Measuring Skill of Numerical Weather Prediction

Garrett Blaine Wedam

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington
2008

Program Authorized to Offer Degree:
Department of Atmospheric Sciences

University of Washington

Graduate School

This is to certify that I have examined this copy of a master's thesis by

Garrett Blaine Wedam

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee members:

_____

Lynn McMurdie

_____

Clifford F. Mass

_____

Gregory Hakim

Date: _____

In presenting this thesis in partial fulfillment of the requirements for a master's degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Any other reproduction for any purposes or by any means shall not be allowed without my written permission.

Signature: _____

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

1. INTRODUCTION

Winter weather affects life, limb and property every year in the United States. Nearly all injuries resulting from winter storms (95%) are due to vehicle collisions and individuals caught outdoors unsuspecting of dangerous weather (National Weather Service 2001). With accurate warning and preparation systems in place, injuries and property losses can be minimized. The crucial first step in prevention is the ability to provide accurate public forecasts. Today's forecasts are considerably more accurate than at any previous time, and this increase in skill is tied to rapid improvement in numerical weather prediction (NWP). All metrics for measuring model forecast skill show great improvement during recent decades and the magnitude of the most extreme failures observed has been significantly reduced (Harper et al. 2007; Simmons and Hollingsworth 2002; Kalnay 1998). However, large model errors still occur with some regularity, and because NWP has evolved to become the primary tool for forecasters, large model errors lead to incorrect public forecasts (McMurdie and Mass 2004).

Figure 1.1 shows an example of a 72-hr North American Mesoscale (NAM) model forecast verifying at 0000 UTC December 24, 2006 that possessed very large errors of forecast sea level pressure in the northeast Pacific. 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. The major synoptic features associated with the poor forecast are a misplaced center of low sea level pressure at approximately 45°N, 148°W and a forecast center of low pressure that did not verify at 52°N, 125°W.

For forecasts verifying at the same time, different models can exhibit different solutions. Sometimes one or more operational forecast models will exhibit very large errors, while others do not. Such was the case for forecasts verifying on 0000 UTC December 24, 2006 on the East Coast of the United States. In this case, a low center developed off the coast of Maine that was not forecast by some models. Table 1.1 shows the errors in sea level pressure at Mt. Desert Rock Station (MDRM1) on the

coast of Maine for four forecasting models: European Centre for Medium-Range Weather Forecasts model (ECMWF) (Bengtsson and Simmons 1983, Simmons and Hollingsworth 2002), the Canadian Meteorological Centre's Global Environmental Mesoscale model (CMC-GEM) (Cote et al. 1998), and the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) (Kalnay et al. 1990) and NAM (Black 1994, Black et al. 2005) models. NAM had the largest forecast SLP errors, but CMC-GEM and ECMWF also had quite large maximum errors for some forecast lead times, while forecasts by the GFS model had much greater skill. Errors in numerical weather prediction arise from uncertainty in initial conditions (Lorenz 1963; Anthes 1986) and imperfect physical parameterizations (Frank 1983; Grell et al. 1991). Because different modeling systems have different methods of assimilating data and parameterizing physics, it would not be surprising to regularly observe differences in forecast skill. Indeed, a timeline (Fig. 1.2) showing SLP errors exceeding a "large error criterion" shows that sometimes all models meet the criterion, while other times just one does, or some meet the criterion and some do not.

Forecast failures with societal impacts, such as in the previous example, are observed with some regularity. In order to gain insight into the origin and nature of model forecast errors, this study examines differences in frequency and magnitude of errors for different models and different geographical regions. Knowing the quality of different models would be of immediate use to forecasters reviewing varying model solutions. Information on model errors and their regional variability is useful in guiding 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 day one to fourteen forecasts (Shapiro and Thorpe 2004, Parsons et al. 2007).

