

WESTERN REGION TECHNICAL ATTACHMENT

NO. 98-39

NOVEMBER 10, 1998

WESTERN REGION MODEL DIAGNOSTICS

Kirby Cook - WRH SSD, Salt Lake City, UT

Introduction

Every forecast produced by the National Weather Service is a blend of Numerical Weather Prediction (NWP) and a forecaster's skill and experience in interpreting the current state and evolution of the atmosphere. Quite often, an operational forecaster places the most faith in a particular model due to a subjective judgment in its recent performance. In order to provide the forecaster with more useful guidance on model performance in addition to furthering understanding of NWP, operational forecasts from the Eta, AVN, and MRF models are evaluated in a diagnostic mode daily. Forecasts from each model are compared to current upper-air and surface observations from sites in the Western United States. Measures of model skill are calculated for single forecasts as well a trend in skill over a seven-day period. These skill measures are plotted and placed on the Western Region webpage (<http://www.wrh.noaa.gov>) for use by operational forecasters on a real-time basis.

Data and Methods

Operational forecasts from the 32-km Eta, AVN, and MRF models are obtained several times daily (once in the case of the MRF) in GRIB format via the OSO server. Upon receipt, these grids are processed into General Meteorological Package (GEMPAK) gridded files. The 32-km Eta and AVN are both obtained on 80-km operational grids (grid 211) and the MRF is received and processed on a 381-km grid (grid 201). More detailed descriptions of these models are provided by Mittelstadt (1998) and Nelson (1998). Forecasts are compared against observations, and model bias error (BE) is calculated as defined by the equation:

$$BE = \frac{1}{N} \sum_{n=1}^N (x^f - x^o)$$

where N is the total number of forecasts and the superscripts f and o denote forecast and observed values, respectively. A positive bias error indicates a tendency to overforecast a particular variable, and conversely, a negative value indicates a tendency to underforecast that variable.

In the case of the upper-air diagnostics, model forecasts of temperature ($^{\circ}$ C), dew point ($^{\circ}$ C), relative humidity (%), wind (ms^{-1}) and geopotential height (m) on all available levels are bilinearly interpolated to 21 upper-air sites (UIL, OTX, SLE, MFR, BOI, TFX, GGW, UNR, RIW, SLC, GJT, LKN, REV, OAK, VBG, MYF, DRA, FGZ, TUS, ABQ, EPZ) in the western United States. Before interpolation, model wind forecasts are broken down into North-relative u and v components. The interpolated model forecasts are then written to GEMPAK upper-air files. Once the forecasts have been translated to each site, comparisons are made on each level, between observations and all valid forecasts. Bias errors are calculated at each site for all forecast hours as well as seven-day period averages and plotted against pressure in the vertical for each upper-air site.

Model forecasts from the Eta and AVN are also interpolated to 36 surface sites (TUS, PHX, FLG, MYF, LAX, SAC, SFO, EKA, DRA, RNO, EKO, CDC, SLC, PIH, BOI, MFR, SLE, PDX, PDT, BIL, GGW, HLN, MSO, GTF, GEG, SEA, UIL, DEN, GJT, PUB, BIS, RAP, CYC, RIW, ABQ AND LBF) within or surrounding Western Region and compared to observations at 0000, 0600, 1200, and 1800 UTC daily. Eta 2 m temperatures ($^{\circ}$ C), 2 m dew points ($^{\circ}$ C) and 10 m winds (ms^{-1}) as well as AVN 2 m temperatures ($^{\circ}$ C) and 10 m winds (ms^{-1}) are bilinearly interpolated to each surface observation site and stored in GEMPAK surface files for comparison to observations. Again, prior to interpolation, model winds are broken down into North-relative u and v components. There is no vertical interpolation or adjustment performed on the forecast data to account for differences in model terrain and actual terrain. Bias errors are calculated for each parameter at each site and forecast hour, as well as a domain average. The model biases for each parameter (temperature, dew point and wind) are plotted on a surface map for each parameter at the 0000 and 1200 UTC validation times. Wind barbs are generated from the resultant winds, which are derived from the component wind vectors, rather than an actual direction and speed. Additional meteogram plots of forecast versus observation are also produced at 6 hourly intervals for each surface site.

