3.5 **Spatial Bias Errors in the Operational NCEP Eta Model**

Kimberly L. Elmore[1], Michael E. Baldwin[2] and David M. Schultz[3]

1. Introduction

Forecasters are often interested in understanding the most likely errors in numerical model guidance. While forecasters understand that blindly applying such errors is not appropriate, such knowledge provides additional insight into the numerical guidance that is often useful.

Forecasters are typically provided RMS error values for various forecast parameters such as geopotential, wind components, temperature and mixing ratio. While RMS errors might provide some rather vague guidance, they do not provide any spatial information about the errors that might be expected, they do not provide any spatial information about the variability of those errors, and they do not provide any information about whether or not the forecasts are biased. Knowledge about the spatial distribution of errors would likely be more helpful than knowledge of the simple RMS error. In addition, knowledge about the predictability, if any, of such spatial errors based on the initial pattern of some arbitrary field, 500 hPa height, for example, might help forecasters refine their expectations of the likeliest model errors and consider the ramifications to their forecast if such errors are present. Such knowledge could also be useful to model developers, to help them unravel any systematic model errors.

2. Data

The data consist of an archive of the operational National Centers for Environmental Prediction (NCEP) Eta model, maintained at the Storm Prediction Center (SPC) and the National Severe Storms Laboratory (NSSL). These data start on 26 Jan 2001 and end on 31 Mar 2002. Because the operational Eta model changed grid resolution during this period, all data have been interpolated to the Automated Weather Integration and Processing System (AWIPS) 212 grid with a 40 km horizontal resolution. Geopotential, *u,* and *v* are archived for 850, 700, 500 and 300 hPa, but at 700 and 850 hPa, mixing ratio (derived from temperature, pressure, and dewpoint), and temperature are also archived. Only that part of the grid that fully encompasses the continental United States (CONUS), a fraction of southern Canada and a fraction of northern Mexico is archived.

Data for this analysis consist of 24-h forecasts from the 1200 and 0000 UTC analyses and the verifying analyses. Because this work examines the spatial error characteristics, the gridded analysis is used as the verification data. This is different from the work done by White et al. (1999) where verification was performed at specific soundings sites within the Rocky Mountains. The present work is directed towards overall spatial error patterns over the CONUS. Most previous work dealing with Eta model verification examines either a few soundings or a limited region. Little, if any, work has appeared that examines the entire CONUS.

Using the Eta analysis for verification has some drawbacks. The most serious concern is how much the 24-h forecast affects the verifying analysis. Fortunately, the nature of the 4D Eta data assimilation system (EDAS), which includes a large amount of nonmodel data, tends to mitigate this effect. If only sounding data are used for verification, the exercise reverts to a sounding validation study with only five sounding levels including the surface, which is hardly of interest. A different model's initial conditions could have been used, but the EDAS is considered to be the most sophisticated and advanced system currently in use, so substituting data from a less advanced, less sophisticated system seems even harder to justify. Because only rather large-scale errors are examined, it is likely that using the verifying Eta analysis is at least as good as any other choice, and is likely to be the best available for a gridded verification.

3. Generating the verification statistics

The mean error value for all fields within the entire dataset and each of the four seasons is computed for each data point. The seasons are defined as follows: Winter is December through February, Spring is March through May, Summer is June through August, and Autumn is September through November. The variance and the lag-1 autocorrelation are also computed at each grid point.

A standard *t*-test is performed at each individual grid point to determine the 95% confidence interval about the mean. However, the sample size is adjusted for serial correlation of the errors. The effective sample size, n', is given approximately by

$$n' \cong n\frac{1-\rho_1}{1+\rho_1},$$

where ρ_1 is the lag-1 autocorrelation and n is the sample size (Wilks 1995).

Spatial correlation exists that is not accounted for here. The spatial correlation in the data will reduce the effective sample size further (reducing the degrees of freedom) but by an unknown amount. The effect will be to make computed confidence intervals too small, but the there seems to be no good way to account for this effect. Later, some possible ramifications of this will be discussed.

---

1. Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma/National Severe Storms Laboratory
Corresponding Author Address: 1313 Halley Circle, Norman, OK 73069, elmore@nssl.noaa.gov.

2. Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma/Storm Prediction Center/National Severe Storms Laboratory, Norman, OK.

3. Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma/National Severe Storms Laboratory, Norman, OK.

The results shown here are limited to the 500-hPa level for mean bias errors and confidence intervals over the entire data set.

