Evaluation of the 29-km Eta Model. Part I: Objective Verification at Three Selected Stations

PAUL A. NUTTER* AND JOHN MANOBIANCO

ENSCO, Inc., Cocoa Beach, Florida, and Applied Meteorology Unit, NASA/Kennedy Space Center, Florida

(Manuscript received 30 March 1998, in final form 15 September 1998)

ABSTRACT

This paper describes an objective verification of the National Centers for Environmental Prediction 29-km Eta Model from May 1996 through January 1998. The evaluation was designed to assess the model's surface and upper-air point forecast accuracy at three selected locations during separate warm (May–August) and cool (October–January) season periods. In order to enhance sample sizes available for statistical calculations, the objective verification includes two consecutive warm and cool season periods.

The statistical evaluation identified model biases that result from inadequate parameterization of physical processes. However, since the model biases are relatively small compared to the random error component, most of the total model error results from day-to-day variability in the forecasts and/or observations. To some extent, these nonsystematic errors reflect the variability in point observations that sample spatial and temporal scales of atmospheric phenomena that cannot be resolved by the model.

On average, Meso Eta point forecasts provide useful guidance for predicting the evolution of the larger-scale environment. A more substantial challenge facing model users in real time is the discrimination of nonsystematic errors that tend to inflate the total forecast error. It is important that users maintain awareness of ongoing model updates because they modify the basic error characteristics, particularly near the surface. While some of the changes in error were expected, others were not consistent with the intent of the model updates and further emphasize the need for ongoing sensitivity studies and localized statistical verification efforts.

Objective verification of point forecasts is a stringent measure of model performance, but when used alone, is not enough to quantify the overall value that model guidance may add to the forecast process. Therefore, results from a subjective verification of the Meso Eta Model over the Florida peninsula are discussed in the companion paper by Manobianco and Nutter.

1. Introduction

For several years, Model Output Statistics (MOS; Glahn and Lowry 1972, Carter et al. 1989) from models such as the National Centers for Environmental Prediction's (NCEP) Medium Range Forecast and Nested Grid Models have been used prevalently as sources of localized point forecast guidance. Given an adequately populated sample of runs in which the model configuration is not changed, MOS provides added value to the forecast process by statistically accounting for characteristic strengths and weaknesses in model forecasts at specific locations. However, NCEP is now entering an era where improvements in modeling capabilities are occurring so rapidly (McPherson 1994) that traditional applications of MOS may no longer be appropriate for newer models. On the other hand, the combination of data assimilation techniques, refinements in model physics, and advances in computing efficiency (McPherson 1994) are enabling the possibility for ever more accurate deterministic model point forecasts.

In order to maximize the benefits of point forecast guidance from newer models within an environment of ongoing changes, it is helpful for both model users and developers to maintain an objective awareness of the model's error characteristics at given locations. For example, the local development of techniques that help correct identifiable model errors in real time could improve objective point forecast accuracy (e.g., Homleid 1995; Stensrud and Skindlov 1996; Baldwin and Hrebenach 1998). Moreover, periodic examination of model error characteristics could help developers diagnose and correct possible deficiencies in the model's physical parameterizations.

In the spring of 1996, the Applied Meteorology Unit (AMU) began an evaluation of the NCEP 29-km Eta (Meso Eta) Model in order to document its error characteristics for the U.S. Air Force 45th Weather Squadron (45WS), the National Weather Service (NWS) at Melbourne, Florida (MLB), and the NWS Spaceflight Me-

* Current affiliation: School of Meteorology, University of Oklahoma, Norman, Oklahoma.

Corresponding author address: John Manobianco, ENSCO, Inc., 1980 N. Atlantic Ave., Suite 230, Cocoa Beach, FL 32931.
E-mail: johnm@fl.ensco.com

teorology Group (SMG). The mission of the AMU is to evaluate and transition new technology, tools, and techniques into the real-time operational weather support environment for the NWS MLB, SMG, and 45WS (Ernst and Merceret 1995). The NWS MLB is responsible for making daily regional forecasts and for providing warnings of hazardous weather across east-central Florida (Friday 1994). The 45WS provides forecast and weather warning support for ground processing and launch operations of the space shuttle and other expendable vehicles primarily at the Kennedy Space Center (KSC), Cape Canaveral Air Station (CCAS), and Patrick Air Force Base in east-central Florida (Boyd et al. 1995; Hazen et al. 1995). Among other responsibilities, SMG provides weather support for normal space shuttle end-of-mission and possible launch abort landing scenarios at locations around the world including KSC and Edwards Air Force Base (EDW), California (Brody et al. 1997).

The objective portion of the Meso Eta evaluation was designed to assess the model's point forecast accuracy at three selected locations that are important for NWS MLB, 45WS, and SMG operational concerns. Objective verification of point forecasts is a stringent measure of model performance, but when used alone, is not enough to quantify the overall value that model guidance may add to the forecast process. This is especially true for models with enhanced spatial and temporal resolution that may be capable of generating meteorologically consistent, though not necessarily accurate, mesoscale weather phenomena (e.g., Cortinas and Stensrud 1995). With this in mind, the AMU also performed a subjective verification of Meso Eta Model forecasts to help quantify the added value, which cannot be inferred solely from an objective evaluation. Results from the AMU's subjective verification of the Meso Eta Model over the Florida peninsula are discussed in a companion paper (Manobianco and Nutter 1999).

In this paper, results from the objective component of the Meso Eta Model verification at EDW, the Shuttle Landing Facility, Florida (TTS), and Tampa International Airport, Florida (TPA), are discussed. Emphasis is placed on establishing the Meso Eta Model's basic warm and cool season error characteristics at these three locations and on determining if model updates between the evaluation periods led to statistically significant changes in forecast accuracy. The TTS and EDW stations are selected because they are the primary and secondary landing sites for the shuttle. The TPA site is chosen to compare model errors at two coastal stations on the eastern (TTS) and western (TPA) edges of the Florida peninsula. Note that model sensitivity tests necessary to isolate the exact sources of forecast errors following Manning and Davis (1997) are beyond the scope of this study.

The paper is organized as follows. A brief overview of the Eta Model and its configuration is presented in section 2. Procedures for data collection and statistical analysis are described in section 3. Detailed statistical results describing Meso Eta surface forecast accuracy during 1996 and 1997 are presented in sections 4 and 5, respectively. Results for upper-air forecast accuracy are described in section 6. Finally, the results are summarized for more generalized applications in section 7, followed by a concluding discussion in section 8.

TABLE 1. Eta Model attributes from Black (1994), Janjic (1994), and Rogers et al. (1996).

Dynamics
Model top = 25 mb
Time step = 72 s
Semistaggered Arakawa E grid
Gravity wave coupling scheme
Silhouette-mean orography
Split-explicit time differencing
Physics
Explicit gridscale cloud and precipitation
Modified Betts–Miller convective adjustment
Mellor–Yamada (2.5) for free atmosphere vertical turbulent exchange
Mellor–Yamada (2.0) near ground
Geophysical Fluid Dynamics Laboratory radiation scheme
Viscous sublayer over water

2. Eta model overview

The primary mesoscale modeling efforts at NCEP are focused on the development of the Eta Model (Rogers et al. 1995). The original version of the Eta Model with a horizontal resolution of 80 km and 38 vertical layers replaced the Limited-Area Fine Mesh model in June 1993 (Black 1994). In October 1995, NCEP increased the horizontal resolution of the operational ''early'' Eta Model from 80 to 48 km. At the same time, a cloud prediction scheme (Zhao et al. 1997) was implemented and initial analyses were produced using the Eta Data Assimilation System (Rogers et al. 1996). In August 1995, NCEP also began running a mesoscale version of the Eta (Meso Eta) Model with a horizontal resolution of 29 km and 50 vertical layers (Mesinger 1996). Following model upgrades on 31 January 1996 (Chen et al. 1996; Janjic 1996a–c; Betts et al. 1997), the early and Meso Eta Model configurations became identical except for resolution and data assimilation procedures. The relevant numerics and physics of the Eta Model are summarized in Table 1.

