Weather Regimes and Forecast Errors in the Pacific Northwest

LYNN A. McMURDIE AND JOSEPH H. CASOLA
Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 11 June 2008, in final form 17 December 2008)

ABSTRACT

Despite overall improvements in numerical weather prediction and data assimilation, large short-term forecast errors of sea level pressure and 2-m temperature still occur. This is especially true for the west coast of North America where short-term numerical weather forecasts of surface low pressure systems can have large position and central pressure errors. In this study, forecast errors of sea level pressure and temperature in the Pacific Northwest are related to the shape of the large-scale flow aloft. Applying a hierarchical limited-contour clustering algorithm to historical 500-hPa geopotential height data produces four distinct weather regimes. The Rockies ridge regime, which exhibits a ridge near the axis of the Rocky Mountains and nearly zonal flow across the Pacific, experiences the highest magnitude and frequency of large sea level pressure errors. On the other hand, the coastal ridge regime, which exhibits a ridge aligned with the North American west coast, experiences the highest magnitude and frequency of large 2-m minimum temperature errors.

1. Introduction

Numerical weather prediction accuracy has increased dramatically over the last 20 yr due to improvements in computer architecture, atmospheric models, observational data, and data assimilation (Kalnay et al. 1998). Simmons and Hollingsworth (2002) and Harper et al. (2007) demonstrated that current 4-day forecasts of 500-hPa heights and sea level pressure for the Northern Hemisphere are about as accurate as 3-day forecasts of 10 yr ago. Four-dimensional variational data assimilation (Rabier 2005) and the ensemble Kalman filter (Evenson 2003) continue to advance numerical weather prediction.

Despite these hemispheric-scale improvements in the performance of numerical models, individual events of large short-term forecast errors still occur, especially near the west coast of North America and the offshore waters. McMurdie and Mass (2004) documented that the National Centers for Environmental Prediction’s (NCEP’s) Eta Model experienced large sea level pressure forecast errors (greater than 10 hPa) along the west coast of North America 10–15 times a winter. These errors were shown to be associated with large position and central pressure errors of mobile midlatitude cyclones. Charles and Colle (2009) and Wedam et al. (2009) both show that forecast errors of sea level pressure are larger along the west coast compared to the east coast of North America for several operational numerical models.

Analysis and forecast errors over the Pacific and the west coast of North America not only affect local forecast accuracy, but can also affect short- and medium-range forecast skill well downstream, such as eastern North America or the North Atlantic. Hakim (2003, 2005) and Chang (2005) demonstrated that eastward-propagating wave packets are a dominant source of forecast errors over the North Pacific and they can propagate farther downstream and affect forecasts as far away as Europe. Reynolds and Gelaro (2001) found large initial condition sensitivity over the North Pacific for Northern Hemisphere 2-day forecasts. Elmore et al. (2006) found that analysis errors in the NCEP Eta initializations over the west coast were not always corrected in subsequent forecast cycles despite the additional data from the U.S. rawinsonde network. Recognizing the importance of the Pacific region for short- to medium-range weather prediction, The Observing System Research and Predictability Experiment (THORPEX) Pacific-Asian Regional Campaign (T-PARC), which occurred in the fall of 2008, was designed to increase our
understanding of the dynamical connections between weather events over the western Pacific and the development of high-impact weather events over North America and to increase the 1–14-day predictability of these events (Shapiro and Thorpe 2004; Parsons et al. 2007).

The focus of this study is to determine the influence of the shape of the upper-level flow, or weather regime, on the magnitude and frequency of short-term forecast errors of sea level pressure and 2-m temperature for the Pacific Northwest region by using the NCEP Global Forecast System (GFS) and North American Mesoscale (NAM) models. We will examine forecast errors occurring during the winters of 2002–08, classifying each day using a hierarchical limited-contour clustering algorithm that has been applied to the 500-hPa geopotential height data. The algorithm and patterns are identical to those outlined in Casola and Wallace (2007). The goal of this study is to examine the potential factors that contribute to decreased predictability over the Pacific Northwest and offshore waters.