Upper-air Diagnostics

Model bias errors generated from forecast comparisons against upper-air observations are broken down into single-forecast diagnostics and period (seven-day) diagnostics. These upper-air measures may be used to evaluate forecast skill both temporally and spatially. Single-forecast diagnostics represent a "snapshot" of model skill with respect to the latest observations, as shown in Fig.1. In this case, the Eta 12-h, 24-h and 36-h temperature forecasts valid at Great Falls, MT. on 12 UTC, October 29, are too cold (by as much as 3° C) between 300 and 450 hPa, as well as below 700 hPa. On the other hand, the 48-h temperature forecasts are too warm below 500 hPa. Similarly, model 12-h, 24-h, and 36-h

relative humidity (dew point) forecasts are too dry (cold) between 500 and 700 hPa, while 48-h forecasts are too wet (warm) throughout the entire air-column. The Eta is also under forecasting the magnitude of both the u and v component winds and over forecasting geopotential heights at all levels. It is interesting to note, although not surprising, that for most fields, the initial analysis (F000) has the smallest bias, while the 48-h forecast produces the largest. Using single-forecast diagnostics, a forecaster can make a quantitative judgment on the skill of recent model initializations as well as its performance at larger lead times. Because the single-forecast diagnostics represent a “snapshot” of model skill, these measures are vulnerable to varying weather regimes. By evaluating model biases at different sites and levels, an understanding of model skill is made available in varying situations. For instance, in the case described above, the forecast skill with respect to relative humidity (bias errors as large as 40%) at Great Falls represents not only the highest bias locally, but rather the lowest skill with respect to this parameter in several days, indicating the model’s inability to correctly forecast this parameter in a changing weather regime.

Similarly, the period diagnostics represent model skill with respect to both lead time and forecast parameter. However, this skill is represented as an average over a seven day interval, allowing the forecaster to examine overall model performance over the last week, with a much more generalized view on model skill. In addition, the period diagnostics can be used in conjunction with measures provided by the single-forecast diagnostics to produce a more detailed analysis of current model performance. In the case discussed above, model skill with respect to relative humidity, as evaluated against the most recent observations, is relatively poor (Fig. 1). However, model biases for the same parameter over the last seven days (Fig. 2) are not only different, but slightly better. This is, in fact, due to averaging varying biases over the period, and illustrates the “regime dependent” nature of the single-forecast diagnostics.

Surface Diagnostics

Model surface diagnostics are generated from forecast comparisons against surface observations every six hours. Like the upper-air diagnostics, bias error is used as a measure of model skill both temporally and spatially at the lowest forecasted level. Diagnostics created from Eta 2 m temperature forecasts valid on October 25, 1998 are shown in Fig. 3. Bias errors from the model initialization (red), 12-h forecast (orange), 24-h forecast (yellow), 36-h forecast (green), and 48-h forecast (blue) valid at 1200 UTC are displayed at each site as well as a domain average (Fig. 3). In this case, the average bias error at all lead times is negative, indicating a tendency for the Eta to underforecast the temperature in this instance. In addition, the average bias is 2° C colder for the initialization than for any of the other forecast hours valid at this time. This illustrates the utility of the surface diagnostics in identifying model accuracy at the initialization as well as error growth with lead time. The surface diagnostics may also be used to identify spatial biases connected to resolution issues in complex terrain. For example, the bias errors in the above case are smaller in magnitude at the lower elevation sites (LAX, SFO, SLE,

PDX, UIL and SEA) than those for sites further inland. Quite simply, the Eta model terrain is much closer to the actual terrain at sites at or near sea-level than those in the Rocky Mountains. Since no correction for the discrepancies between model terrain and actual terrain was employed, these biases were not removed from the diagnostics (nor are they removed from the actual forecasts). Like the single-forecast upper-air measures, the surface diagnostics are regime dependant, and can be used to analyze model skill in varying situations. Differences between biases at various sites can indicate model skill with respect to a changing airmass, or simply the diurnal forcing under the stability of an upper-level ridge. Additional plots of forecast versus observation are produced for each site, an example of which, is shown as a meteogram for Great falls in Fig. 4. Model forecasts of temperature are illustrated as color coded positive symbols, while observations appear as black asterisks. In this case, the tendency of the Eta model to under forecast temperature (Fig. 3) at this site is reinforced by plotting the actual forecast against the most recent observation. In addition, all available forecasts (out to 48-h) are plotted as well, allowing the forecaster to easily extrapolate biases out in time.

Summary

Model upper-air and surface diagnostics are produced on a real-time basis, providing forecasters with up-to-date measures of model skill both aloft and at the surface. These measures can be a valuable prognostic tool in addition to providing insight into current NWP that is crucial in today's changing operational environment. Future diagnostic products will include a period surface validation as well as varying displays of existing products. Additional work will continue to address the evolving needs of the forecaster, and will include new types observations and methods of validation.