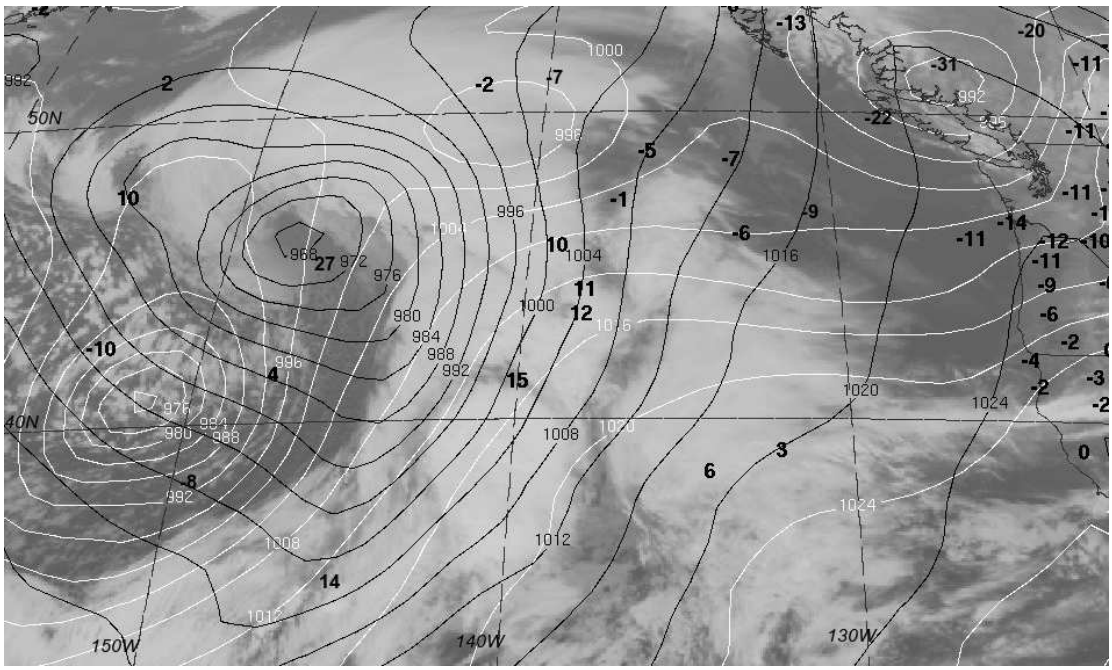


Figure 1.1 Infrared satellite imagery, NCEP NAM 72-h forecast of sea level pressure valid 0000 UTC 24 December 2006 (white contours), NCEP NAM 00-h analysis of sea level pressure for 0000 UTC 24 December 2006 (black) and sea level pressure errors (forecast – observed) calculated at observation sites (black numbers).