2. Data analysis

a. Classifying days based on the shape of the circulation

To determine the effect of the shape of the circulation on the frequency and magnitude of forecast errors, the winter days between November 2002 and March 2008 were segregated into weather regimes. Following Casola and Wallace (2007), four regime patterns (Fig. 1) were formed by applying a clustering algorithm to 5-day averages of the wintertime [December–March (DJFM)] 500-hPa geopotential height data for the North Pacific region (20°–90°N, 150°E–60°W) for the period 1958–99 using the NCEP–National Center for Atmospheric Research (NCAR) reanalysis dataset (Kalnay et al. 1996). The clustering algorithm uses Ward’s method (Ward 1963; Wishart 1969) to segregate records into groups based upon the latitudinal coordinates of the 540-dam geopotential height contour in the sector. Thus, each regime pattern in Fig. 1 represents the average of a large number of 5-day records that have 540-dam contours with a similar shape and location.

The regimes can be distinguished from one another based on the location and amplitude of their respective ridges and troughs. The offshore trough has a ridge near the Bering Sea and a trough to the west of the North American west coast. The Alaskan ridge exhibits a high-amplitude ridge centered in the Gulf of Alaska and northwesterly flow along the North American west coast. The coastal ridge has a ridge aligned with the North American west coast, and the Rockies ridge exhibits a low-amplitude ridge along the Rocky Mountains with nearly zonal flow across much of the Pacific.

Each day within the 2002–08 study period was classified using these regime patterns. For each day, a 540-dam contour was computed, also using the NCEP–NCAR reanalysis 500-hPa geopotential height data. The contours for each day were generated using the 3-day average of the 500-hPa geopotential height field for the surrounding days, rather than just the daily field. Pattern correlations were calculated between the 540-dam contour for each 3-day average during the 2002–08 winters [November–March (NDJFM)] and the four 540-dam contours corresponding to the regimes. The pattern correlation value can range from −1 to 1, and expresses the amount of similarity between the 540-dam contour from a 3-day record and a particular regime. As described in Cheng and Wallace (1993), it is defined as

\[
 rpq = \frac{\sum_{150^\circ E}^{60^\circ W} \phi_p \phi_q}{\left(\sum_{150^\circ E}^{60^\circ W} \phi_p^2 \sum_{150^\circ E}^{60^\circ W} \phi_q^2\right)^{1/2}},
\]

where \( \phi_p \) is the difference between the latitude coordinates of the 540-dam contour from a 3-day record within the 2002–08 study period and the latitude coordinates of the 540-dam contour from the climatological mean, and \( \phi_q \) is the analogous difference between the latitude coordinates of the 540-dam contour for one of the regimes and that of the climatological mean. Each day within the 2002–08 study period was assigned to the regime with which it had the highest pattern correlation value, provided that this value was greater than the arbitrarily chosen value of 0.5. Days that exhibited pattern correlation values below 0.5 for all four regimes were placed in a fifth category, called transition days.

Of the 908 days in the 2002–08 study period, 194 (22%) were classified as offshore trough days, 117 (13%) were classified as Alaskan ridge days, 133 (15%) were classified as coastal ridge days, 155 (17%) were classified as Rockies ridge days, 297 (32%) were classified as transition days, and 12 (1%) were not classified, since

\[1\] Experiments with an arbitrarily chosen subset of the 1958–99 data showed that using the surrounding days to calculate the 540-dam contour increases the number of days that are identified as being a part of a regime (as opposed to being a transition day). Five-day averages were also tested and found to perform slightly better than the 3-day averages; however, the choice was made to do as little averaging as necessary.
their 3-day averages incorporated days falling outside of the winter as defined here (e.g., 1 November or 31 March).

b. Forecast errors

The NCEP NAM and GFS forecasts are compared to observations in the Pacific Northwest for six consecutive winters (NDJFM) from 2002–03 through 2007–08. The 24-, 48-, and 72-h forecasts of sea level pressure and 2-m temperature from both models at the 0000 and 1200 UTC forecast cycles are included in the study. The forecasts were interpolated to the observation location using a bilinear interpolation scheme. Over the time period of the study, both NCEP models experienced model and data assimilation upgrades, most notably the NAM switched from the Eta Model (Black 1994) to the Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF-NMM) in June 2006 (Rogers et al. 2005; Black et al. 2005). Information regarding the model upgrades can be found in Wedam et al. (2009) and Charles and Colle (2009).