References

- Nelson, J.A., 1998: The AVN/MRF Model Enhancements. WR-Technical Attachment 98-22.
- Mittelstadt, J., 1998: The Eta-32 Model. WR-Technical Attachment 98-03.

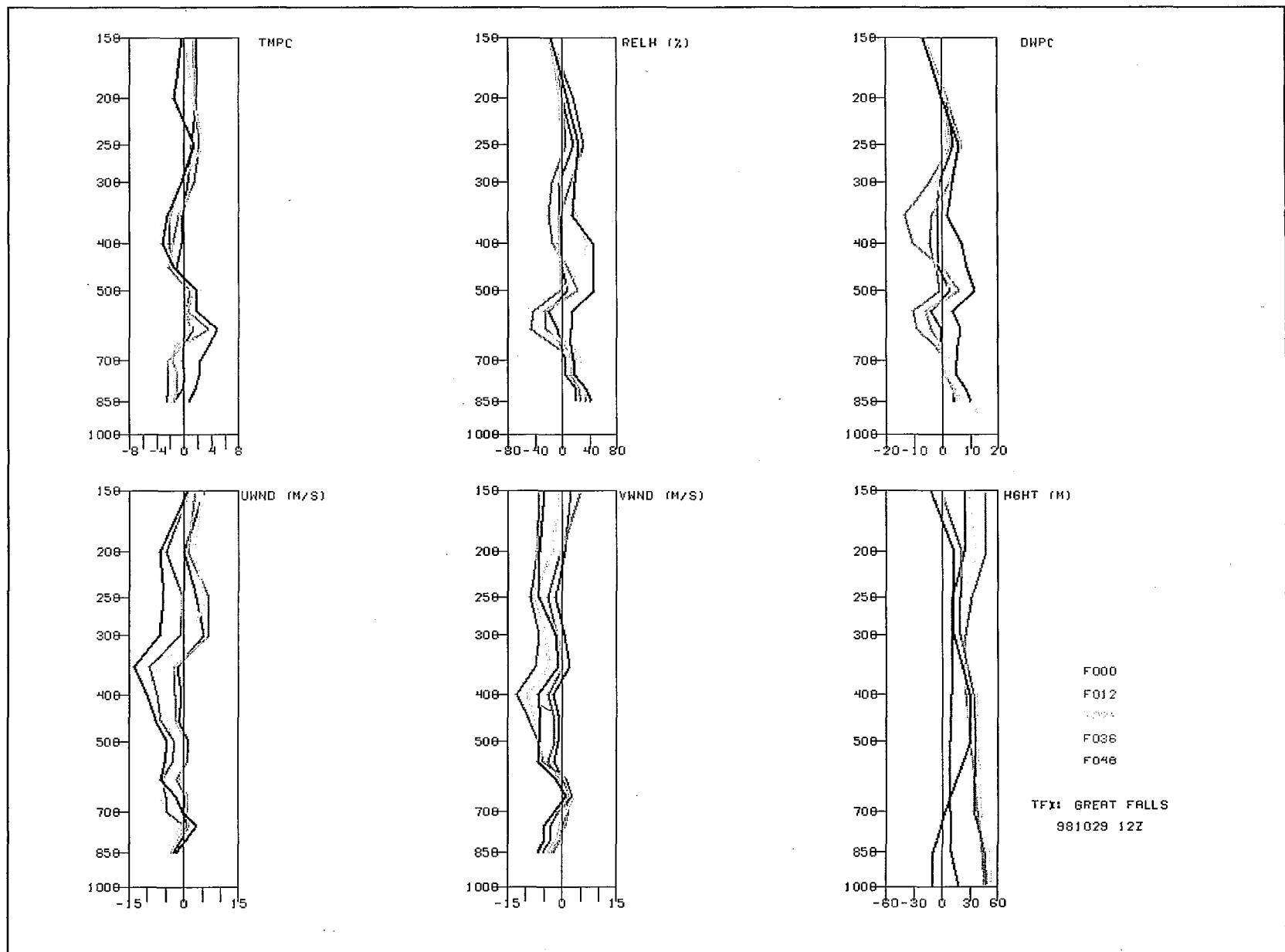


Figure 2 Example of a single-forecast upper-air diagnostic plot for Eta forecasts of temperature ($^{\circ}$ C), relative humidity (%), dew point ($^{\circ}$ C), u-component winds (ms^{-1}), v-component winds (ms^{-1}), and geopotential height (m) valid at Great Falls, MT. (TFX) on October 29th, 1998 at 1200 UTC. Model bias errors are shown for the initialization (red), 12-h forecast (orange), 24-h forecast (yellow), 36-h forecast (green) and 48-h forecast (blue).

