.3	-0.0	0.1	0.3
Eta	0.2	0.0	0.1	-0.1	0.1	-0.0	0.0	0.1
MAE (ms^{-1})								
WRF-Hot	1.1	1.2	1.4	1.6	1.8	1.6	1.3	1.3
WRF-Cold	1.2	1.3	1.2	1.5	1.7	1.5	1.4	1.3
Eta	1.4	1.3	1.5	1.8	2.0	1.9	1.5	1.4
Bias Corrected RMSE (ms^{-1})								
WRF-Hot	1.5	1.6	1.8	2.1	2.2	2.0	1.7	1.6
WRF-Cold	1.4	1.7	1.6	1.9	2.1	1.9	1.7	1.7
Eta	1.8	1.6	1.9	2.2	2.5	2.3	1.8	1.8

Wind v-component Forecast Length (hours)								
Model	**3-h**	**6-h**	**9-h**	**12-h**	**15-h**	**18-h**	**21-h**	**24-h**
Bias (ms^{-1})								
WRF-Hot	0.4	0.5	0.4	0.3	0.3	0.5	0.7	0.9
WRF-Cold	0.8	0.7	0.3	0.3	0.5	0.7	0.9	0.9
Eta	0.8	0.6	0.7	1.1	1.5	1.6	1.4	1.3
MAE (ms^{-1})								
WRF-Hot	1.2	1.3	1.4	1.6	1.8	1.6	1.5	1.4
WRF-Cold	1.3	1.4	1.3	1.5	1.7	1.6	1.5	1.4
Eta	1.3	1.2	1.6	2.0	2.1	2.0	1.8	1.7
Bias Corrected RMSE (ms^{-1})								
WRF-Hot	1.6	1.7	1.9	2.1	2.3	2.0	1.8	1.6
WRF-Cold	1.4	1.6	1.7	2.0	2.2	1.9	1.6	1.5
Eta	1.5	1.5	1.9	2.1	2.2	1.9	1.7	1.6

A.3 Temperature

Overall statistics for temperature from the 0600 UTC run for 3-, 6-, 9-, 12-, 15-, 18-, 21-, and 24-hour forecasts of the WRF-Hot, WRF-Cold, and Eta from 1 June–31 August 2003 are presented in Table 3. The results indicate at all forecast hours little improvement in the WRF-Hot temperature forecasts as compared to the Eta. The bias values for the Eta at forecast hours less than 9 are smaller than those values for the WRF-Hot. The WRF-Hot has a 1–2.6-degree cool bias at hours 9 to 18, which is the coolest bias among all of the models. The largest errors for all models occur at forecast hours 9, 12, and 15. Although the WRF-Hot has the largest cold bias, the variation in MAE and bias-corrected RMSE errors between all the models is extremely small.

Table 3. Verification scores generated from 1 June–31 August 2003 for temperature forecasts from the 0600 UTC run of the WRF-Hot, WRF-Cold, and Eta models.

Temperature Forecast Length (hours)								
Model	**3-h**	**6-h**	**9-h**	**12-h**	**15-h**	**18-h**	**21-h**	**24-h**
Bias (degrees)								
WRF-Hot	1.1	0.6	-1.9	-2.6	-2.1	-1.0	0.4	0.9
WRF-Cold	1.2	1.1	-1.1	-1.8	-1.5	-0.5	0.8	1.1
Eta	0.3	-0.7	0.2	-0.7	0.2	1.0	1.0	0.7
MAE (degrees)								
WRF-Hot	1.5	1.2	2.4	3.1	3.0	2.0	1.4	1.4
WRF-Cold	1.6	1.4	1.9	2.6	2.6	1.8	1.5	1.5
Eta	1.2	1.1	1.8	2.1	2.1	1.9	1.6	1.5
Bias Corrected RMSE (degrees)								
WRF-Hot	1.5	1.5	2.2	2.8	3.1	2.3	1.7	1.5
WRF-Cold	1.5	1.4	2.2	2.6	2.9	2.4	1.9	1.5
Eta	1.5	1.1	2.3	2.6	2.7	2.2	1.8	1.7

10. References

Albers, S., J. McGinley, D. Birkenheuer, and J. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analysis of clouds, precipitation, and temperature. *Wea. Forecasting*, **11**, 273–287.

Barth, M.F., P.A. Miller, and A.E. MacDonald, 2002: MADIS: The Meteorological Assimilation Data Ingest System. *Symposium on Observations, Data Assimilation, and*

Probabilistic Prediction, Orlando, FL, Amer. Meteor. Soc., 20–25.

Black, T.L., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting*, **9**, 265–278.

Bogenschutz, P., P. Ruscher, P. Welsh, J. Mahoney, J. McGinley, M. Kay, B. Shaw, J. Smart, J. Savadel, and J. McQueen, 2004: Summer season verification of the first NWS operational WRF model forecasts from the NOAA Coastal Storms Initiative project in northeast Florida. *17th Conf. on Probability and Statistics in the Atmospheric Sciences*, Seattle, WA, Amer. Meteor. Soc., CD-ROM J12.1

Ferrier, B., Y. Lin, D. Parrish, M. Pondeca, E. Rogers, G. Manikin, M. Ek, M. Hart, G. DiMego, K. Mitchell, H. Chuang, 2003: *http://wwwt.emc.ncep.noaa.gov/mmb/tpb.spring03/tpb.html.*

Mahoney, J.L., J.K. Henderson, B.G. Brown, J.E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002: The Real-Time Verification System (RTVS) and its application to aviation weather forecasts. *10th. Conference on Aviation, Range, and Aerospace Meteorology, Portland,* OR, Amer. Meteor. Soc., 323–326.

Miller, P.A. and M.F. Barth, 2003: Ingest, integration, quality control, and distribution of observations from state transportation departments using MADIS. *19th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, 10.11.

Schaefer, J.T., 1990: The Critical Success Index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.

Shaw, B., M. Kay, J. Mahoney, J. McGinley, J. Smart, P. Welsch, A. Wildman, J. Savadel, P. Bogenschutz, and P. Ruscher, 2004: Applying the Weather Research and Forecast (WRF) model to National Weather Service Forecast Office operations. *20th Conference on Wea. Anal. and Forecasting,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, 1.2.

Shaw, B.L., J.A. McGinley, and P. Schultz, 2001: Explicit initialization of clouds and precipitation in mesoscale forecast models. *14th Conference on Numerical Weather Prediction*, Fort Lauderdale, FL, Amer. Meteor. Soc., J87–J91.

Welsh, P.T., A. Wildman, B. Shaw, J. Smart, P. Ruscher, J. Savadel, and J. McGinley, 2004: Implementing the Weather Research and Forecast (WRF) model in the National Weather Service Forecast Office (WFO). *20th Conference on Wea. Anal. and Forecasting/16th Conference on NWP*, Seattle, WA, CD-ROM 12.3.

Wilks, D.S., 1995: Statistical methods in the atmospheric sciences. *Academic Press*, San Diego, CA, 464 pp.