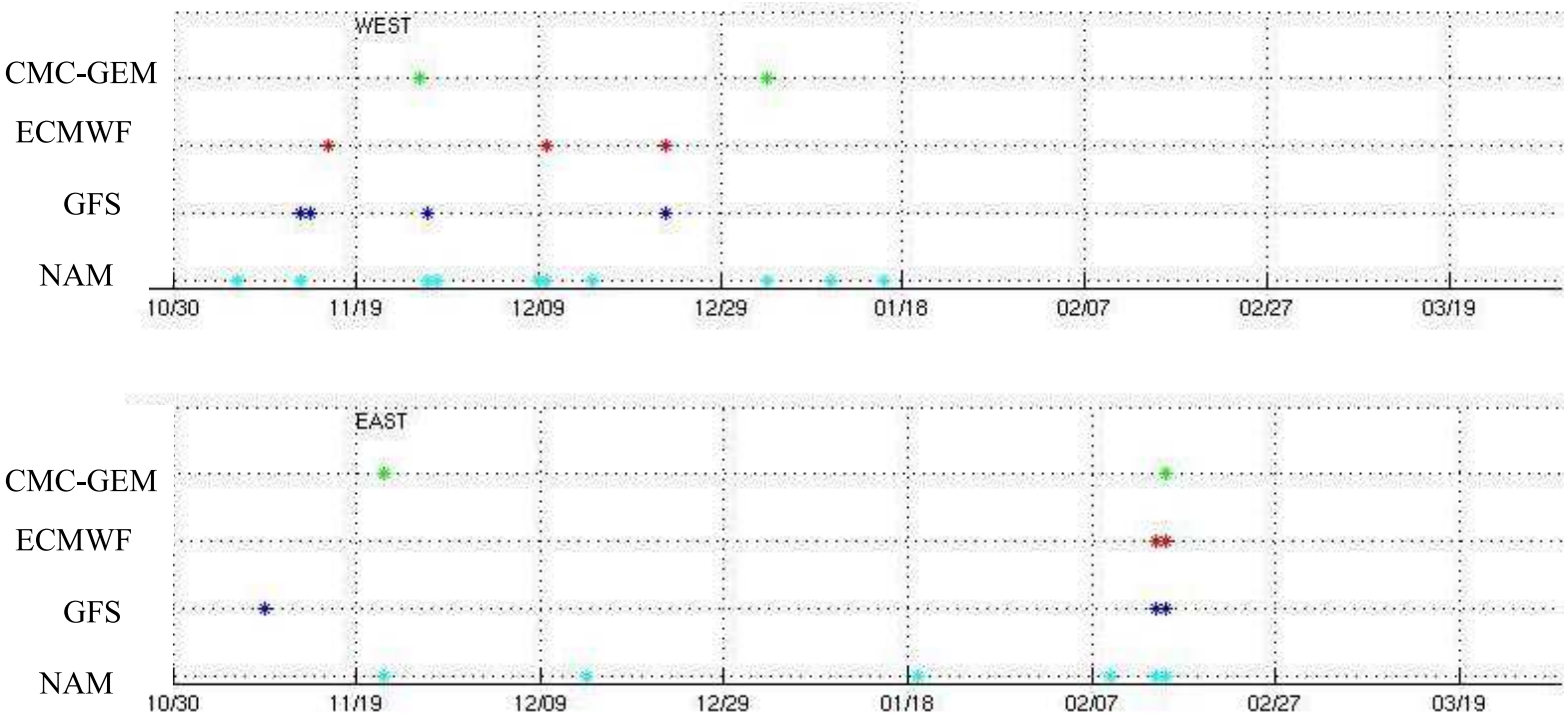


Figure 1.2 Timeline of large sea level pressure errors averaged over 11 buoys on the West and East Coasts of the United States during the cool season of 2005/2006. Asterisks represent averaged 24-hour forecast errors greater than 5 hPa, and the different colors correspond to different operational models.

Table 1.1 Sea level pressure errors at Mt. Desert Rock, ME for the 72-hr forecast verifying at 0000 UTC December 24, 2006 for four operational forecast models.

Model Name	Error (hPa)
72-hour forecasts	
NAM	24.1
CMC-GEM	14.4
ECMWF	10.0
GFS	3.2
48-hour forecasts	
NAM	13.7
CMC-GEM	9.1
ECMWF	7.6
GFS	1.5
24-hour forecasts	
NAM	4.7
CMC-GEM	2.4
ECMWF	2.7
GFS	-0.2

2. BACKGROUND

2.1 HISTORY OF NUMERICAL WEATHER PREDICTION

The history of NWP is thoughtfully documented and detailed in a number of recent publications (Kalnay 1998; Harper et al. 2007) and this section of this paper will simply serve to briefly outline major aspects of the evolution of NWP.