NCEP implemented two major changes to the Eta Model's physical parameterizations during the AMU's objective evaluation period. On 18 February 1997, components of the soil, cloud, and radiation packages were updated in both models (Betts et al. 1997, hereafter BE97; Black et al. 1997, hereafter BL97; EMC 1997). These modifications were designed to help control excessive net shortwave radiation at the ground that led indirectly to a bias in the diurnal range of surface temperatures. In addition, the updates helped control ex-

TABLE 2. Definition of seasonal verification periods and notable Eta Model updates.

Verification period	Date began	Date ended	Notable Eta Model changes (EMC 1997)
1996 warm season	1 May 1996	31 Aug 1996	
1996 cool season	1 Oct 1996	31 Jan 1997	Radiation, cloud fraction, soil moisture, etc. (18 Feb 1997)
1997 warm season	1 May 1997	31 Aug 1997	Corrected PBL depth computation (19 Aug 1997)
1997 cool season	1 Oct 1997	31 Jan 1998	

cessive mixing of the planetary boundary layer (PBL) and a negative (dry) bias in surface dewpoint temperatures. On 19 August 1997, calculation of the model's PBL depth was adjusted to correct for an underestimation of vertical moisture transport out of the lowest model layers (EMC 1997). A portion of the results shown in section 5 indicate that combined effects of these changes led to identifiable and statistically significant changes in forecast accuracy for a few selected parameters.

3. Data and analysis method

The AMU's objective and subjective verification was originally designed to consider 29-km Eta Model forecast errors over separate 4-month periods from May through August 1996 (warm season) and from October 1996 through January 1997 (cool season). Given the ongoing changes to the Eta Model configuration and the small sample sizes obtained from these limited 4-month verification periods, the objective portion of the evaluation was extended to include secondary warm and cool season periods from May through August 1997 and October 1997 through January 1998, respectively. The correspondence between these twin-seasonal evaluation periods and relevant Eta Model updates is described in Table 2. The most substantial modifications were implemented in February 1997 at a time that falls between the 1996 and 1997 datasets. The timing of this update is convenient for the identification of changes in forecast accuracy, particularly for variables influenced by boundary layer processes.

Forecasts from the 0300 and 1500 UTC Meso Eta Model cycles were obtained via the Internet from the National Oceanic and Atmospheric Administration's (NOAA) Information Center (NIC) server.[1] During the evaluation period, the server provided forecasts of surface and upper-air parameters available at 1-h intervals, for projections out to 33 h in advance at more than 500 stations. NCEP extracted these surface and upper-air station forecasts from the Meso Eta Model grid point nearest to the existing observation sites.

[1] At the time of writing, forecast point and gridded data could be obtained from the NIC Web site via anonymous ftp (nic.fb4.noaa.gov).

Hourly surface observations from TTS, TPA, and EDW are used to verify Meso Eta point forecasts of 10-m wind speed and 2-m temperature and dewpoint temperature. Upper-air forecasts of wind speed, temperature, and mixing ratio are verified by using rawinsonde observations from CCAS (XMR), Tampa Bay (TBW), and EDW. Log–linear interpolation of data is used between reported pressure levels for verification at 25-mb intervals from 1000 to 100 mb. While surface forecasts are verified hourly, upper-air forecasts are verified only for those hours coinciding with the available rawinsonde release times.

Note that the surface and upper-air station forecasts obtained from the NIC server correspond to the rawinsonde observation sites located at XMR, TBW, and EDW. The surface observations at TTS and TPA are not collocated with their respective rawinsonde observation sites at XMR and TBW. However, the available sites are separated by not more than about 30 km (i.e., the Meso Eta Model grid spacing). In order to avoid confusion, all subsequent references to rawinsonde and surface verification will use the rawinsonde station identifiers XMR, TBW, and EDW.

The statistical measures used to quantify model forecast errors are the bias (forecast − observed), root-mean-square (rms) error, and error standard deviation. For interpretation of results, it is helpful to recognize that the total model error includes contributions from both systematic and nonsystematic sources. Systematic errors (model biases) are usually caused by a consistent misrepresentation of such factors as orography, radiation, and convection. Nonsystematic errors are indicated by the error standard deviation and represent the random error component caused by initial condition uncertainty or inconsistent resolution of scales between the forecasts and observations. While it could be possible to partially correct for known systematic errors by subtracting the bias, the nonsystematic errors are rather unpredictable in nature and may contribute to a degraded daily forecast product. In order to determine if model updates lead to a statistically significant annual change in forecast accuracy, a Z statistic (Walpole and Meyers 1989) is calculated for a given parameter and compared with the normal distribution using a 99% confidence level. Additional details regarding statistical calculations are provided in the appendix.

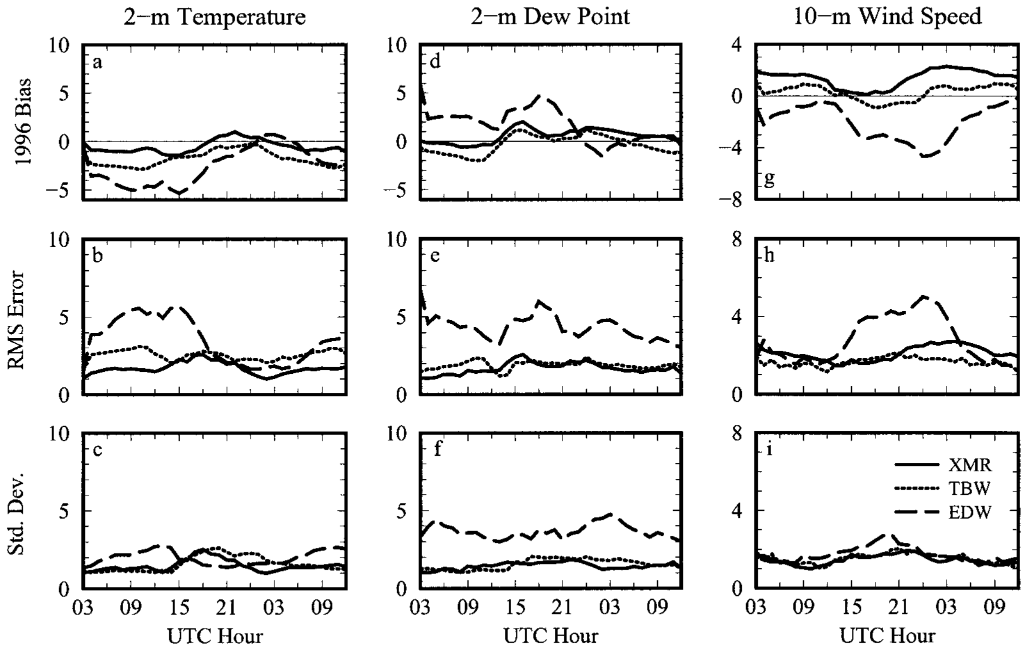

FIG. 1. Bias, rms error, and error standard deviation for 2-m temperature and dewpoint temperature (°C) and 10-m wind speed (m s^{-1}) forecasts from the 0300 UTC Meso Eta cycle. Results are plotted for the 1996 warm season as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed).

For quality control, gross errors in the data are screened manually and corrected, if possible. Errors that are greater than three standard deviations from the mean error (bias) are excluded from the final statistics. This procedure is effective at flagging bad data points and removes less than 1% of the data.

