Comparison of Model Forecast Skill of Sea Level Pressure along the East and West Coasts of the United States

GARRETT B. WEDAM, LYNN A. MCMURDIE, AND CLIFFORD F. MASS
Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 20 May 2008, in final form 12 November 2008)

ABSTRACT

Despite recent advances in numerical weather prediction, major errors in short-range forecasts still occur. To gain insight into the origin and nature of model forecast errors, error frequencies and magnitudes need to be documented for different models and different regions. This study examines errors in sea level pressure for four operational forecast models at observation sites along the east and west coasts of the United States for three 5-month cold seasons. Considering several metrics of forecast accuracy, the European Centre for Medium-Range Weather Forecasts (ECMWF) model outperformed the other models, while the North American Mesoscale (NAM) model was least skillful. Sea level pressure errors on the West Coast are greater than those on the East Coast. The operational switch from the Eta to the Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF-NMM) at the National Centers for Environmental Prediction (NCEP) did not improve forecasts of sea level pressure. The results also suggest that the accuracy of the Canadian Meteorological Centre’s Global Environmental Mesoscale model (CMC-GEM) improved between the first and second cold seasons, that the ECMWF experienced improvement on both coasts during the 3-yr period, and that the NCEP Global Forecast System (GFS) improved during the third cold season on the West Coast.

1. Introduction

Major investments in numerical weather prediction and observation systems have resulted in significant improvements in the forecast skill of synoptic-scale models during the past several decades. For example, Simmons and Hollingsworth (2002) compared forecasts and analyses for sea level pressure and 500-hPa heights throughout the extratropical Northern Hemisphere and found the equivalent of a 1-day increase in forecast skill for three operational models during the 1990s. Harper et al. (2007) and Kalnay et al. (1998) documented the increasing skill of the National Centers for Environmental Prediction’s (NCEP) operational models for 500-mb heights and other forecast parameters.

But even with recent improvements in numerical weather prediction, major failures of short-to-medium-range forecasts still occur regularly. McMurdie and Mass (2004) showed that NCEP’s North American Mesoscale (NAM) model experienced large 0–48-h forecast errors over the northeast Pacific and West Coast, with errors in sea level pressure greater than 10 hPa occurring 10–20 times per cold season for 48-h forecasts. They found that such large errors were generally associated with misplaced and underforecast surface low pressure systems. Large sea level pressure errors can also be produced by timing errors. For example, Colle et al. (2001) found trough passage timing errors along the Pacific Northwest coast as large as 15 h. More recently, Charles and Colle (2009) used NCEP’s Global Forecasting System (GFS) analyses to verify GFS and NAM model forecasts of cyclones over different regions of North America. For short-term (24–72 h) forecasts they found that the NAM had larger errors in central pressure than the GFS and that this skill disparity was most pronounced over the northeast Pacific.

To illustrate a recent major forecast problem over the northeast Pacific, Fig. 1 shows the errors of a 72-h NAM forecast verifying at 0000 UTC 24 December 2006. A clear inconsistency between the satellite imagery (at the verification time) and the sea level pressure forecast of a major low pressure center highlights the poor quality of the forecast. A large 27-hPa error in sea level pressure
at approximately 45°N, 148°W is associated with a misplaced center of low sea level pressure, while a 31-hPa error, located near 52°N, 125°W, was produced by a forecast center of low pressure that did not verify.

To gain insight into the origin and nature of model forecast errors, the frequencies and magnitudes of such errors needs to be documented for different models and regions. Knowing the relative quality of different models would be of immediate use to forecasters assessing varying model solutions. Information on model errors and their regional variability is an important first step toward model improvement and is valuable to international programs such as The Observing System Research and Predictability Experiment (THORPEX), which calls for a collaborative effort between operational and academic communities to accelerate the increase in skill of 1–14-day forecasts (Shapiro and Thorpe 2004; Parsons et al. 2007).

Recent prior work highlighting the accuracy of operational models has either focused on verification of one numerical model in one region only (McMurdie and Mass 2004), or did not use observations for model verification (Charles and Colle 2009). Therefore, the goal of this work is to document the frequency and magnitude of model mean sea level pressure errors for four operational numerical models along both coasts of the United States using an observed quantity. Mean sea level pressure is chosen for verification as it is a primary quantity for defining key atmospheric features such as synoptic-scale storms and frontal troughs. We focus on model accuracy for the two coastal regions to avoid potential errors due to sea level pressure reduction over complex terrain and to examine whether there are regional differences in model performance. Several modeling systems are evaluated to determine the generality of the results.