For the sea level pressure verification portion of the study, coastal and offshore buoys and the National Oceanic and Atmospheric Administration (NOAA) Coastal-Marine Automated Network (C-MAN) stations within a box from 40°–55°N to 125°–140°W were used in the study and the locations are shown in Fig. 2. Sea level pressure observation sites were limited to coastal and offshore locations in order to avoid potential errors due to sea level pressure reduction over complex terrain and to assess the quality of the forecasts before storms made landfall. The largest difference between the model terrain heights and the true terrain height at the observation sites was 276 m, and most differences were 50 m or less. The model reduces the surface pressure to sea level, and sea level reduction for locations less than 300 m gives reliable results (Pauley 1998). As an additional check, the analysis was repeated using only the stations whose model terrain heights differed from the true terrain height by 50 m or less. The conclusions based on those results were the same as those presented in this paper. The station data were quality controlled to remove unrealistic values, such as surface pressures greater than 1050 hPa (see Colle et al. 1999).

Forecast error is defined as the model forecast of sea level pressure at the station minus the observed sea level pressure. At each station, a time series of forecast errors are produced for both models and all forecast lead times: a total of six time series per station. For each combination of models and forecast lead times, a daily mean error is calculated by averaging the forecast errors at all the stations for a particular day using both the 0000 and 1200 UTC forecast cycles. The daily mean absolute error is calculated analogously, except the absolute values...
of the forecast errors at all the stations for a particular day are averaged together.

In addition, a large error event is defined as follows. At each observation site, a forecast error is called a large error when the absolute value of that error exceeds 2 times the standard deviation of the errors calculated from the time series. A large error event occurs when two or more stations for a particular day, model, and forecast lead time meet this criterion. Frequently, large error events had large errors at more than two stations.

For the 2-m surface temperature portion of the study, land-based stations from several data networks are used for verification. These networks include the North American surface airways observation network, the C-MAN network (information online at http://www.ndbc.noaa.gov/cman.php), National Weather Service hydrometeorological sites (http://www.weather.gov/oh/hads), remote automated weather stations (RAWS; http://www.fs.fed.us/raWS/), the avalanche center stations (http://www.nwac.gov), the Washington State Department of Ecology network (https://fortress.wa.gov/ecy/enviwa/), and the Hanford (http://hms.pnl.gov) observations sites. Approximately 500 stations located in Washington, Oregon, North California, Idaho, and southern British Columbia, Canada, are included and are distributed as shown in Fig. 3. Since the model terrain is a smoothed version of the true terrain, there are stations where the model estimate of the station elevation differed from the true station elevation. In those circumstances, the model estimate of 2-m temperature was corrected to the station elevation by assuming a lapse rate of 6.5°C (km)$^{-1}$ (Mass et al. 2002). Daily mean errors and daily mean absolute errors were calculated by averaging the mean and mean absolute errors at each of the 500 stations for the 0000 and 1200 UTC verification times separately since these times are near the time of maximum and minimum temperature in the Pacific Northwest. Large error events for temperature were defined as days when the absolute value of the...
domain-averaged temperature was greater than one standard deviation calculated from the six seasons of daily mean absolute errors and mean errors, and not at individual stations. This criteria for large errors emphasized large errors occurring over a significant portion of the domain rather than anomalous conditions occurring at a few stations.

c. Uncertainty in the error statistics

Many of the results are displayed with confidence intervals, or error bars. Two sets of error bars are used. The first set represents the 95% confidence interval for the entire dataset (not segregated by regime) and the second set represents the 95% confidence interval for a particular regime. For the mean absolute error and mean error calculations, the error bars are given by

$$CI = \frac{t_{CRIT} \sigma}{\sqrt{N/2}}.$$

where CI refers to the confidence interval, and $\sigma$ and $N$ are the standard deviation and the total number of days, respectively, for the entire dataset (first group of error bars) or the standard deviation and number of days of the regime (second group of error bars). The critical $t$ value ($t_{CRIT}$) is computed from a $t$ distribution for a confidence level of 0.975 and degrees of freedom equal to the number of days divided by two. The divisor of two was chosen because the $e$-folding time of the autocorrelation of the forecast error time series was between 1 and 2 days.