4. 1996 surface results

In the following section, Meso Eta point forecast error characteristics for 2-m temperature and dewpoint temperature and 10-m wind speed are established for the 1996 warm and cool seasons. Although statistics were calculated separately for the 0300 and 1500 UTC forecast cycles, only those from the 0300 UTC cycle are shown here. Results from the 1500 UTC cycle provide little additional information since positive or negative biases occur with comparable magnitudes at approximately the same time of day in both forecast cycles. Moreover, averaging data from both the 0300 and 1500 UTC cycles as a function of forecast duration tends to cancel out the diurnally varying errors. For most interests, the error characteristics described here for the 0300 UTC forecast cycle apply equivalently for the 1500 UTC cycle.

a. The 1996 warm season

1) THE 2-M TEMPERATURE

During the 1996 warm season, biases in 2-m temperature at XMR and TBW follow a diurnal cycle as the mean errors range from about $-3°$ to 1°C (Fig. 1a). The amplitude of the diurnal cycle is larger at EDW, with cold biases reaching almost -6°C during the early part of the forecast. More generally, the forecasts are excessively cold during nighttime hours and slightly warm during the afternoon hours.

Since forecast biases and corresponding rms errors are comparable in magnitude at EDW (Figs. 1a,b), the larger contribution to the total error for this location evidently is derived from a systematic model error. One possible explanation for this apparent model deficiency at EDW may be that the forecast point data extracted from the model are almost 250 m lower than the actual station elevation. The results at all three locations are also consistent with those from BE97 and BL97 who found an excessive range of summer temperatures due to radiation errors in the 1996 version of the 48-km Eta Model.

2) THE 2-M DEWPOINT TEMPERATURE

Warm season biases in 2-m dewpoint temperature at XMR and TBW are less than $\pm$2°C (Fig. 1d). Biases at EDW are positive during the first 21 h of the forecast cycle (Fig. 1d). When viewed in conjunction with the 2-m temperature bias in Fig. 1a, the net result is that forecasts for EDW are too cold and moist over this period.

The studies by BE97 (their Fig. 10b) and BL97 (their Fig. 4b) indicate excessive amounts of 2-m specific humidity in the forecasts at time zero using regionally

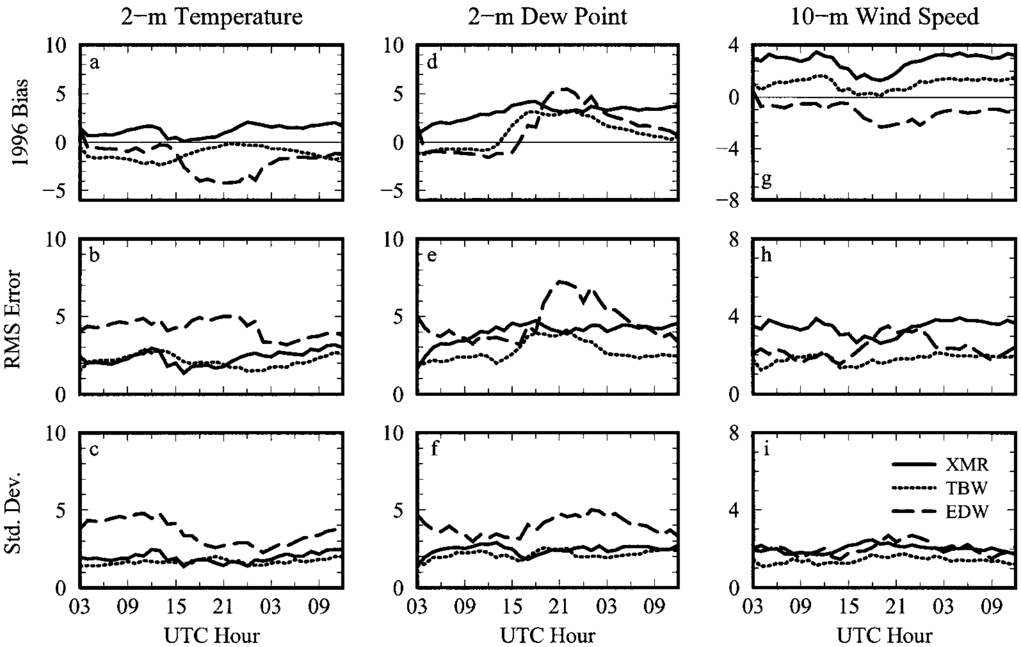

FIG. 2. Same as Fig. 1 but for the 1996 cool season.

averaged data during the summer. Their results also reveal that specific humidity levels are underforecast on average throughout the remainder of the forecast cycle. The zero-hour dewpoint errors shown here for EDW are consistent with results from those studies but the enduring positive (moist) bias indicates that regionally averaged statistics can mask important error characteristics that are specific to particular locations.

Some of the difficulties in forecasting dewpoint temperatures at EDW could relate to problems with PBL mixing and/or incorrect specification of soil moisture processes as discussed by BE97. Such difficulties would likely be exacerbated by the station elevation error at EDW and also by postprocessor errors while translating mixing ratios into 2-m dewpoint temperatures.

3) THE 10-M WIND SPEED

Warm season biases in the 10-m wind speed forecasts range from -5 to 0 m s^{-1} at EDW and from 0 to 2 m s^{-1} at XMR (Fig. 1g). The biases at TBW are less than ± 1 m s^{-1}. More generally, the 10-m wind speed forecasts at TBW are fairly good on average while those at XMR are slightly fast. At EDW, the wind speed forecasts are excessively slow, especially during the daytime hours. The relatively large increase in the magnitudes of biases and rms errors at EDW between about 1500 and 0300 UTC reflects a period during which systematic model errors compose the larger portion of the total forecast error (compare Figs. 1g–i).

b. The 1996 cool season

1) THE 2-M TEMPERATURE

During the 1996 cool season, 2-m temperature forecasts at EDW are on average about $-4°$ to 0°C colder than observed (Fig. 2a). The 2-m temperature forecasts at XMR (TBW) are on average 0° to 2°C warmer (colder) than observed. Over the first 12 h of the forecast cycle, large error standard deviations at EDW (Fig. 2c) suggest that nonsystematic errors contribute to a substantial portion of the total model error. During the middle part of the forecast cycle from about 1500 to 0300 UTC, the larger negative bias at EDW indicates that systematic model errors contribute more strongly to the total error.

Comparison of the warm and cool season error characteristics at each station reveals that the model errors vary locally by season. For example, the cool season temperature bias at XMR is positive (warm) throughout the forecast cycle (Fig. 2a) and does not clearly indicate the diurnal fluctuations shown during the previous warm season (Fig. 1a). At EDW, the nonsystematic error component is greater during the cool season (Figs. 1c, 2c) while the strong cold bias shifts toward the middle part of the forecast cycle. More detailed sensitivity studies are necessary to identify possible sources that contribute to seasonal changes in both systematic and nonsystematic model errors.

2) THE 2-M DEWPOINT TEMPERATURE

Cool season biases in 2-m dewpoint temperature at all three stations are mostly larger (wetter) than those of the previous warm season (Figs. 1d, 2d). The biases at TBW range from about $-1°$ to 3°C while at XMR, a moist bias of 3° to 4°C is evident throughout much of the forecast cycle. Qualitatively, the difference in error characteristics at XMR and TBW is notable given their relative proximity. Model biases at EDW follow similar fluctuations with time during both seasons, but reach slightly higher maximum values of around 6°C during the cool season at 2100 UTC.

The overall increase of cool season dewpoint temperature biases contributes to a corresponding growth in rms error at all three locations (Fig. 2e). This result suggests that systematic errors in Meso Eta Model dewpoint temperature forecasts are larger during the cool season for these stations. The strong moist biases shown in Fig. 2d for XMR, TBW, and EDW are not consistent with the dry bias demonstrated by BE97 and BL97. However, their studies utilized regionally averaged summertime data from the 1996 version of the 48-km Eta Model. In general, the results emphasize the need for localized, seasonally stratified verification to help understand the source of model errors and to enhance the utilization of point forecast guidance.

3) The 10-m wind speed

The existing positive (fast) bias for 10-m wind speed forecasts during the warm season at XMR increases by about 1 m s^{-1} during the cool season (Figs. 2g, 1g). Wind speed biases at TBW are comparable during both seasons while the slow bias at EDW improves in the cool season.