2. Datasets and methods

In this study, model forecasts of sea level pressure are compared to pressure observations at coastal buoys and Coastal-Marine Automated Network (C-MAN) stations along the east and west coasts of the United States. Only buoys and coastal land stations were chosen to minimize terrain effects and sea level pressure reduction problems. Eleven sites along each coast are included in the study for three cold seasons: November 2005–March 2006, November 2006–March 2007, and November 2007–March 2008. Sea level pressure is used because it is routinely measured at surface observation stations and directly related to the development and movement of deep tropospheric circulations and structures.

The four operational numerical models considered in this study are the European Centre for Medium-Range Forecasts model (ECMWF) (Bengtsson and Simmons 1983; Simmons and Hollingsworth 2002), the Canadian Meteorological Centre’s Global Environmental Meso-scale model (CMC-GEM) (Côté et al. 1998), and the NCEP GFS (Kalnay et al. 1990) and NAM (Black 1994; Black et al. 2005) models. The major characteristics of each model are given in Table 1. Each model was upgraded during the period of study. For example, the

Model forecast errors in this paper are defined as the differences between the interpolated forecast sea level pressure and the observed sea level pressure at the specified observation sites. In addition to the true forecast error, the calculated errors reflect instrument error, model error due to an incorrect terrain height assignment, and errors in the interpolation of the model output.

<table>
<thead>
<tr>
<th>Date</th>
<th>Model</th>
<th>Operational change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec 2005</td>
<td>CMC-GEM</td>
<td>Assimilation of additional satellite data</td>
</tr>
<tr>
<td>Feb 2006</td>
<td>ECMWF</td>
<td>Increased horizontal resolution to T799 and increased vertical levels to 91 and raised top of model to 0.01 hPa; ozone and gridpoint humidity used in 4DVAR and revised coefficients in ozone chemistry</td>
</tr>
<tr>
<td>Jun 2006</td>
<td>NAM</td>
<td>WRF-NMM replaces the Eta Model; data assimilation switch from Eta 3D Variational Analysis (EDAS) to GSI</td>
</tr>
<tr>
<td>Aug 2006</td>
<td>GFS</td>
<td>New orography and land–sea mask, improved snow analysis, and new ozone physics; fixed error in downward longwave radiation scheme at earth's surface</td>
</tr>
<tr>
<td>Sep 2006</td>
<td>ECMWF</td>
<td>Revisions to cloud scheme; introduction of turbulent orographic form drag scheme; reduction of ocean surface humidity from 100% to 98%; revised assimilation of rain-affected radiances and variational bias correction of satellite radiances</td>
</tr>
<tr>
<td>Oct 2006</td>
<td>CMC-GEM</td>
<td>Increase in horizontal and vertical resolutions; improved physical parameterization, condensation, and precipitation package; new surface scheme implemented</td>
</tr>
<tr>
<td>Dec 2006</td>
<td>ECMWF</td>
<td>Introduced new satellite data including winds from the Multifunctional Transport Satellites (MTSAT) and GPS radio occultation data from the Challenging Minisatellite Payload (CHAMP), the Gravity Recovery And Climate Experiment (GRACE), and the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC)</td>
</tr>
<tr>
<td>Dec 2006</td>
<td>NAM</td>
<td>Improvements made to the convective parameterization and cloud microphysics</td>
</tr>
<tr>
<td>May 2007</td>
<td>GFS</td>
<td>Unified the NCEP 3DVAR assimilation system under the GFS; changed vertical coordinate to hybrid sigma pressure; added new observing systems</td>
</tr>
<tr>
<td>Jun 2007</td>
<td>ECMWF</td>
<td>Improved moist linear physics; revised subgrid-orography scheme; added new shortwave radiation scheme and other changes</td>
</tr>
<tr>
<td>Nov 2007</td>
<td>ECMWF</td>
<td>Made significant changes to the model physics, including the convection scheme, new soil hydrology scheme, new radiosonde temperature and humidity bias correction, assimilation of additional satellite assets, introduced four new pressure levels into the lower troposphere</td>
</tr>
</tbody>
</table>
differs between coasts. The differing coastal variance that the spatial distribution of sea level pressure variance on the two coasts are not found at the same latitudes and would ignore the fact that the synoptically active regions on the two coasts in other ways (such as based on latitude) in error characteristics. To match observation locations synoptic activity in order to examine regional differences between coasts are made between stations of similar high-pass variance matching assures that comparisons their latitudes and longitudes is given in Table 3. The of 1% of the observations missing. A list of stations and stations had nearly complete data, with less than one-quarter nearly identical results (not shown). The chosen 22 sta-

tions were chosen with nearly identical results (not shown). The chosen 22 stations using the total variance, producing and one from the West Coast) were chosen with nearly to eliminate variance on monthly to seasonal time scales. The 30-day high-pass filter was arbitrarily chosen shorter than 30 days) were calculated over the three cold seasons. The same analysis is performed for the mean error (bias) and we define a coastal mean error (CME) as the mean of all the forecast errors for all 11 buoys on a coast for a particular model, forecast hour, and forecast cycle. Because errors based on averaging do not distinguish between a sample containing a small number of large errors or one containing a large number of small errors, we have also calculated the number of times a model exceeds an arbitrarily chosen large error criterion. One model forecast time could potentially have up to 11 large errors on a coast if all of the observation sites exceed the large error criterion.