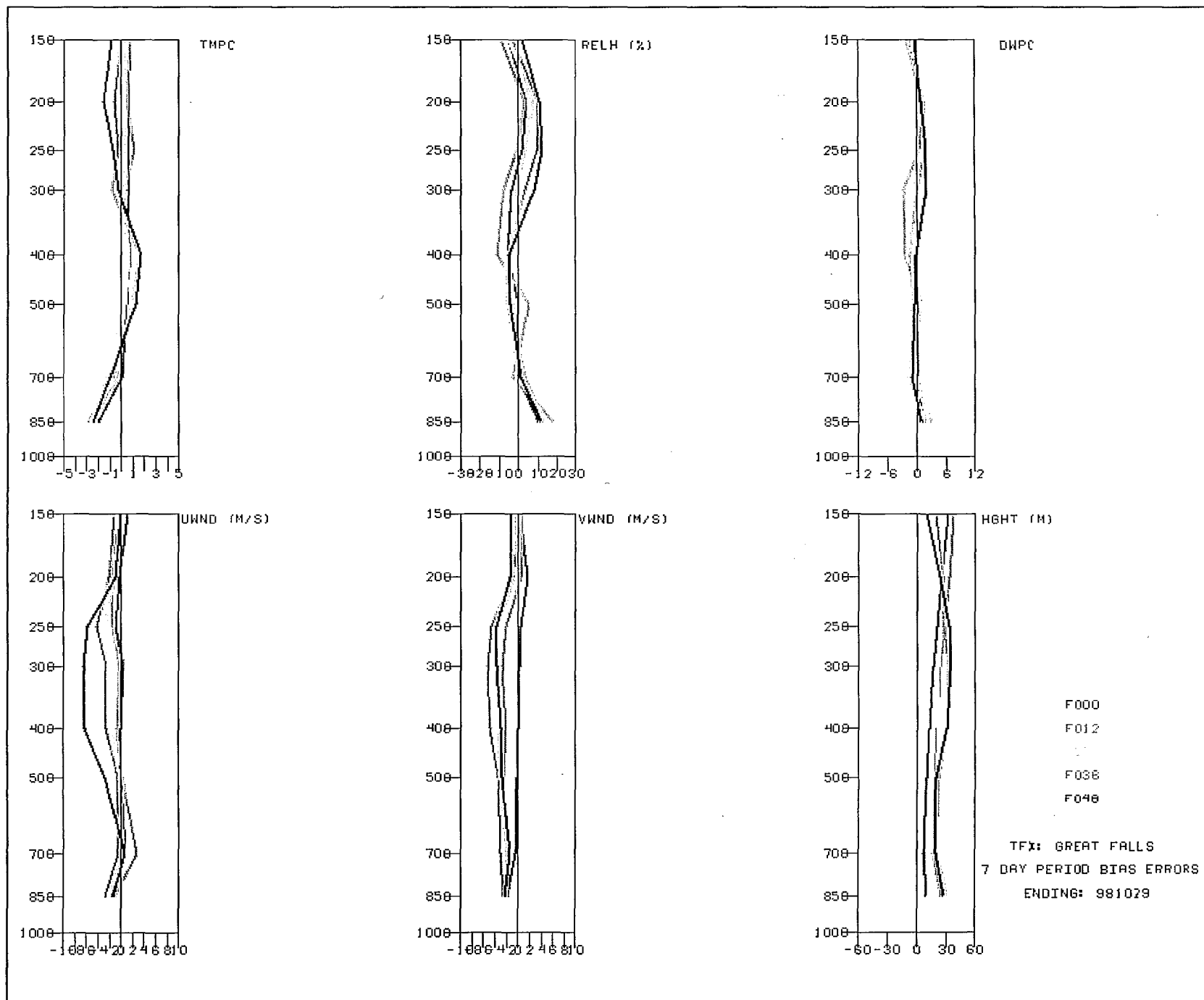


Figure 3 Example of period upper-air diagnostic plot for Eta model forecasts of temperature ($^{\circ}$ C), relative humidity (%), dew point ($^{\circ}$ C), u-component winds (m s^{-1}), v-component winds (m s^{-1}), and geopotential height (m) valid at Great Falls, MT. (TFX). Average bias errors for the seven-day period ending on October 29th, 1998 are shown for the initialization (red), 12-h forecast (orange), 24-h forecast (yellow), 36-h forecast (green), and 48-h forecast (blue).