5. Impact of model updates on surface forecast accuracy

The results in section 4 for the 1996 warm and cool seasons demonstrate that Meso Eta Model error characteristics vary by location, season, and time of day even under the same model configuration. The original statistical evaluation was extended for a second year, in part, to enhance the quality of results by providing larger sample sizes. However, the NCEP model updates implemented between the 1996 and 1997 evaluation periods (Table 2) created an opportunity to examine changes in forecast errors and to determine if such changes were driven primarily by the model updates. As described below, a comparison of results between the 1996 and 1997 evaluation periods further highlights the necessity for model users to maintain an awareness of forecast accuracy at specific locations in lieu of ongoing changes.

Comparison of statistical results from 1996 and 1997 (full details not shown) reveals that many of the model biases described in section 4 were altered in 1997 while the nonsystematic error component remained mostly unchanged. Therefore, only the annual changes in model biases, or systematic errors, are described here. However the annual changes in bias are statistically significant (see appendix), the difference could be attributed to changes in the seasonal mean forecasts, observations, or a combination thereof. When the differences are explained largely by annual changes in mean forecast values, it is likely that the model updates led to the change in bias during 1997. Otherwise, the difference may reflect merely the interannual variability in the observations.

The Eta Model updates implemented in February 1997 were designed to decrease low-level temperatures and increase the low-level moisture (BL97). The results shown in Fig. 3 reveal that the updates led to identifiable and statistically significant changes in systematic error at XMR, TBW, and EDW between the 1996 and 1997 warm seasons. Some of these results are as follows.

- The existing cold, moist bias in 2-m temperature and dewpoint temperature forecasts at EDW became worse in 1997 (Figs. 3a,d). The decrease in the mean forecast temperature and increase in the mean forecast moisture is consistent with the February 1997 model updates. However the change exposes a more serious model error at this location, possibly related to the incorrect specification of station elevation.
- The 2-m dewpoint temperature forecasts at XMR and TBW became drier on average during the 1997 warm season (Fig. 3d). This change is opposite the response anticipated from the model updates. Additional sensitivity studies would be required to identify an alternative local forcing mechanism that may influence the systematic error for this location.
- Annual changes in the 2-m temperature biases at XMR and TBW are not statistically significant throughout most of the forecast cycle (Fig. 3c).
- The 10-m wind speed biases did not change in response to the model updates (Figs. 3g–i). Note that the model updates were not designed explicitly to affect wind components.

The Meso Eta Model upgrades also led to identifiable and statistically significant changes in systematic error between the 1996 and 1997 cool seasons. However, just as Meso Eta forecast error characteristics vary by season for each location, so did the nature of their response to model updates. Selected highlights from Fig. 4 include the following.

- A daytime warm bias was introduced for 2-m temperature forecasts at TBW (Figs. 4a,c). This increase in systematic error was not anticipated since the model updates were designed to reduce temperatures.
- There were no statistically significant changes in the error characteristics for temperature forecasts at EDW during the cool season (Figs. 4a–c). This is in contrast to the large reduction in temperature noted during the warm season (Figs. 3a–c).
- The existing moist bias in 2-m dewpoint temperature forecasts was reduced at XMR and TBW (cf. Fig. 2d

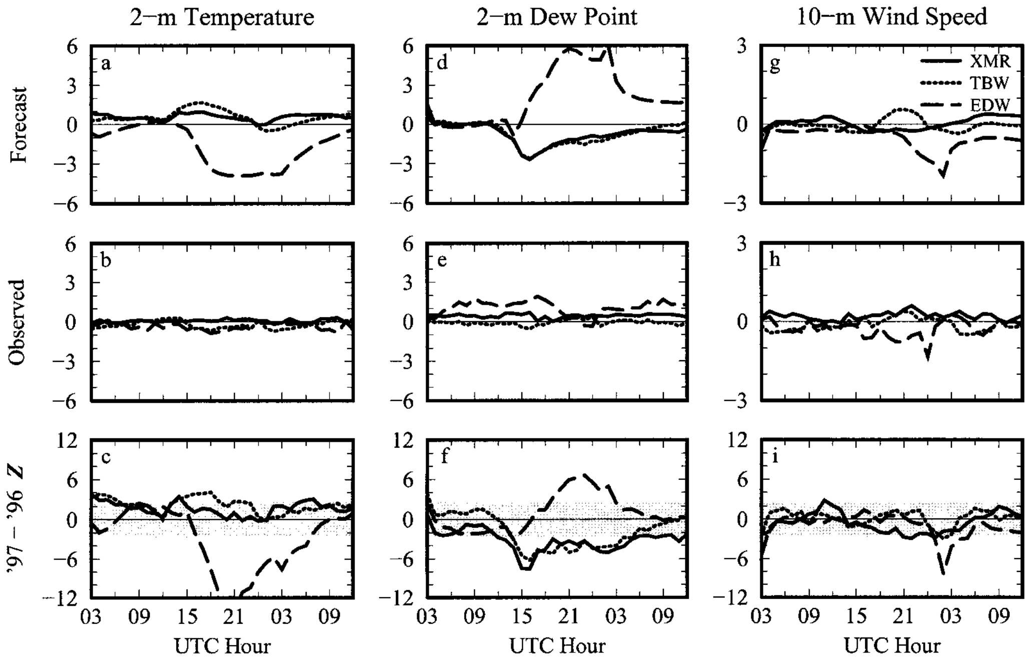

FIG. 3. Comparison of annual changes (1997–1996) in mean forecasts and observations for 2-m temperature and dewpoint temperature (°C) and 10-m wind speed (m s^{-1}) during the warm season. Values of the standardized Z statistic (bottom row) that lie outside the shaded region indicate that annual changes in forecast bias are considered statistically significant at the 99% confidence level (see the appendix). Results are plotted as a function of time at XMR (solid), TBW (dotted), and EDW (dashed).

and Figs. 4d–f). Since the model updates were designed to *increase* moisture, additional sensitivity studies would be required to identify an alternative local forcing mechanism that might have caused this change in systematic error.

6. Upper-air results

The AMU's original Meso Eta evaluation (MN97) was extended, in part, to enhance the quality of results by increasing sample sizes. For the surface parameters discussed previously in sections 4 and 5, it was not reasonable to combine data from 1996 and 1997 because model updates produced identifiable and statistically significant changes in forecast accuracy. However, examination of error statistics for the upper-air forecasts at XMR, TBW, and EDW reveals only subtle changes in their characteristics between 1996 and 1997 (annually stratified results not shown). This is not surprising since the Eta Model changes implemented in February and August 1997 (Table 2) were designed primarily to improve deficiencies in forecasts for surface and boundary layer variables (BE97, EMC 1997). For these reasons, all upper-air data collected during 1996 and 1997 are pooled into their respective warm and cool season periods to develop generalized profiles of Meso Eta error characteristics at XMR, TBW, and EDW.

The model's systematic error growth during the forecast cycle is minimal at all three locations (demonstrated in section 6d). Hence, for most operational forecast interests, all the data may be combined into a single dataset regardless of duration. Moreover, since the systematic error growth is minimal, the error characteristics outlined below apply, on average, at any time during the forecast period. This generality does not apply to the surface data where the error characteristics varied with time of day, and does not necessarily apply at locations other than XMR, TBW, or EDW.

a. Temperature

Warm season temperature biases at EDW are less than ±1°C (Fig. 5a). At XMR and TBW, forecast temperatures below 700 mb are about 1°C colder than observed whereas above 700 mb they are about 1°–2°C warmer than observed. The net effect for warm season forecasts at the Florida stations is a tendency toward a thermally stable model atmosphere.

Warm season rms errors range from about 1° to 2.5°C and are largest in the upper troposphere (Fig. 5b). The corresponding error standard deviations of 1°–2°C (Fig. 5c) reveals that nonsystematic errors compose a substantial portion of the total error. For comparison, the typical uncertainty in rawinsonde temperature observations is about 0.6°C (Hoehne 1980; Ahnert 1991). This fact suggests that about half the nonsystematic error component could include contributions from measurement uncertainty.