The $t$ distribution has been used to calculate the confidence intervals since the histograms of the MAE and ME data are nearly normally shaped (not shown). Confidence intervals were also calculated using a non-parametric bootstrap approach; however, they were nearly identical to the confidence intervals calculated using the $t$ distribution.

For large-error events, the error bars are given by a normal approximation to a binomial distribution:

$$CI = t_{CRIT} \sqrt{\frac{N(P(1-P)}{2}}.$$

where $N$ is the number of days in the regime ($N/2$ is the degrees of freedom), $P$ is the probability of a large error for a particular regime or the whole dataset, and $t_{CRIT}$ is obtained from a $t$ distribution for a confidence level of 0.975.

Following Ramsey and Schafer (2002), we focus the discussion of the results on 1) the cases where the mean estimate associated with a regime is not included in the confidence interval of the mean estimate associated with climatology, and visa versa, and 2) cases where two sets of confidence intervals do not overlap. These two cases indicate when the error statistics associated with a particular regime are potentially physically meaningful, as they are highly unlikely to be replicated via random sampling from subsets of the full dataset.

3. Results

a. Sea level pressure errors and weather regimes

The means of the daily mean absolute error and mean error from all the buoys for the six winter seasons for 48-h forecasts by the GFS and NAM separated by weather regimes are shown in Fig. 4. Confidence intervals for the entire dataset for the GFS and NAM and for each weather regime are also plotted in Fig. 4. For simplicity, the mean of the daily mean absolute error and the mean of the daily mean error for each regime will be referred to hereafter as MAE and ME, respectively. In Fig. 4 and subsequent figures, the forecast errors and the regimes are matched at the verification time. Repeating the analysis by matching forecast errors at the verification time with regimes at the initialization time produced similar results (not shown).

The Rockies ridge regime has the largest MAE for both models, whereas the coastal ridge regime has the lowest MAE. The Alaskan ridge regime also has a lower MAE for the GFS model, but not the NAM model. None of the other regimes have error statistics that are distinguishable from the error statistic for the entire dataset. The results for the other forecast lead times, 24 and 72 h, are nearly identical to those shown in Fig. 4 (not shown).

The ME results demonstrate significant differences between the two NCEP models. The NAM has a positive bias for all weather regimes, except the offshore trough. The Rockies ridge ME is the largest and is almost 1 hPa larger than the mean ME of the entire dataset of 0.7 hPa for the NAM model. The GFS model has a negative bias across all the weather regimes with the largest bias for the offshore trough regime and transition days. The Alaskan ridge and coastal ridge regimes, on the other hand, do exhibit smaller negative biases than the mean of the dataset.

The frequencies of large error events for each weather regime for the GFS model and all three forecast times are shown in Fig. 5. When the large error statistics for each regime are compared to the large error statistics for the entire period, the Rockies ridge pattern exhibits the highest frequency (~50%) of large error events compared to the frequency of 33% for the
entire dataset. The coastal and Alaskan ridge exhibit much smaller frequencies of large error events. The results are nearly the same for the NAM model with the Rockies ridge regime exhibiting the highest frequency of large error events and the coastal ridge the lowest. However, the NAM also experienced a slightly higher frequency of large errors for the offshore trough regime compared to the GFS (not shown). For several of the

FIG. 4. NCEP GFS (dark gray) and NAM (light gray) 48-h forecast errors of sea level pressure for each regime: (a) MAE and (b) ME. The gray error bars correspond to the entire dataset and the black error bars correspond to individual regimes.

FIG. 5. Frequency of sea level pressure large error events for the GFS model and all forecast lead times (24, 48, and 72 h) separated by regime. The gray error bars correspond to the entire dataset and the black error bars correspond to individual regimes.
regimes, the percentage of large errors is slightly higher for 24-h forecasts than for 48- or 72-h forecasts. This is likely due to the way that a forecast was flagged as a large forecast error. Since the criteria required that the forecast at a station to be more than 2 times the standard deviation of the errors for the particular model and forecast hour, the criteria is a larger value for 48- and 72-h forecasts than for 24-h forecasts.

b. 2-m temperature errors and weather regimes

The MAE and ME of the 48-h forecasts of the 1200 UTC 2-m surface temperature by the GFS and NAM separated by weather regimes are shown in Fig. 6. In contrast to the sea level pressure results, the MAE for temperature is largest for the coastal ridge regime and is 0.5°C greater than the MAE of the entire dataset of about 2.7°C. The Alaskan ridge pattern also exhibits a large MAE, but it is not distinguishable from the MAE of the entire dataset for both models. The Rockies ridge and offshore trough both exhibit MAEs that are smaller than the entire dataset for both models. The results for other forecast lead times (24 and 72 h) are very similar to those shown in Fig. 6 (not shown).