4. Results

Three 500-hPa fields are considered: geopotential, *u,* and *v*. To provide some common ground on which to start, RMS and bias errors for these three fields are shown in Table 1. The bias errors are well below 1 m s^{-1}

Table 1: 500 hPa 24 h Forecast Errors

Field	1200 UTC RMS	1200 UTC Bias	0000 UTC RMS	0000 UTC Bias
Height	16.1 m	-4.4 m	15.5 m	-5.1 m
u	3.4 m s^{-1}	0.05 m s^{-1}	3.2 m s^{-1}	0.04 m s^{-1}
v	3.4 m s^{-1}	-0.10 m s^{-1}	3.3 m s^{-1}	-0.05 m s^{-1}

for the wind components and below 10 m for the height, which are quite respectable. Large RMS errors indicate occasional large bias errors scattered about within these fields. The simplicity of reducing the verification error statistics to single statistics is certainly attractive, but they provide no insight as to whether or not the errors have any spatial structure because the inherent dimensionality of gridded forecasts has been discarded.

Preserving the dimensionality allows for some more insight into the nature of forecast errors, but incurs a significant cost due to the increased complexity or dimensionality. If the mean bias errors at each gridpoint are preserved, then any mean spatial structure becomes evident. In addition, if an estimate of the variance is included, then confidence intervals can be placed about the bias error estimate. Confidence intervals may help establish if the bias errors are statistically significant and, since they are essentially scaled variance, show where the forecast errors are most and least reliable. However, statistical significance does not equal physical significance. For example, height errors of 2 m may be significant at the 95% confidence level, but may have no physical significance.

The first example is for the 1200 UTC 24-h forecast for 500-hPa height (Fig. 1). Compared to the analysis valid at the forecast time, 500-hPa heights are high by 55 m over the UT-AZ border and low by about 14 m over north-central GA. Based on the confidence interval thickness, bias errors greater than about 0.3 m are statistically significant at the 95% level. Note that even if the estimate of variance is low by a factor of four, making the confidence intervals twice as large, bias errors greater than about 1 m are still significant at the 95% level.

Because the 95% confidence interval represents a scaled variance, structure in the variance is evident as variations in the confidence-interval thickness. Hence, while almost all of the bias errors are statistically significant, they are most reliable in areas where the confidence interval thickness is smallest and least reliable where it is largest. For the 1200 UTC 24-h forecast, the bias error variance tends to be smallest over the FL peninsula, along the Gulf Coast and along the East Coast north to VA. The bias-error variance is largest over the

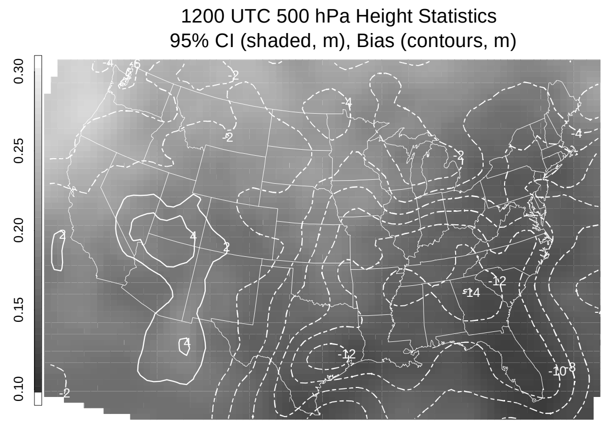

Figure 1. Error statistics for the 1200 UTC 24-h 500-hPa height forecast for the NCEP Eta model. The shading indicates the thickness, in m, of the 95% confidence interval (CI) for the mean bias error. The shaded values are interpreted based on the gray scale on the left. The white contours show the mean 500-hPa height errors. Positive errors (solid) are for forecast heights greater that observed, and negative errors (dashed) are for forecast heights less than observed.

northwest coast. Thus, the expected forecast error is for 500-hPa heights to be low over southeast CONUS, and most reliably low over the southeast coast from the southern tip of TX up through VA. Forecast 500-hPa heights are expected to be slightly high over the southwestern states. The expected nature of the errors is least predictable over the northwest coast.