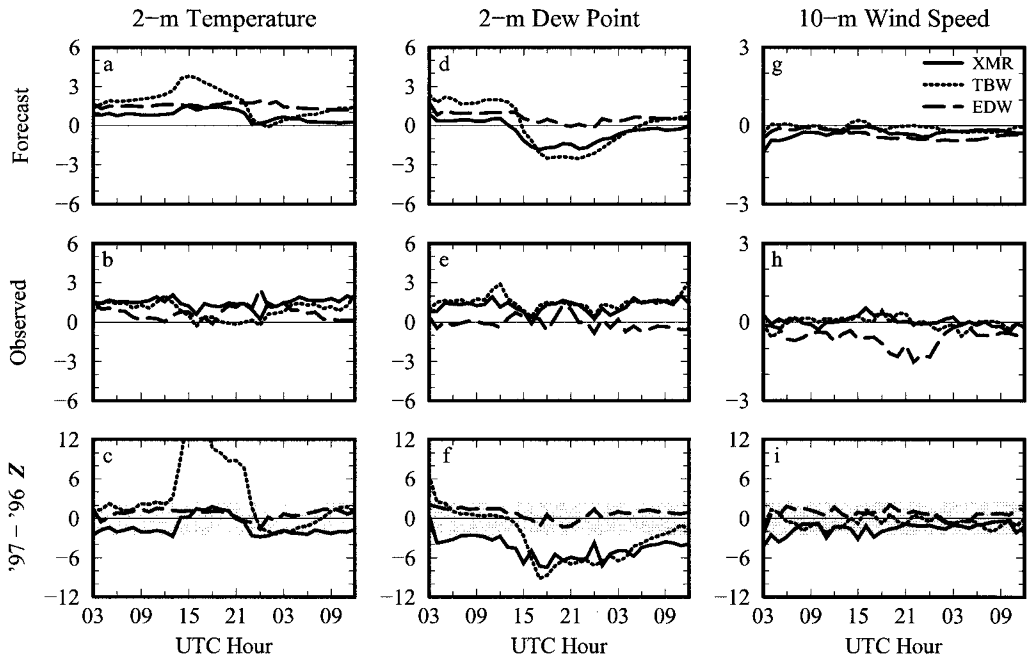

FIG. 4. Same as Fig. 3 but for the cool season.

During the cool season, temperature forecasts at EDW exhibit a negative (cold) bias below 700 mb that exceeds −4°C near the surface (Fig. 5d). At XMR and TBW, temperature biases are less than 1°C except around the 700-mb level and above the tropopause. Examination of individual forecast and observed soundings at XMR throughout the cool season (not shown) reveals that the 700-mb cold bias appears primarily because model forecasts of the lower-tropospheric inversion height are frequently at a higher level than where they are actually observed.

In the lower and middle troposphere, rms errors for cool season temperature forecasts at EDW are larger those at XMR and TBW (Fig. 5e). Since biases are small above 700 mb at EDW, the relatively large error standard deviations suggest that a greater portion of the total rms error is caused by a large amount of day-to-day variability in the forecast errors (Fig. 5f). This cool-season result for EDW could be more representative of Meso Eta Model error characteristics in midlatitudes where the day-to-day error variability likely is greater than at lower-latitude stations such as XMR and TBW.

b. Mixing ratio

Warm season mixing ratio biases at XMR and TBW (Fig. 6a) indicate that Meso Eta forecasts are on average about 1 g kg^{-1} too dry below 700 mb. Conversely, mixing ratio biases at EDW are about 0.5 g kg^{-1} greater than observed. Between 700 and 500 mb, forecasts at all three locations indicate a negative (dry) bias while above 500 mb they tend to retain excessive amounts of moisture. In combination with the negative lower-tropospheric temperature biases discussed previously, these results suggest that warm season model forecasts at XMR and TBW are typically more stable than observed. Cool season mixing ratio biases at all three locations reveal excessive moisture near the surface with a rapid vertical transition to a layer with less moisture than observed (Fig. 6d).

Rms errors for the warm season (Fig. 6b) drop from around 2.5 g kg^{-1} at low levels (1.5 g kg^{-1} at EDW) to near zero at 200 mb, where there is very little water vapor present in the atmosphere. In the cool season, rms errors follow a similar profile at all three stations starting with values of 2 g kg^{-1} near the surface (Fig. 6e). Since the error standard deviations shown in Figs. 6c and 6f are more than double the magnitude of the mixing ratio biases, nonsystematic errors account for roughly 50%–75% of the total rms error. Results shown in Figs. 6b and 6e are consistent with those of Rogers et. al (1996), who show 24-h rms errors in specific humidity from 48-km Eta Model forecasts across the United States during September 1994 ranging from nearly 2 g kg^{-1} at 1000 mb to less than 0.1 g kg^{-1} at 250 mb (see their Fig. 7). Note that these calculations for mixing ratio errors are not normalized by magnitude and are therefore not representative of percent errors as the mixing ratio tends toward zero in the upper troposphere.

c. Wind speed

Warm season wind speed biases are generally less than ±1 m s^{-1} (Fig. 7a). The exception occurs at EDW

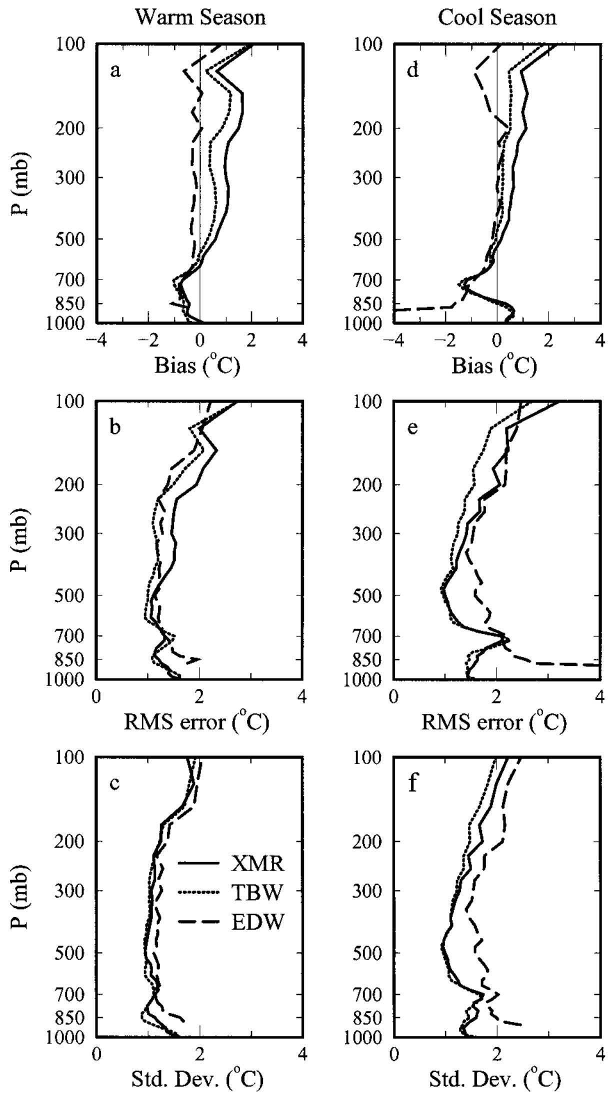

FIG. 5. Bias, rms error, and error standard deviation (°C) of Meso Eta temperature forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column [(a)–(c)] while errors for the cool season are shown in the right column [(d)–(f)].