The MEs plotted in Fig. 6 indicates some regime-dependent model biases. The coastal ridge regime exhibits the largest positive bias for the NAM, implying that this model often predicts minimum temperatures to be warmer than observed during that regime. Both the offshore trough and Alaskan ridge regimes have large negative biases for the GFS model, implying that the model often predicts the minimum temperatures to be colder than observed for these regimes, especially the Alaskan ridge. Similar biases to those shown in Fig. 6 were found for the 24- and 72-h forecast lead times (not shown).

The MAE and ME 0000 UTC 2-m surface temperatures of both models and all forecast hours exhibit similarities to the 1200 UTC results in Fig. 6 except that the MAE for the offshore trough and Alaskan ridge and occasionally the Rockies ridge regimes are not different from the MAE of the entire dataset (not shown). Unlike at the 1200 UTC verification time, the ME for 0000 UTC has a large negative bias for all regimes and forecast hours and is especially large for the GFS (~1.5°C). This implies that the GFS tends to forecast colder maximum temperatures than observed for all weather regimes.

When the temperature data are segregated into stations located east and west of the Cascade Mountains, the results at 1200 UTC are nearly identical to those shown in Fig. 6 (not shown). The only difference is that the MAEs are larger for stations located east of the Cascades.

The frequencies of large error events by regime verifying at 1200 UTC for the GFS model and all three
forecast times are shown in Fig. 7. The coastal ridge regime has a higher frequency of large error events compared to the other regimes and the offshore trough has the fewest. The frequencies of large error events for the Alaskan and Rockies ridge regimes and transition days are similar to the frequency of large error events for the entire dataset. The NAM exhibits very similar results (not shown). The frequency of large error events at 0000 UTC is similar to those shown in Fig. 7 for both models, but the coastal ridge regime has more frequent large error events than the mean for the entire dataset for only the 24- and 72-h forecasts.

c. Examples of large error events

In this section, we present two examples of large forecast error events to illustrate how large forecast errors are manifested in a synoptic-scale setting. The first example is an intense midlatitude cyclone that approached the Pacific Northwest. The daily 500-hPa geopotential height field obtained from the NCEP–NCAR reanalysis data valid at the initialization and verification times are shown in Fig. 8. On 9 November 2007, the flow is nearly zonal and strong across the Pacific with ridging over the Rockies. Short-wave troughs are evident in the flow at approximately 160°W and 160°E. The 3-day 500-hPa pattern centered at this time was classified as a Rockies ridge with a pattern correlation of 0.6. At verification time, the short waves embedded in the upper-level flow have amplified and, subsequently, the Pacific Northwest experienced strong southwesterly flow. The infrared satellite image of the storm and the GFS 72-h sea level pressure forecast (contours) and sea level pressure errors (numbers) are given in Fig. 9. The 72-h GFS forecast failed to indicate a deepening surface low and only analyzed a weak low 700 km south of the actual position of the storm. Forecast errors in the vicinity of the storm exceeded 14 hPa in a large region near 50°N, 130°W. Forecasts by other operational models also experienced large forecast errors of cyclone position and depth. Subsequent forecast cycles by all operational models (i.e., the 24- and 48-h forecasts) valid for the same time continued to experience large sea level pressure errors. This particular storm deepened as it came ashore across Vancouver Island and produced strong damaging winds in the northern Puget Sound region and significant power outages occurred (see Storm Data; information available online at http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwEvent~Storms). Because of the inferior numerical guidance for the storm track and intensity, it was difficult to predict if, where, and when strong winds would occur. Although this example may be a more intense storm than is typically experienced in the Pacific Northwest, the types of forecast errors, such as the position and intensity of surface lows, are representative of most events that are characterized by large forecast errors.
The second example illustrates the conditions typically found for large forecast errors of 2-m temperature. In Fig. 10, the daily 500-hPa geopotential field obtained from the NCEP–NCAR reanalysis data on 29 November 2002 exhibits a large high-amplitude ridge located along the West Coast and two short-wave troughs to the west at 150°W and 170°E. The 3-day 500-hPa pattern centered at this time was classified as a coastal ridge with a pattern correlation of 0.8. This event occurred during a string of several days classified as a coastal ridge. The 72-h forecast errors of 2-m temperature by the GFS model valid at this time (Fig. 11) show a significant portion of the verification region having large errors and several stations with errors of 10°C or larger. There are large portions of the domain where the forecast errors are large and positive, such as along the Puget Sound region in Washington, the eastern Washington basin, and the Willamette Valley in Oregon. Near-surface wind speeds reported at many of the locations plotted in the figure were light, 2.5 m s$^{-1}$ or less. In these regions the
model-forecasted minimum temperatures are too high in valleys where nocturnal cold pools formed under the quiescent conditions of a strong upper-level ridge. High-elevation stations experienced both large positive and negative errors. In these regions, a combination of the model underestimating the strength of the inversion produced by radiational cooling and our assumption of how to handle the mismatch of station and model terrain potentially contributed to the forecast errors. Subsequent forecast cycles (i.e., at 48- and 24-h lead times) also experienced widespread large temperature forecast errors. This case is typical of the large error events during the coastal ridge regime where the forecast errors of temperature are large and the observed surface winds are light.