The 0000 UTC 24-h Eta forecast errors are significantly different (Fig. 2). All of the bias errors are nega-

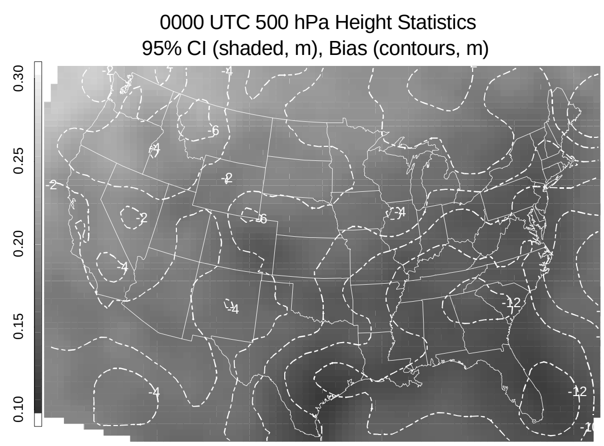

Figure 2. Same as Fig. 1, but for the 0000 UTC 24-h 500-hPa height forecast.

tive, but none are as large as for the 1200-UTC forecast. The most negative errors, -13 m, are still over the southeast. The least negative errors, -2 m, are also still over the southwestern states. The error variance is also generally smaller, but is particularly reduced over OK and eastern CO.

Wind errors are broken into *u* and *v* components. The wind errors appear to be primarily ageostrophic. For the 24 h forecast valid at 1200 UTC (Fig. 3) there are two large regions of significant bias. One region is over eastern ID/western MT, where the *u* wind component is high

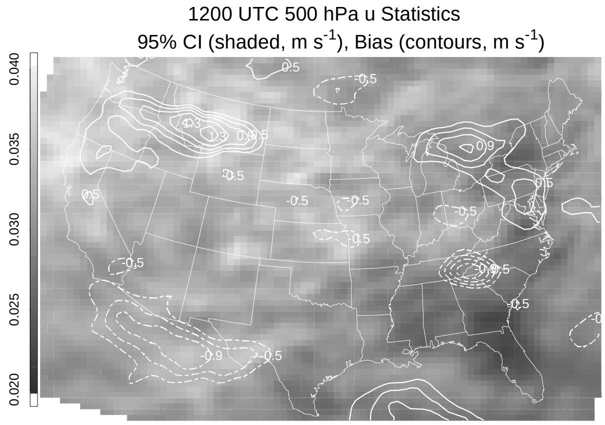

Figure 3. Same as for Fig. 1, but for the *u* wind component. Only values greater than 0.5 m s^{-1} are contoured.

by as much as 1.4 m s^{-1}. A second region is along northern Mexico, west of the Big Bend, where the *u* component is too low by about 1 m s^{-1}. The *u* component is also biased nearly 1 m s^{-1} negative over a relatively small region over western NC, near the highest part of the Appalachian mountain chain.

The height error variance displays a general large-scale trend from low values in the southeast to high values in the northwest, as for the height error variance. Overall, the 95% confidence interval for the *u* wind component is quite small, on the order of 0.05 m s^{-1}, so almost all bias errors are statistically significant, though they may not all be physically significant.

Similar patterns are noted for the 24-h forecast valid at 0000 UTC, though, as for heights, the bias errors are smaller (Fig. 4). The main regions of positive and nega-

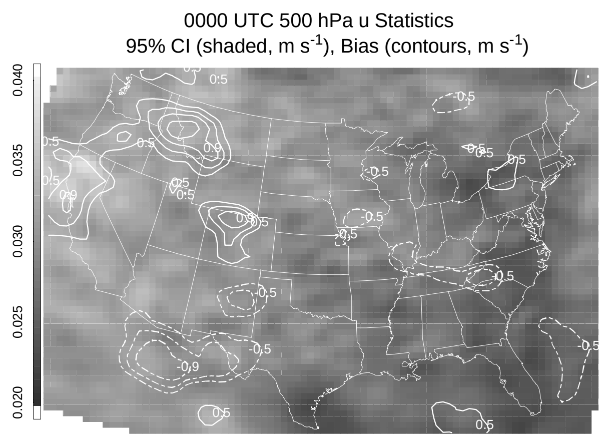

Figure 4. Same as Fig. 3, but for 24-h forecasts valid at 0000 UTC.

tive bias errors are still in evidence, as is the small region of negative bias errors over western NC. However, the region of positive bias now has an southward extension, to near the San Francisco, off of the western end, and a second area of + 1 m s^{-1} bias appears over north central CO. As for the 1200 UTC data, the error variance is smaller overall and still shows a general trend from low values in the southeast to larger values in the northwest.