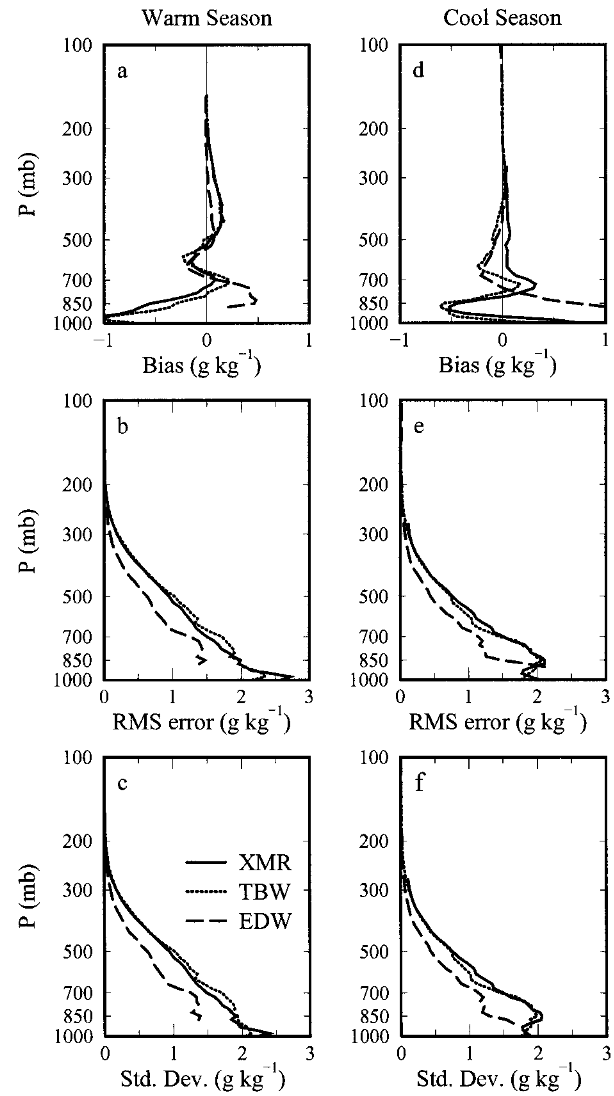

FIG. 6. Same as Fig. 5 but for mixing ratio (g kg^{-1}).

where lower-tropospheric wind speed forecasts are about 2 m s^{-1} slower than observed. This result is consistent with the negative (slow) bias in 10-m wind speed forecasts identified at EDW (Fig. 1g). Below 400 mb, warm season rms errors range from about 2 to 4 m s^{-1}. (Fig. 7b). Rms errors around the 200-mb level are larger with values approaching 6 m s^{-1}. Since forecast biases are small and uncertainties in rawinsonde wind speed measurements are about 3.1 m s^{-1} (Hoehne 1980), much of the total rms wind speed error at lower levels includes contributions from nonsystematic sources such as observational uncertainty.

During the cool season, forecast wind speeds at XMR and TBW are about 1 m s^{-1} slower (faster) than observed in the middle (upper) troposphere (Fig. 7d). At EDW, wind speed biases range from 1 to 3 m s^{-1} except near the surface where forecast wind speeds remain slow. Cool season rms errors at XMR and TBW are comparable to those found during the warm season and again, receive large contributions from nonsystematic error components including observational measurement uncertainties (Fig. 7e). At EDW, cool season rms errors above 700 mb are nearly double those of the warm season with increased contributions from both systematic and nonsystematic errors.

d. Forecast error growth

Since rawinsonde observations are available only twice daily under normal circumstances, it is not possible to observe the temporal evolution of upper-level forecast errors on an hourly basis throughout the forecast cycle.[2] However, separate examination of seasonal forecast errors at three 12-h intervals (not shown) reveals that upper-level errors do fluctuate slightly with forecast duration although their vertical profiles remain qualitatively similar. Unlike the surface error characteristics, diurnal oscillations are not evident in the upper-air forecast errors above the lowest few levels. A paired Z statistic is applied to determine if seasonal mean changes in upper-level model biases during a 24-h period represent a statistically significant systematic error growth (see the appendix). Although a few exceptions are noted in Fig. 8, the mean 24-h systematic error growth for temperature, mixing ratio, and wind speed generally is not statistically significant at the 99% confidence level. Hence, for the purpose of establishing error characteristics for operational use at XMR, TBW, and EDW, it is justifiable to blend all available data together regardless of forecast duration.

7. Summary

From May 1996 through January 1998, the AMU conducted warm and cool season evaluations of Meso Eta surface and upper-air point forecast accuracy at XMR, TBW, and EDW. These three locations were selected because they are important for 45WS, NWS MLB, and SMG operational concerns. Each warm and cool season verification period extends from May through August and October through January, respectively. By extending the evaluation for a second consecutive year, it was possible to identify statistically significant changes in systematic error that developed in response to the February and August 1997 model updates (BL97; EMC 1997). The twin-season comparison of forecast accuracy is helpful for model users by highlighting the model's characteristic strengths and weaknesses before and after the incorporation of model updates. Such results are also helpful for model development efforts and emphasize the need for ongoing analysis of model errors at specific locations.

a. Surface results

The surface error statistics presented in Figs. 1 and 2 vary widely by location, season, and time of day. The results are utilized most effectively by considering the model biases for each parameter separately and making the appropriate adjustments to the forecast guidance. For example, the fact that the Meso Eta 10-m wind speed forecasts are too fast on average at XMR (Figs. 1g, 2g) suggests that forecasts could be improved by adjusting such guidance to lower speeds. Similar adjustments should be made by local forecasters to accommodate the biases identified for other parameters.

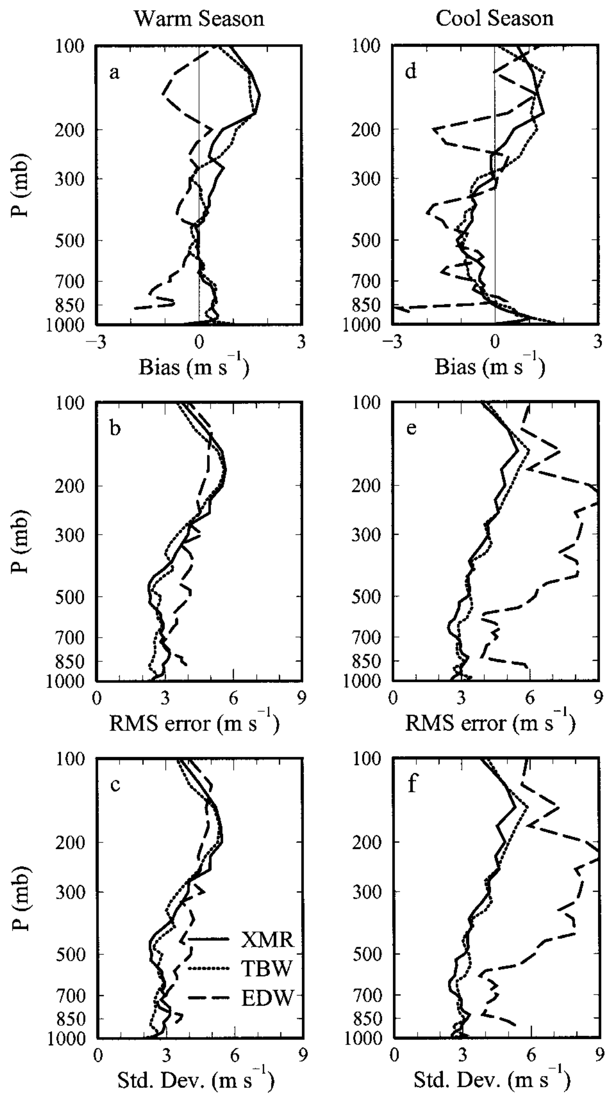

FIG. 7. Same as Fig. 5 but for wind speed (m s^{-1}).

The random error component indicates that there is substantial day-to-day variability in forecast accuracy. For many parameters, the random errors are larger than the corresponding biases, or systematic model errors. The random errors prevent perfect forecast guidance and are caused by a combination of measurement uncertainty and the model's inability to resolve localized phenomena such as wind gusts, temperature gradients, or the effects of thunderstorms. While it is possible to partially adjust for model biases, it is much more difficult to accommodate the variability in forecast errors on any given day. It might help users to compare the latest forecast guidance with current observations and make appropriate adjustments in real time.