3. Results
The monthly CAEs for three forecast hours, all four models, and both coasts are given in Fig. 4 for the three cold seasons. It is clear that the monthly CAEs are larger on the West Coast than the East Coast for most months, forecast lead times, and models. This is most
evident in Fig. 4c where the difference between the East Coast and West Coast CAEs is plotted. In Fig. 5, CAEs averaged over the three cold seasons are shown for each model and forecast lead time. In all cases, the West Coast three-cold-season CAEs are larger than the East Coast CAEs at the 99.9% confidence level (using Student’s $t$-test statistics). In addition, the frequency of large forecast errors is larger for the West Coast than the East Coast (Fig. 6). A sea level pressure error is defined as “large” when it is greater than 3, 5, or 7 hPa for forecasts with 24-, 48-, and 72-h lead times, respectively. Comparison of Figs. 6a and 6b reveals that forecasts on the West Coast meet the error criterion at least 2 times more frequently than those on the East Coast.

Figures 4–6 show that consistent differences in forecast accuracy exist among the models; specifically, the ECMWF performs best and the NAM performs worst. ECMWF is the only model for which no forecast errors meet the large error criterion for any month. Forecasts from both NAM models (the Eta and WRF-NMM) consistently exceeded the error criterion more frequently than did the other models. The NAM model had a greater number of forecasts meeting the large error criterion and the highest CAE for all lead times whether evaluated by month, by cold season, or for the entire study period. The differences between the models are substantially smaller for 24-h forecasts on the East Coast than on the West Coast.

---

1 These thresholds are arbitrary, but chosen to distinguish frequent small errors from infrequent large errors. Other thresholds (such as 5, 7, or 9 mb for 24-, 48-, and 72-h forecasts) were examined and the conclusions drawn from those results were similar to those discussed here (not shown).
An alternate method of calculating large errors is to define a forecast to have a large error when at least one station on a coast exceeds the threshold. With this method, the NAM still has the most errors while ECMWF has the least, and the West Coast has more errors than the East Coast (not shown).

The CMEs, which can be thought of as biases, vary significantly by model, month, coast, and forecast lead time (Fig. 7). Each model and coast has a significant CME for most months, but the largest biases are for the GFS on the West Coast during the end of the second cold season. The bias on the East Coast is generally

---

**FIG. 4.** (a) West Coast monthly CAEs in hPa for forecasts with (top) 24-, (middle) 48-, and (bottom) 72-h lead times for the ECMWF (red), CMC-GEM (green), NAM (cyan), and GFS (blue) models. (b) As in (a) but for the East Coast. (c) The difference between (a) and (b). Positive values indicate greater West Coast CAEs.
smaller than on the West Coast, but the bias values relative to the CAE values are similar for both coasts. In general, bias contributes up to 50% of the CAE. The ECMWF usually has a positive bias, while the GFS model typically possesses a negative bias. During the first cold season, the NAM (Eta) had a positive bias on the West Coast, whereas after the switch to WRF-NMM, the biases were negative the second cold season and near zero the third. On the East Coast, the NAM had very small biases for all cold seasons.

Figure 8 shows histograms of CAE on both coasts for the ECMWF and NAM 48-h forecasts. This figure clearly shows the differences in forecast accuracy between the two coasts and two models. The overall mean CAE is about 25% larger on the West Coast for both models, and the spread of errors on the West Coast is larger than on the East Coast (e.g., the standard deviation is 1.05 hPa for the NAM model on the East Coast versus 1.35 hPa for the NAM on the West Coast). For the largest CAEs, the difference is even greater: the West Coast 95th percentile level (a CAE value that is exceeded once every 20 forecasts) is about 40% greater on the West Coast than the East Coast for both models.

Also apparent in Fig. 8 is a marked disparity in forecast accuracy between the models. The ECMWF model had CAEs exceeding 6 hPa on the West Coast only once during the entire study period. However, the West Coast NAM had CAEs reaching that criterion at least 10 times. The results are similar on the East Coast, with ECMWF CAEs never exceeding 4 hPa, but NAM CAEs exceeding that criterion over 30 times.