4. Summary and discussion

In the prior section, it was shown that the shape of the upper-level flow pattern modulates short-term forecast errors along the West Coast. The results can be summarized as follows:

- The Rockies ridge regime, which is characterized by ridging along the Rockies and nearly zonal flow across the Pacific, exhibits the largest average MAE of sea level pressure and the highest frequency of large error events of sea level pressure errors for both models and all forecast lead times.
- The coastal ridge regime, which is characterized by ridging along the West Coast, exhibits the largest average MAE of 2-m temperature at 1200 UTC (i.e., minimum temperature) and the highest frequency of large error events of 2-m temperature errors for both models and all forecast lead times.

In the remainder of this section, we explore the physical factors that may explain why the shape of the upper-level flow would be associated with large short-term forecast errors of sea level pressure and 2-m temperature.
As shown in Blackmon et al. (1977) and Wedam et al. (2009), variance in daily sea level pressure is an indicator of synoptic storm activity. In Fig. 12, the observed sea level pressure variance calculated from the NCEP–NCAR reanalysis data for each regime is composited for the study period. In Fig. 12, the large values of observed sea level pressure variance associated with the Rockies ridge are displaced farther to the east and closer to the verification region than for the other regimes. This indicates that mobile synoptic storms are active in the area close to or within the verification regime during the Rockies ridge regime. Considering that the Rockies ridge pattern is similar to the zonal flow patterns examined by Nutter et al. (1998) and Roebber and Tsonis (2005), our study confirms that such patterns are less predictable than higher-amplitude flows. The link between zonal flow and low predictability makes sense physically as forecast errors would arise from the movement of upper-level short waves and their surface low pressure centers embedded within the zonal flow.

The coastal ridge stands out as the regime exhibiting small forecast errors of sea level pressure and large forecast errors of 2-m temperature for both models and all forecast lead times. As seen in Fig. 12, the coastal ridge pattern is associated with relatively small sea level pressure variance across the North Pacific and especially along the West Coast. Therefore, the coastal ridge is associated with reduced storminess and reduced sea level pressure errors. However, the 2-m temperature errors during the coastal ridge regime are large. During this regime, the ridge axis is over the verification region and the reduced synoptic activity allows valley cold pools to form nocturnally in regions of complex terrain. In addition, temperature observations may be inaccurate and less representative of nearby locations compared to situations with more horizontal and vertical mixing (Myrick and Horel 2006). Our results concur with other studies that showed lower forecast accuracy of surface temperature during persistent cold-pool events in regions of complex topography especially during the nighttime and early morning hours (Hart et al. 2004; Myrick and Horel 2006; Cheng and Steenburgh 2007). There are several potential sources contributing to poor numerical model performance during this type of synoptic situation, such as inaccurate boundary layer parameterization and differences in model and observation station elevation.