The fundamental theories describing NWP were developed decades before they were put to practical use. Vilhelm Bjerknes (1904) introduced the concept of using known equations to describe the evolution of the Earth's atmosphere, and Lewis Richardson (1922) provided a method for solving them numerically. He conceded that such an approach would take 64,000 accountants working together with calculators to just keep pace with the weather. Both a feasible method of performing the necessary calculations and a political climate favoring such a major investment into weather forecasting were finally brought together at the close World War II. Along with the advent of the digital computer, there was a ten-fold increase in the number of trained meteorologists due to the war, and thousands of newly returned pilots desired accurate aviation forecasts (Harper et al. 2007). With both the ability and desire for advancing forecasting methods, the Meteorology Project began at Princeton in 1946 to determine the feasibility of NWP, and was closely followed by the launching of two additional modeling efforts that were all aimed to develop NWP. One was Carl-Gustov Rossby's team of meteorologists working in Stockholm, and the other was by the United States Air Force. After several years of development and armed with modest successes using barotropic and baroclinic models, these teams of theoretical meteorologists joined with operational meteorologists in a successful call to the skeptical United States Weather Bureau to fund the development of operational computer-generated forecasting, and NWP quickly advanced during the 1960s (Harper et al. 2007).

Starting from the time computers were first utilized in operational forecasts in 1954, major investment in NWP and observation systems have resulted in steady and significant improvements to model forecast skill. An examination of S1 scores

(which are a measure of forecast accuracy that examines errors in pressure gradients) reveals that – from the 1960's to 1990 – the accuracy of 72-hour forecasts were 10-20 years behind the accuracy of 36-hour forecasts (Harper et al. 2007). Even in recent decades, improvement to weather forecasts has been considerable, as evidenced in a study by Simmons and Hollingsworth (2002) that compared forecasts and analyses for sea level pressure and 500 mb height throughout the extra-tropical Northern Hemisphere and found the equivalent of a one-day increase in forecast skill for three operational models during the 1990s. Harper et al. (2007) and Kalnay et al. (1998) documented increasing skill of the National Centers for Environmental Prediction's (NCEP) operational models for 500 mb heights and other forecast parameters.

Numerical weather prediction is hailed as atmospheric science's greatest achievement of the 20th century. Even after 50 years of rapid development, today's opportunities to increase accuracy and expand applications of NWP seem as plentiful as ever. In fact, the first issues that pioneering forecast modelers needed to address to pursue more accurate forecasts – limited computing resources, development of data assimilation techniques amidst expanding data assets, and development of parameterization techniques (Harper et al. 2007) – are still aspects of NWP that, when improved, yield most fruitful returns.

2.2 ERRORS IN NUMERICAL WEATHER PREDICTION

Prior to the mid-1980s, failure of models to predict rapid cyclogenesis routinely resulted in unpredicted storms with major societal impacts (Sanders and Gyakum 1980). Some prominent examples include the Queen Elizabeth II storm of September, 1978 (Gyakum 1983a, b; Anthes et al. 1983; Uccellini 1986), and the eastern Pacific storm of November 1981 (Reed and Albright 1986; Kuo and Reed 1988). Some of the poor forecasts were exacerbated by poor observing networks that allowed even the short-term forecasts to miss an event. But even with physics improvements and modern observing techniques (such as global satellite imagery and broad radar coverage), major failures of NWP models can still lead to major short-

term public forecast errors, as when a “surprise” windstorm caused considerable damage in populated regions of Oregon and grounded the New Carissa cargo ship on the Oregon coast in March 1999 (McMurdie and Mass, 2004).

Although the recent improvements to forecast models have all but eliminated the model forecast failures on the scale seen prior to the mid-1980s, major model failures of short-to-medium range forecasts still occur regularly. This was demonstrated by McMurdie and Mass (2004) when they showed that NCEP’s NAM model had large 0- to 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 under-forecast surface low-pressure systems. Large sea level pressure errors can also be produced by timing errors. For example, Colle (2001) found trough passage timing errors along the Pacific Northwest coast as large as 15 hours.

2.3 GEOGRAPHICAL DIFFERENCES FOR MID-LATITUDE CYCLONES – THE UNITED STATES EAST AND WEST COASTS

The United States East and West Coasts have different synoptic characteristics. Notably, the East Coast is characterized by some of North America’s highest rates of cyclogenesis (Roebber 1984) and anticyclone-lysis, and the West Coast is characterized by frequent cyclolysis (Zishka and Smith 1980) and very low rates of cyclogenesis (Roebber 1984).