Results shown in section 5 indicate that changes to

[2] The 50-MHz wind profiler data at KSC/CCAS are available every 5 min but are not used for the objective portion of this study because similar data are not available at TBW or EDW.

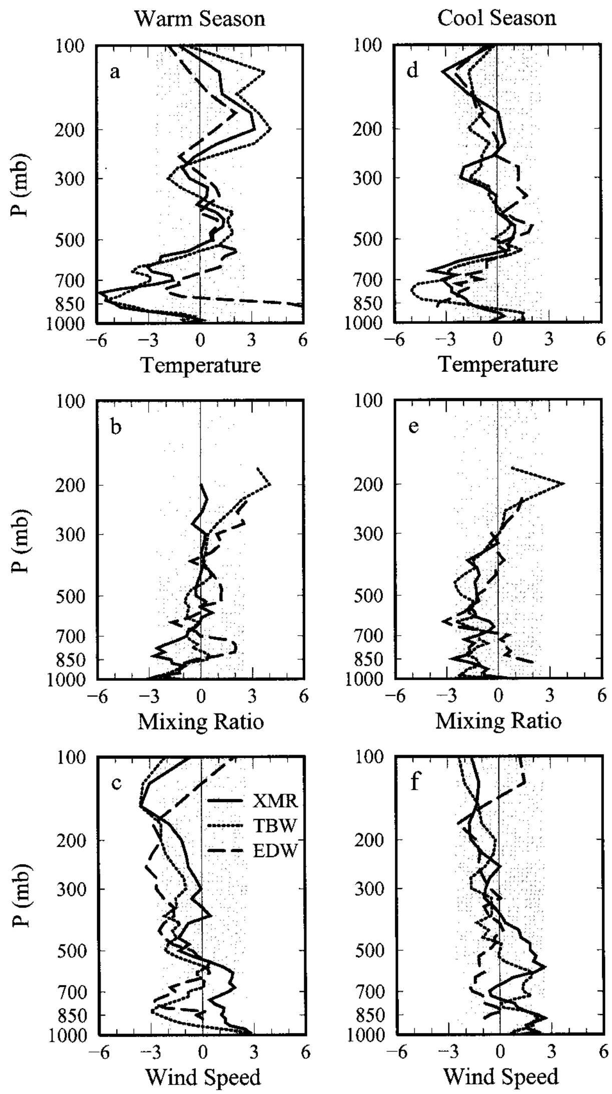

FIG. 8. Paired Z statistic plotted as a function of pressure level for Meso Eta forecast errors at XMR (solid), TBW (dotted), and EDW (dashed). The nondimensional statistic is shown for temperature (a), (d), mixing ratio (b), (e), and wind speed (c), (f). Warm season values are shown in the left column [(a)–(c)] while cool season values are shown in the right column [(d)–(f)]. Paired Z values that lie outside the shaded region indicate that 24-h systematic error growth is statistically significant at the 99% confidence level (see the appendix).

the model's physical parameterizations produced identifiable and statistically significant changes in forecast accuracy at XMR, TBW, and EDW. Some changes enhanced forecast accuracy while others actually led to greater systematic errors. It is important that model users maintain an awareness of ongoing model changes. Such changes are likely to modify the model's basic error characteristics, particularly near the surface.

b. Upper-air results

The basic error characteristics for Meso Eta forecasts of upper-level temperature, mixing ratio, and wind speed were established in section 6 for XMR, TBW, and EDW. On average, the forecast soundings at XMR and TBW during the warm season are too stable. The height of the lower-tropospheric inversion at XMR and TBW was misrepresented during the cool season. Forecast biases for wind speed are small at all three locations, but the random error component dominates the day-to-day variability. Given this variability, real-time assessment of forecast accuracy is necessary on any given day to help users determine if the model forecasts are consistent with current observations.

The statistics for the upper-air parameters did not reveal annual changes in forecast error that could be attributed solely to the February and August 1997 model updates. Moreover, the model's systematic error growth during the forecast cycle is minimal for XMR, TBW, and EDW (section 6d). Since the error characteristics were similar at all verification times during both 1996 and 1997, it was reasonable to combine all forecast–observation pairs into a single dataset for verification. Hence, on average, the error characteristics outlined in section 6 apply throughout the forecast period for XMR, TBW, and EDW.

8. Discussion

The detailed statistical results presented in sections 4, 5, and 6 are specific to Meso Eta surface and upper-air point forecasts at XMR, TBW, and EDW. The basic error characteristics vary by station, and may not be representative of errors at other geographic locations. For example, in a preliminary investigation of temperature errors, Colby (1998) demonstrated that the maximum temperature biases for Meso Eta forecasts during 1996 occurred over the central United States while the minimum biases were found over the southeastern United States and surrounding waters. Most generally, the results presented here for XMR, TBW, and EDW demonstrate the value of conducting ongoing station verification efforts.

It is important that forecasters maintain an ongoing awareness of model updates and the effects that such changes will have on point forecast accuracy within their area of responsibility. The ongoing model updates are well tested and designed to improve forecast accuracy (e.g., BL97). Indeed, Colby's (1998) study confirmed that the February 1997 Meso Eta Model updates produced a dramatic reduction of the lower-tropospheric temperature biases when averaged across the United States. However, the results shown here demonstrate that the planned changes do not always yield the expected improvements at every location.

In recent years, information documenting model updates has been made available regularly on the Internet. Much of the information needed for writing this paper and maintaining an understanding of the model changes was obtained from a list of frequently asked questions (FAQ) written on the Internet expressly for this purpose

(EMC 1997). As forecasters discover localized model deficiencies through ongoing real-time statistical verification strategies, results should be documented regularly and shared with model developers. As expressed by Manning and Davis (1997), ''These statistics would provide additional information to model users and alert model developers to those research areas that need more attention.'' The additional and complementary need for subjective verification strategies in mesoscale models is discussed in the companion paper (Manobianco and Nutter 1999).

Acknowledgments. The authors thank Dr. F. Merceret (NASA KSC), Mr. R. Lafosse (SMG), Mr. T. Oram (SMG), Mr. William P. Roeder (45WS), and Mr. John Weems (45WS) for their comments and suggestions on an earlier version of the manuscript. The constructive comments from two anonymous reviewers substantially improved the presentation.

APPENDIX

Statistical Measures

The statistical measures used here to quantify model forecast errors are the bias, rms error, and standard deviation. If Φ represents any of the parameters under consideration for a given time and vertical level, then forecast error is defined as $\Phi' = \Phi_f - \Phi_o$, where the subscripts f and o denote forecast and observed quantities, respectively. Given N valid pairs of forecasts and observations, the bias is computed as

$$\overline{\Phi'} = \frac{1}{N}\sum_{i=1}^{N} \Phi'_i, \quad \text{(A.1)}$$

the rms error is computed as

$$\text{rmse} = [\text{mse}]^{1/2} = \left[\frac{1}{N}\sum_{i=1}^{N} (\Phi'_i)^2\right]^{1/2}, \quad \text{and} \quad \text{(A.2)}$$

the standard deviation of the errors is computed as

$$\sigma' = \left[\frac{1}{N}\sum_{i=1}^{N} (\Phi'_i - \overline{\Phi'})^2\right]^{1/2}. \quad \text{(A.3)}$$

In Eq. (A.3), N is used rather than $N - 1$ so that a decomposition following Murphy [1988, Eq. (9)] could be applied to the mse:

$$\text{mse} = \overline{\Phi'}^2 + \sigma'^2. \quad \text{(A.4)}$$

Therefore, the total model error consists of contributions from model biases ($\overline{\Phi'}^2$) and random variations in the forecast and/or observed data (σ'^2). Note that if the model bias or systematic error is small, most of the mse is due to random, nonsystematic type variability in the errors. Murphy's (1988) decomposition of the mse considered individually the error contributions from the model bias and from the sample variances and covariance of the forecasts and observations. Here, Eq. (A.4) represents an algebraic simplification of that decomposition and quantifies the portion of the mse that is due to the bias and the variance of the forecast *errors*.