The errors for the *v* component for the 24-h forecast valid at 1200 UTC and 0000 UTC (Figs. 5 and 6, respec-

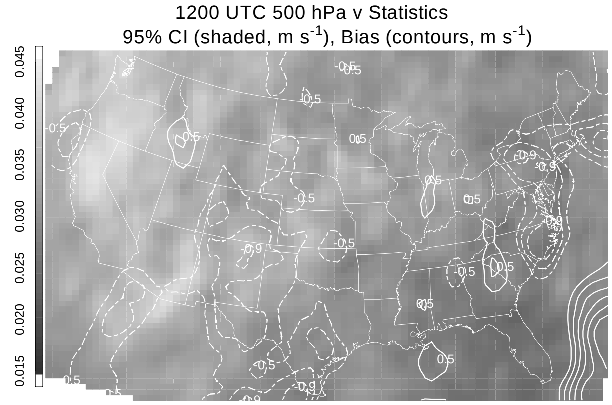

Figure 5. Same as Fig. 3, but for the *v* component.

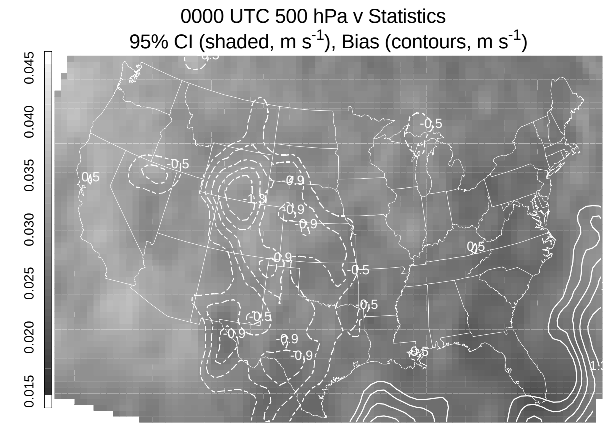

Figure 6. Same as Fig. 4, but for the *v* component.

tively). The major feature is consistent positive bias (too much southerly component) off of southeast coast at 1200 and 0000 UTC and, in addition, over the Gulf of Mexico at 0000 UTC.

For the 24-h forecast valid at 1200 UTC, there is a broad band of positive bias extending from the central TX Gulf coast into western CO. Unlike height and *u,* the bias in *v* is more pronounced for the 0000-UTC forecast than for the 1200-UTC forecast. For the 0000-UTC forecasts, the same broad region of positive *v* bias extends from the Big Bend area northward into central WY., where the maximum bias errors are -1.4 m s^{-1}. Qualitatively, these wind errors appear to be supergeostrophic when considered with the 500-hPa height errors, because the 500-hPa height error gradient is less over this region at 0000 UTC than at 1200 UTC.

The confidence interval thickness is overall slightly larger than for the *u* component, indicating that the *v* errors are slightly more variable than the *u* errors. However, the same general trend exists, with the smallest error variance of the Southeast and the largest over the Northwest. For the most part, the 500-hPa 24-h forecast *v* component is biased low over the CONUS.

5. Conclusions and future work

While the current presentation contains compelling statistics, it is difficult to deduce the physical processes that might explain the observed errors without considering other levels. Also, an overall mean error pattern at any single level has limited utility if the seasonal error patterns differ significantly. However, the ability to deduce the physical nature of at least some of the bias errors is clearly present and should prove beneficial to both forecasters and model developers.

In the future, data will be partitioned based on cluster analysis. Starting with all of the gridded initial conditions, a field (such as 500-hPa geopotential) is chosen on which to generate clusters. Hence, sets of initial patterns deemed "similar" are generated. For each cluster, the mean error in the 24-h forecast is computed. These errors are then compared to the grand mean bias errors, or to any seasonal bias errors, to determine if they are significantly different, based on initial conditions. In a like manner, clusters of similar error patterns can be generated and mapped back to initial conditions.

This approach will be used to examine if the error patterns themselves have any predictability beyond the mean values. Certain initial patterns may lead to similar error patterns that can be significantly different from the mean. Such knowledge would be helpful to forecasters. Using the converse, clusters of similar particular error patterns can be examined to determine if there is any commonality in the initial conditions that generate them. Additionally, the largest error magnitudes in any given cluster (for example, all errors greater than the 75^{th} percentile) can be examined to see if there is some underlying predictability unique to the largest errors.

References

White, B. G., J. Paegle, W. J. Steenburgh, J. D. Horel, R. T Swanson, L. K. Cook, D. J. Onton, and J. G. Miles, 1999: Short-term forecast validation of six models. *Wea. and Forecasting,* **14,** 84-108.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, New York, NY, 467 pp.