Because this study examines errors on the East and West Coasts of the United States, it is worth noting that previous studies of model forecast errors have indicated that the characteristics of the errors on the two coasts are not the same. In one such study comparing two models, Colle and Charles (2008) used NCEP’s GFS analyses to verify GFS and NAM model forecasts of cyclones over different regions of North America. They found larger central pressure errors in the northeast Pacific and North

American west coast for short-term forecasts (24- to 72-h) but larger errors in the North Atlantic for forecasts greater than 96-h.

3. COASTAL DIFFERENCES IN OBSERVED SEA LEVEL PRESSURE BY SPECTRAL FREQUENCY

3.1 INTRODUCTION

Before forecast accuracy can be fairly compared between two geographical regions, differences in the meteorology must be taken into account. Because this study focuses on differences in the skill of model forecasts of sea level pressure on the West and East Coasts of the United States, this section examines differences in the characteristics of observed sea level pressure on the two coasts. The observed synoptic-scale meteorology on the two coasts is quite different, as discussed in section 2.3. Whether due to the passage of a front, or to changes in the relative proximity of low and high pressure centers, synoptic-scale meteorology is often the cause of sea level pressure changes. Because of this, one would not necessarily expect time series of sea level pressure on the two coasts to exhibit the same characteristics.

This chapter examines previous research in order to draw connections between time series of observed sea level pressure and synoptic-scale meteorological structures. As a means to determine an appropriate method for comparing forecast skill on the East and West Coasts, potential coastal differences in such meteorological structures are examined by performing a spectral analysis on observed sea level pressure on both coasts. To carry out this objective, a seventh-order Butterworth Filter (Guillemin, 1957) was applied to the time series power spectrum. This filter is frequently used in meteorological time series studies for its recursive use and its distinct advantage in having a smooth but very sharp transition between the pass region and suppressed region (Shanks, 1967; Murakami, 1979). Details of this filter, including its response function and use in geophysical analyses are described in Murakami (1979).

3.2 BACKGROUND

Several studies in the 1970s examined the connection between the power spectrum of sea level pressure (or various height levels) and structures related to synoptic meteorology. These studies showed that a predominance of large-scale synoptic waves or a predominance of shorter waves could be recognized by the resulting sea level pressure power spectrums. The large-scale synoptic waves have a lot of low frequency power which rapidly decreases with increasing frequency. Shorter waves, on the other hand, have less low frequency power but have more slowly decreasing power with frequency (Blackmon 1976). More specifically blocking patterns and frequent mid-latitude cyclone passage have characteristic power spectrums. The regions with frequent blocking patterns have greater power in periods between 20 and 60 days, while frequent cyclone passage is characterized by greater power in periods less than ten days: these shorter periods have a variance maximum coincident with Northern Hemispheric storm tracks (Blackmon 1977, Sawyer 1970). It is important to note that, for periods greater than a month, most of the observed variance is related to seasonal variability, El Niño Southern Oscillation, and other phenomena not directly attributed to synoptic-scale weather features (Blackmon 1977, Sinclair et al. 1997).

3.3 METHODS

In this study, a spectral analysis is performed on time series of twice-daily observations of sea level pressure from sixteen stations along the West and fifteen stations along the East Coast of the United States during three November-March cool seasons. The stations are nearly evenly spaced along each coast. In addition, a subset of stations is examined using hourly data as well as full-year data to check for reproducibility and seasonal differences. A seventh-order Butterworth Filter is used to separate the observations into high- and low-pass time series, and the effects of cutoff periods of 6, 10 and 30-days are examined. A chunk length of 128 was used with a Hanning window (Blackman and Tukey, 1959). The resulting resolution

provides an ability to distinguish between periods as similar as 10 and 12 days with an uncertainty corresponding to 17 degrees of freedom. For greater certainty, the subset of stations with hourly data provided over twice the resolution and over 50 degrees of freedom.