To determine the temporal evolution of forecast accuracy for the individual models over the three-cold-season study period, CAEs averaged over each cold season for all models and forecast lead times were calculated. These CAEs were then subtracted from the three-season mean CAE for each model and forecast lead time to define an “anomaly” CAE (Fig. 9). Positive values of the anomaly CAE signify that the particular cold season had a smaller CAE (i.e., more accurate) than the three-cold-season CAE. Three sets of significance tests were then performed for each model and forecast hour. First, the anomaly CAE was compared to the three-cold-season mean CAE, and those with confidence levels of 95% and 90% are indicated in Fig. 9 with a double asterisk (**) and a single asterisk (*), respectively. Second, the first-cold-season anomaly CAE was compared to the third-cold-season anomaly CAE and comparisons with confidence levels of 95% and 90% are indicated in Fig. 9 with a double plus sign (+ +) and a single plus sign (+), respectively. And, finally, for the NAM and CMC only, the first-cold-season anomaly CAE was compared to the second-cold-season anomaly CAE and comparisons with confidence levels of 95% and 90% are indicated in Fig. 9 with a double percent symbol (%%) and a single percent symbol (%), respectively.

Of all the models, ECMWF most clearly demonstrates an increase in accuracy of sea level pressure forecasts over the three cold seasons, with the anomaly CAE for the third cold season on both coasts significantly larger (more accurate) than the first-cold-season CAE for all three forecast lead times at the 95% level. The GFS also demonstrates an increase in accuracy over the study period for the West Coast for lead times of 48 h (95% confidence) and 24 and 72 h (90% confidence). However, there are no significant improvements in GFS forecasts of sea level pressure for the East Coast. These improvements in the ECMWF and GFS models occurred even though the synoptic activity was greater...
during the third cold season (Fig. 10). In Fig. 10 the 30-day high-pass sea level pressure variance calculated from the 11 observation stations on each coast is plotted for each cold season. On the West Coast, the first two seasons had similar observed high-pass variances, whereas the third had a significantly higher variance. On the East Coast, the third cold season also had the highest variance, but it was not as large of an increase over the other two cold seasons as on the West Coast.

In contrast to the ECMWF and the GFS models, the CMC model had a modest increase in sea level pressure accuracy over the three-cold-season study period. The second-cold-season anomaly CAE was significantly larger (more accurate) than the first-cold-season anomaly CAE for the 24-h forecasts on both coasts and for the 48-h forecasts on the East Coast at the 95% level. The other forecast lead times show the same trend, but none of the significance tests achieved the 90% level. The CMC experienced a major upgrade between the first and second cold seasons and this study suggests that the upgrade potentially had a positive impact on 24-h sea level pressure forecasts.

Although the NAM switched from the Eta to the new NMM-WRF model over the course of the study period, there is no evidence that there was an improvement in the forecasts of sea level pressure, especially on the West Coast. Modest improvement is found in the 24-h forecasts of sea level pressure on the East Coast, with the third cold-season CAE significantly better than the first-cold-season CAE at the 90% confidence level.

4. Discussion

This paper has demonstrated that large forecast failures of sea level pressure still occur regularly in four operational models from three different countries and that substantial differences in forecast performance exist among the modeling systems. It was noted that large sea level pressure errors occur more frequently on
the West Coast than the East Coast of the United States for every model for 24-, 48-, and 72-h forecasts, and that the ECMWF model outperforms the other models while the NAM is the worst performer.

As described above, there are notable differences in forecast accuracy between the two coasts. For all models and lead times, the East Coast sea level pressure errors are smallest. It is perhaps not surprising that all four models were more skillful over the East Coast, since that region feels the effects of enhanced model initialization made possible by the upstream data-rich continent. The ECMWF is not only the most accurate model, but possesses the smallest difference in forecast errors between the East and West Coasts. The relatively advanced data assimilation system (the four-dimensional variational data assimilation system, 4DVAR) of the ECMWF may be a factor in producing more accurate initializations over the Pacific Ocean.

Of the four modeling systems examined, the NAM model scores much worse overall for each forecast lead time and coast. The NAM’s comparatively poor performance on the West Coast in this study is consistent with the results of McMurdie and Mass (2004) and Charles and Colle (2009), who showed that the NAM was not as skillful as the GFS in forecasting the positions and intensities of surface low pressure systems. In this study, there was no evidence that the switch from the Eta to the WRF-NMM model resulted in improved sea level pressure forecasts. The CAEs and the frequencies of large forecast errors were similar for the NAM across all three cold seasons, and the NAM was the worst of the three models for all three cold seasons, especially for the West Coast. We did not examine possible causes for the poor performance of the NAM. However, possible contributions to the poor NAM performance include its early data cutoff time, the 6-h old lateral boundary conditions, and the three-dimensional variational data assimilation system (3DVAR).