In this study, the clustering algorithm only identifies different shapes of the upper-level flow patterns and does not take into account the strength of the upper-level flow. To address this limitation, the error data were composited by the magnitude of the upper-level wind speed. However, the relationships between wind speed and error frequency were not independent from the relationships derived from the shape of the flow. Specifically, the days with strong upper-level winds had many large error events, but this category also disproportionately included many Rockies ridge days and very

---

**FIG. 10.** The daily 500-hPa geopotential height field derived from the NCEP–NCAR reanalysis data valid 29 Nov 2002 for an example of a large forecast error event for 2-m temperature that occurred during a coastal ridge regime. The contour interval is 60 m with the 5400-m contour in boldface.
few coastal ridge or Alaskan ridge days. Yet, there are times when the flow cannot be categorized by regime (i.e., a transition day). In those cases, the strength of the upper-level flow is a second option for indicating the likelihood of large forecast errors of sea level pressure.

Although the overall conclusions of this study appear to be independent of forecast model and forecast lead time, some model differences were highlighted in the results. The most robust difference was that the NAM model consistently had larger sea level pressure and temperature errors for most regimes and lead times compared to the GFS (see Figs. 4 and 6). Although both models exhibited the largest errors of sea level pressure for the Rockies ridge regime, the NAM also exhibited large errors for the offshore trough regime. Thus, the NAM had difficulty with both regimes associated with relatively high sea level pressure variance (Fig. 12) and mobile synoptic-scale storms. This study and others (Charles and Colle 2009; Wedam et al. 2009) demonstrate that the GFS model is a superior model to the NAM in the Pacific Northwest. We did not examine possible causes for why the forecast errors were larger for the NAM. However, possible contributions include its early data cutoff time, the 6-h-old lateral boundary conditions, and the three-dimensional variational data assimilation system. Additional information on the differences between the GFS and NAM can be found in Charles and Colle (2009) and Wedam et al. (2009).

The overall conclusions were not significantly affected by model upgrades over the course of the study. The most notable upgrade was the switch from the Eta to the WRF-NMM of the NAM model in June 2006. To see if the switch yielded different results for the Eta and the WRF-NMM, all the results were repeated for the Eta and WRF separately. The overall conclusions based on the results were the same for both models. However,

Fig. 11. The 2-m temperature 72-h forecast errors (°C) (forecast minus observed) from the GFS model valid at 1200 UTC 29 Nov 2002 for the example of a large forecast error event that occurred during a coastal ridge shown in Fig. 12. The boldface numbers are errors with absolute values greater than 3°C.
the WRF appeared to have larger sea level pressure errors for the offshore trough and the Alaskan ridge than the Eta, but this is likely due to the shorter time period for the WRF model. The ME for sea level pressure of the two models was different. The Eta Model had large positive biases for all regimes while the WRF had smaller positive biases for the Alaskan, coastal, and Rockies ridge regimes and negative bias for the offshore trough. The upgrades to the resolution and model physics did not significantly affect the results for the GFS. When the data were subdivided into groups before and after the change of resolution (May 2005), the results were similar for both periods.

Many forecasters use the concept of weather regimes or analogs when describing the weather. This study provides an objective method for defining regimes and quantifies each regime’s forecast performance. The study is not well suited for direct application to everyday weather forecast procedures, but some aspects could be incorporated. For example, computing a 3–5-day average of the 500-hPa forecasts would be useful to aid in the identification of future weather regimes and their potential for yielding large forecast errors. In addition, the presence of strong upper-level jets in the eastern Pacific in forecast model output should alert the forecaster to the potential for larger than average forecast errors. More study is needed to explore other ways to effectively apply these results to guide forecasting procedures.

Acknowledgments. This material is based on work supported by the National Science Foundation under Grant 0450684. We thank the three anonymous reviewers for their thorough and thoughtful comments on this manuscript. The NCEP–NCAR reanalysis data were provided by the National Oceanic and Atmospheric Administration/Earth System Research Laboratory/Physical Science Division (NOAA/OAR/ESRL/PSD) in Boulder, Colorado, from their Web site (http://www.cdc.noaa.gov). Dave Ovens at the University of Washington created and maintains the model verification software used in this study.

REFERENCES