3.4 RESULTS

Compared to the East Coast, a much greater portion of the West Coast's total variance in sea level pressure is due to structures with long time scales (low frequency). Figure 3.1 shows the power spectrums of a pair of stations (one from each coast). The West Coast station (46088) has a peak in the power spectrum at about 22 days, whereas the East Coast station (mdrm1) has a peak at less than ten days. A much greater portion of the variance at East Coast stations is due to high frequency structures (periods less than 10 days), as evidenced by Figs. 3.2 and 3.3a. Interestingly, the portion of variance that is from high-frequency structures increases with increasing total variance on the East Coast, but decreases with increasing variance on the West Coast (Fig. 3.2b). When a Butterworth Filter cutoff period of six days was examined, the results were analogous (not shown). With a 30-day cutoff period (Fig. 3.3b), the majority of the observed variance falls in the high-pass portion for both the East and West Coast stations, but the West Coast's low-pass portion is greater than the East Coast's. Results using hourly data were indistinguishable from results using twice-daily data. An examination of yearly data shows that the East Coast has significantly more variance in the summertime than the West Coast. The low frequency contribution is similar year-round for the Northern East Coast stations, but it drops off dramatically on the West Coast (not shown). West Coast station 46088 and East Coast station mdrm1 both have about 110 hPa^2 total sea level pressure variance during the cool season, but 19 hPa^2 and 60 hPa^2 (respectively) during the remaining 7 months.

3.5 DISCUSSION

From section 3.2, there is indication that on the West Coast the variance contributed from long waves results in a power spectrum that has a strong peak at low frequencies and that the power rapidly drops off with increasing frequency. Similarly, the East Coast power spectrum fits the description of variance contributed from shortwaves. Because much of the power is contributed from structures with periods greater than 20 days for the West Coast, one could surmise that a larger portion of the contribution to the West Coast total sea level pressure variance is due to blocking. Because much of the power is contributed from structures with periods less than 10 days for the East Coast, one could similarly surmise that a larger portion of the contribution to the East Coast total variance is due to cyclones, fronts and short waves.

In order to compare a model's ability to forecast in different regions, one cannot simply compare error statistics of sea level pressure. This approach would invariably lead to a bias of greater errors at the location that experiences greater variance in sea level pressure. When comparing the East and West Coasts, simply matching regions by latitude would ignore the fact that the synoptically active regions on the two coasts are not found at the same latitudes and that the spatial distribution of sea level pressure variance differs between coasts. So instead, regions must be matched by this observed variance. But because the observed variance on the East and West Coasts do not originate from equivalent structures, a decision has to be made about which structures – based on knowledge about relationships between sea level pressure power spectrum and meteorological phenomena – should be included in the calculation of variance. First, models' abilities to forecast mid-latitude cyclones are certainly important, so the variance due to high frequency structures should be included. In this study, it is decided that models' abilities to forecast during different regimes (i.e. blocking) is important, so periods corresponding to this should also be included. However, since contributions to the total variance from structures with periods greater than about 30 days measure significant contributions from seasonal and interannual variability in addition to some blocking patterns, these

should not be included. Therefore, applying a high-pass filter with a cutoff period of 30 days is determined to be the approach that best isolates the occurrence of synoptic-scale meteorological phenomena. Matching stations on the East and West Coasts using this approach assures that stations of similar synoptic activity are selected, allowing for fair examination of regional differences in error characteristics. An additional benefit to this approach is that it results in an intuitive match between the locations on the East and West Coasts that are associated with the most frequent winter weather. This station-matching approach is utilized for the study in Chapter 4.