Tests are applied to the surface data in order to determine if model updates led to statistically significant changes in mean forecast error between the 1996 and 1997 warm and cool season periods. Following the central limit theorem as described in most statistical texts, it is assumed that the sampling distribution for the difference in mean forecast error between 1996 and 1997 is approximately normal. Sample sizes of $O(100)$ for each season enable use of the standardized Z statistic, where

$$Z = \frac{\overline{\Phi'_{97}} - \overline{\Phi'_{96}}}{\{\delta_{96}[(\sigma'_{96})^2/N_{96}] + \delta_{97}[(\sigma'_{97})^2/N_{97}]\}^{1/2}} \quad \text{(A.5)}$$

the variance inflation factor, $\delta = (1 + \rho)/(1 - \rho)$, and ρ is the lag-1 day autocorrelation for each seasonal time series of data. The variance inflation factor helps prevent the overestimation of Z by adjusting the variance of the sampling distribution to account for the influence of serial dependence, or day-to-day persistence, within the seasonal time series average (Wilks 1995). A two-tailed comparison of Z to the normal distribution using a 99% confidence level has critical values of ± 2.58 (Walpole and Meyers 1989). Calculated values of Z that lie outside this critical range indicate that the data are able to support a statistically significant difference between the 1996 and 1997 seasonal mean forecast errors.

The statistical significance of upper-level systematic error growth from early to later stages of the forecast cycle is determined using a paired Z statistic. The paired Z statistic normalizes the seasonally averaged difference in forecast error between two times during the ith cycle by the associated sample standard deviation. The covariance between errors in the early and later stages of the forecast is included because the parameters from the ith cycle are not independent and do not necessarily have equal variances (Walpole and Meyers 1989). Here, the paired Z statistic is denoted by Z', where

$$Z' = \frac{\sum_{i=1}^{N} (\Phi'_{2i} - \Phi'_{1i})}{\{N[(\sigma'_1)^2 + (\sigma'_2)^2 - (\sigma'_{12})^2]\}^{1/2}}. \quad \text{(A.6)}$$

The subscripts 1 and 2 denote variables from the ith forecast cycle verifying at 6–9 h and 30–33 h, respectively. The times used for verification are separated by 24 h and are taken at forecast durations that vary slightly according to balloon release times. Other notations are as above except that $(\sigma'_{12})^2$ denotes the sample covariance. Again using a 99% confidence level, values of Z' that lie outside the critical values of ± 2.58 indicate that the data are able to support a statistically significant 24-h systematic error growth in the upper-air forecasts.

REFERENCES

Ahnert, P. R., 1991: Precision and compatibility of National Weather Service upper-air measurements. Preprints, *Seventh Symp. on Meteorological Observations and Instrumentation,* New Orleans, LA, Amer. Meteor. Soc., 221–226.

Baldwin, M. E., and S. D. Hrebenach, 1998: Experiments with bias-corrected temperature guidance using NCEP's mesoscale Eta model. Preprints, *16th Conf. on Weather Analysis and Forecasting,* Phoenix, AZ, Amer. Meteor. Soc., 388–389.

Betts, A. K., F. Chen, K. E. Mitchell, and Z. I. Janjic, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta Model using FIFE data. *Mon. Wea. Rev.,* **125,** 2896–2916.

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

——, and Coauthors, 1997: Changes to the Eta forecast systems. NWS Tech. Proc. Bull. 441, National Oceanic and Atmospheric Administration/National Weather Service, 16 pp. [Available from National Weather Service, Office of Meteorology, 1325 East–West Highway, Silver Spring, MD 20910.]

Boyd, B. F., W. P. Roeder, J. B. Lorens, D. S. Hazen, and J. W. Weems, 1995: Weather support to pre-launch operations at the Eastern Range and Kennedy Space Center. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 135–140.

Brody, F. C., R. A. Lafosse, D. G. Bellue, and T. D. Oram, 1997: Operations of the National Weather Service Spaceflight Meteorology Group. *Wea. Forecasting,* **12,** 526–544.

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Duan, and A. Betts, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE results. *J. Geophys. Res.,* **101,** 7251–7268.

Colby, F. P., Jr., 1998: A preliminary investigation of temperature errors in operational forecasting models. *Wea. Forecasting,* **13,** 187–205.

Cortinas, J. V., Jr., and D. J. Stensrud, 1995: The importance of understanding mesoscale model parameterization schemes for weather forecasting. *Wea. Forecasting,* **10,** 716–740.

EMC, cited 1997: Environmental Modeling Center: Mesoscale Modeling Branch FAQ. [Available online from http://nic.fb4.noaa.gov:8000/research/FAQ-eta.html.]

Ernst, J. A., and F. J. Merceret, 1995: The Applied Meteorology Unit: A tri-agency applications development facility supporting the space shuttle. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 266–269.

Friday, E. W., Jr., 1994: The modernization and associated restructuring of the National Weather Service: An overview. *Bull. Amer. Meteor. Soc.,* **75,** 43–52.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Hazen, D. S., W. P. Roeder, B. F. Boyd, J. B. Lorens, and T. L. Wilde, 1995: Weather impact on launch operations at the Eastern Range and Kennedy Space Center. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 270–273.

Hoehne, W. E., 1980: Precision of National Weather Service upper air measurements. NOAA Tech. Memo NWS T&ED-16, National Oceanic and Atmospheric Administration/National Weather Service, 23 pp. [Available from National Weather Service, Office of Meteorology, 1325 East–West Highway, Silver Spring, MD 20910.]

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting,* **10,** 689–707.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

——, 1996a: The Mellor–Yamada level 2.5 turbulence closure scheme in the NCEP Eta model. Research Activities in Atmospheric and Oceanic Modeling, Geneva, Switzerland, CAS/JSC WGNE, WMO, 4.14–4.15. [Available from World Meteorological Organization, Case Postale 2300, Geneva CH-1211, Switzerland.]

——, 1996b: The surface layer parameterization in the NCEP Eta Model. Research Activities in Atmospheric and Oceanic Modeling, WMO, Geneva, CAS/JSC WGNE, 4.16–4.17. [Available from World Meteorological Organization, Case Postale 2300, Geneva CH-1211, Switzerland.]

——, 1996c: The surface layer in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 354–355.

Manning, K. W., and C. A. Davis, 1997: Verification and sensitivity experiments for the WISP95 MM5 forecasts. *Wea. Forecasting,* **12,** 719–735.

Manobianco, J. M., and P. A. Nutter, 1999: Evaluation of the 29-km Eta Model. Part II: Subjective verification over Florida. *Wea. Forecasting,* **14,** 18–37.

McPherson, R. D., 1994: The National Centers for Environmental Prediction: Operational climate, ocean, and weather prediction for the 21st century. *Bull. Amer. Meteor. Soc.,* **75,** 363–373.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.,* **77,** 2637–2649.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.,* **116,** 2417–2424.

Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational ''early'' Eta Model: Original 80-km configuration and recent changes. *Wea. Forecasting,* **10,** 810–825.

——, T. L. Black, D. G. Deaven, and G. J. DiMego, 1996: Changes to the operational ''early'' Eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting,* **11,** 391–413.

Stensrud, D. J., and J. A. Skindlov, 1996: Gridpoint predictions of high temperature from a mesoscale model. *Wea. Forecasting,* **11,** 103–110.

Walpole, R. E., and R. H. Meyers, 1989: *Probability and Statistics for Engineers and Scientists.* Macmillan, 765 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Zhao, Q., T. L. Black, and M. E. Baldwin, 1997: Implementation of the cloud prediction scheme in the Eta Model at NCEP. *Wea. Forecasting,* **12,** 697–712.