Relative to other models, NAM’s inferior forecast quality is least pronounced for 24-h forecasts over the East Coast. The monthly CAE for the NAM model is similar to the other models for most months, and the NAM model has a similar frequency of large forecast

---

**FIG. 7.** (a) West Coast monthly CMEs (in hPa) for forecasts with (top) 24-, (middle) 48-, and (bottom) 72-h lead times for the ECMWF (red), CMC-GEM (green), NAM (cyan), and GFS (blue) models. (b) As in (a) but for the East Coast.
errors compared to the other models for the 24-h East Coast forecasts. In addition, the cold-season-averaged CAE for the NAM improved from the first cold season (with the Eta) to the third cold season (with the NMM-WRF) for the 24-h forecasts on the East Coast. This improvement occurred despite the increase in sea level pressure variance during the third cold season.

On 31 March 2008, NCEP implemented additional changes to the WRF-NMM model and the gridpoint statistical interpolation (GSI) data assimilation system, including the use of gravity wave drag–mountain blocking and additional satellite assets. The performance of the NAM and the other models will continue to be monitored to see if these changes have a positive affect on forecasts of sea level pressure.

The other three models (ECMWF, CMC-GEM, and GFS) experienced several significant updates over the course of the study (Table 2). The ECMWF modifications included updates to the model physics, increases in vertical and horizontal resolution, and enhanced data assets. A CMC-GEM update in October 2006 applied an increase in vertical and horizontal resolution, a new physics scheme, a reduced time step, a later data cutoff time, and the use of additional satellite data assets. The GFS experienced modifications to its physics and radiation packages, upgrades to its 3DVAR data assimilation system, and increased satellite data assets.

Each of these models demonstrated some improvement in sea level pressure forecasts over the course of the study, with the ECMWF exhibiting the greatest improvement. The CMC had an increase in accuracy from the first cold season to the latter two cold seasons for 24-h forecasts on both coasts. The GFS had more

FIG. 8. (a) Histogram of West Coast CAEs (in hPa) for the NAM model. The mean and standard deviation for each distribution are given. (b) As in (a) but for the East Coast. (c) As in (a) but for the ECMWF model. (d) As in (b) but for the ECMWF model.
accurate forecasts the third cold season compared to the other two seasons on the West Coast, and the ECMWF had more accurate forecasts the third cold season for all forecast lead times and both coasts. The observed sea level pressure variance was largest during the third cold season for both coasts, making it a more difficult season to forecast. Although we have not examined whether specific model upgrades coincide with particular forecast improvements, these results suggest that the upgrades have made a positive impact.

5. Summary

This paper documents model forecast errors in sea level pressure at observation sites along the east and west coasts of the United States for the 5-month cold seasons.
of 2005–06, 2006–07, and 2007–08 for the ECMWF, NAM (Eta and WRF-NMM), GFS, and CMC-GEM models. The errors are used to compare the relative forecast quality between the two coasts and among the four different models. Major findings include the following:

- The West Coast has larger and more frequent errors than the East Coast.
- NAM consistently underperformed other models and ECMWF consistently outperformed other models based on all metrics in this study.
- The NAM operational switchover from Eta to WRF-NMM did not result in improved sea level pressure forecasts on either coast in terms of mean absolute error or frequency of large errors.
- ECMWF experienced general improvement of sea level pressure forecasts over the study period for all forecast lead times on both coasts. The GFS experienced general improvement for all lead times on the West Coast and the CMC-GEM experienced improvement over the study period for 24-h forecasts on both coasts.

These results provide forecasters and model developers with specific information on the ability of four operational numerical models to forecast sea level pressure. The difference in forecast skill between the East and West Coasts suggests the importance of improved observations and data assimilation systems, since inferior initializations over the Pacific—compared to those over the data-rich North American continent—could well be the origin of the coastal skill differences. An implication of our results is that improved data assimilation approaches, coupled with targeted research programs such as THORPEX, may lead to substantially improved prediction, particularly over the west coasts of continents.

Acknowledgments. This paper represents a portion of the first author’s master’s thesis at the University of Washington. This research was supported by the National Science Foundation under Awards ATM-0450684 and ATM-0504028. We acknowledge Mr. David Ovens for developing and maintaining the model verification systems used at the University of Washington.

REFERENCES