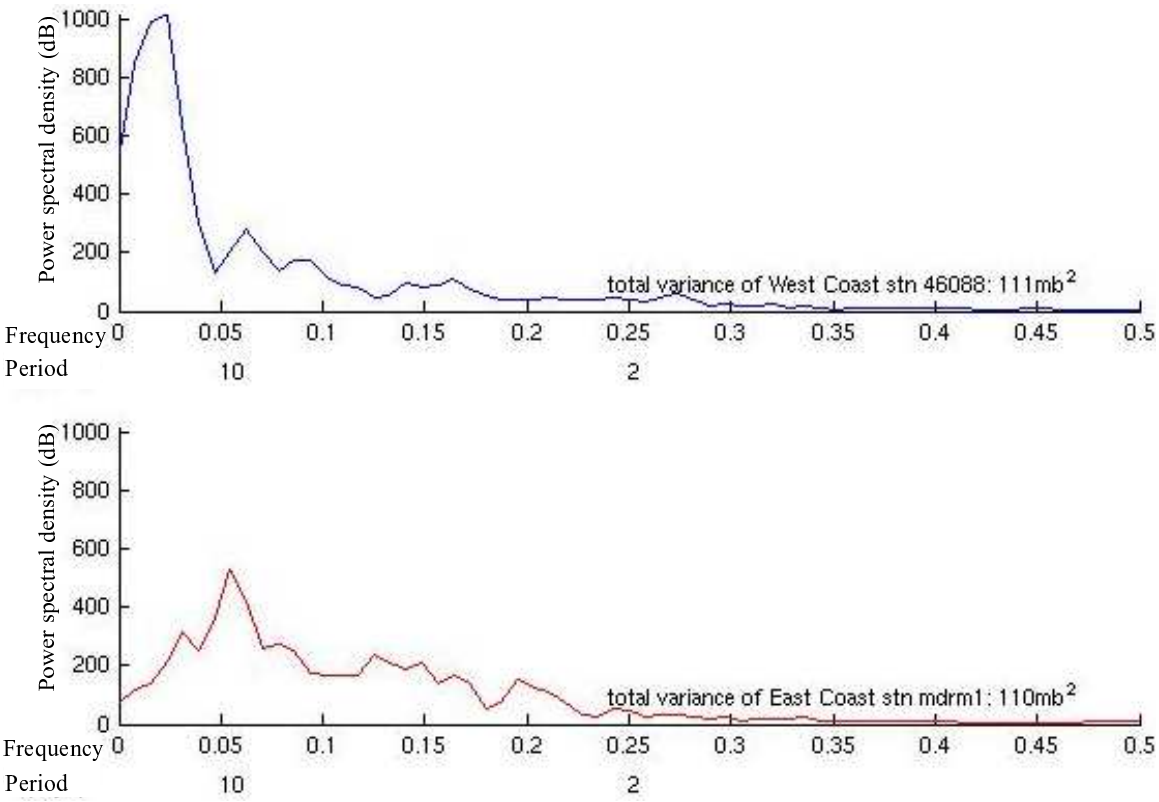
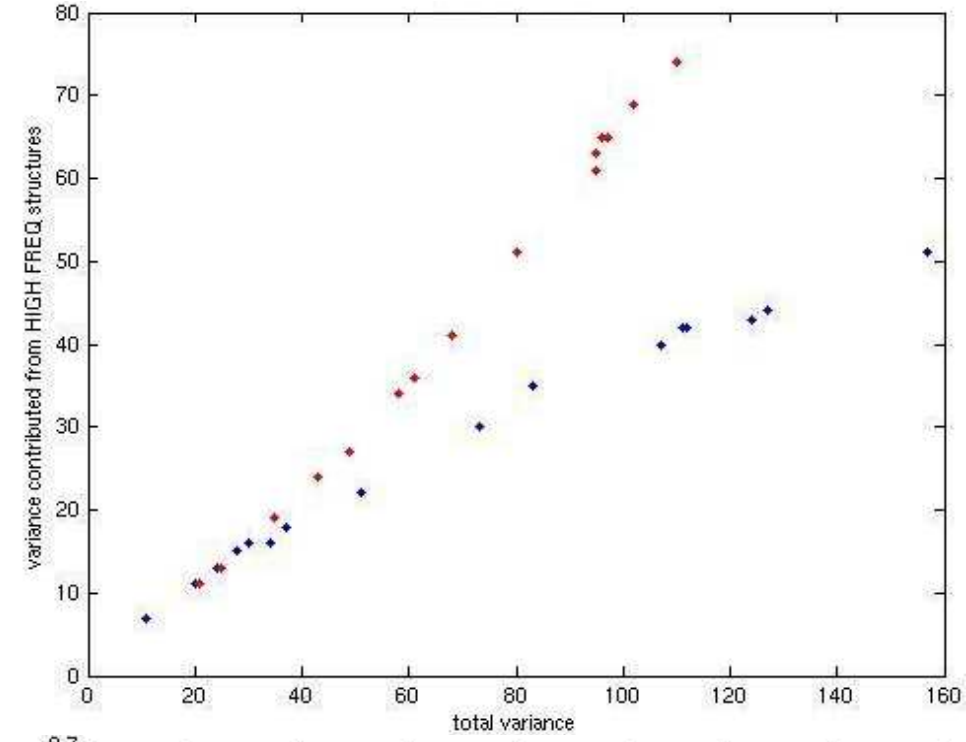


Figure 3.1 Power spectrum [power spectral density (dB)] by frequency (using twice-daily sampling) of West Coast station 46088 (blue) and East Coast station mdrm1 (red) with nearly identical observed sea level pressure variance. The total sample includes about 900 forecasts. The location of 10 and 2 day periods are noted below the x-axis. The total observed variances are noted for each station.

(a)



(b)

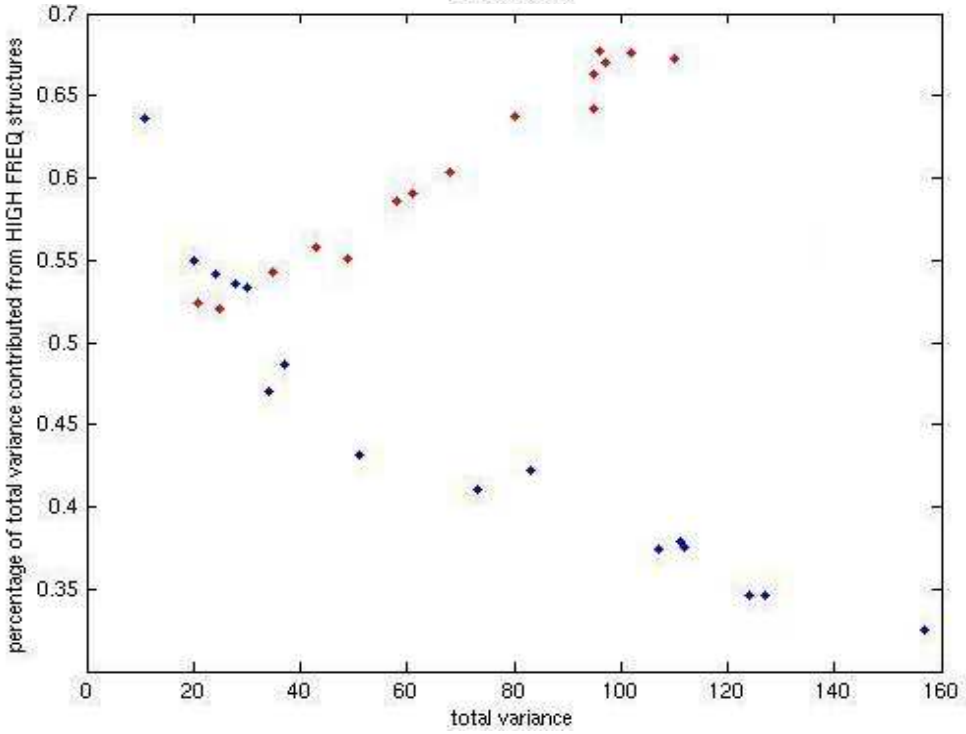


Figure 3.2 (a) High-passed (10-day cutoff period) sea level pressure variance as a function of total variance for East Coast stations (red) and West Coast stations (blue). (b) Percentage of total variance that is high-passed as a function of total variance.

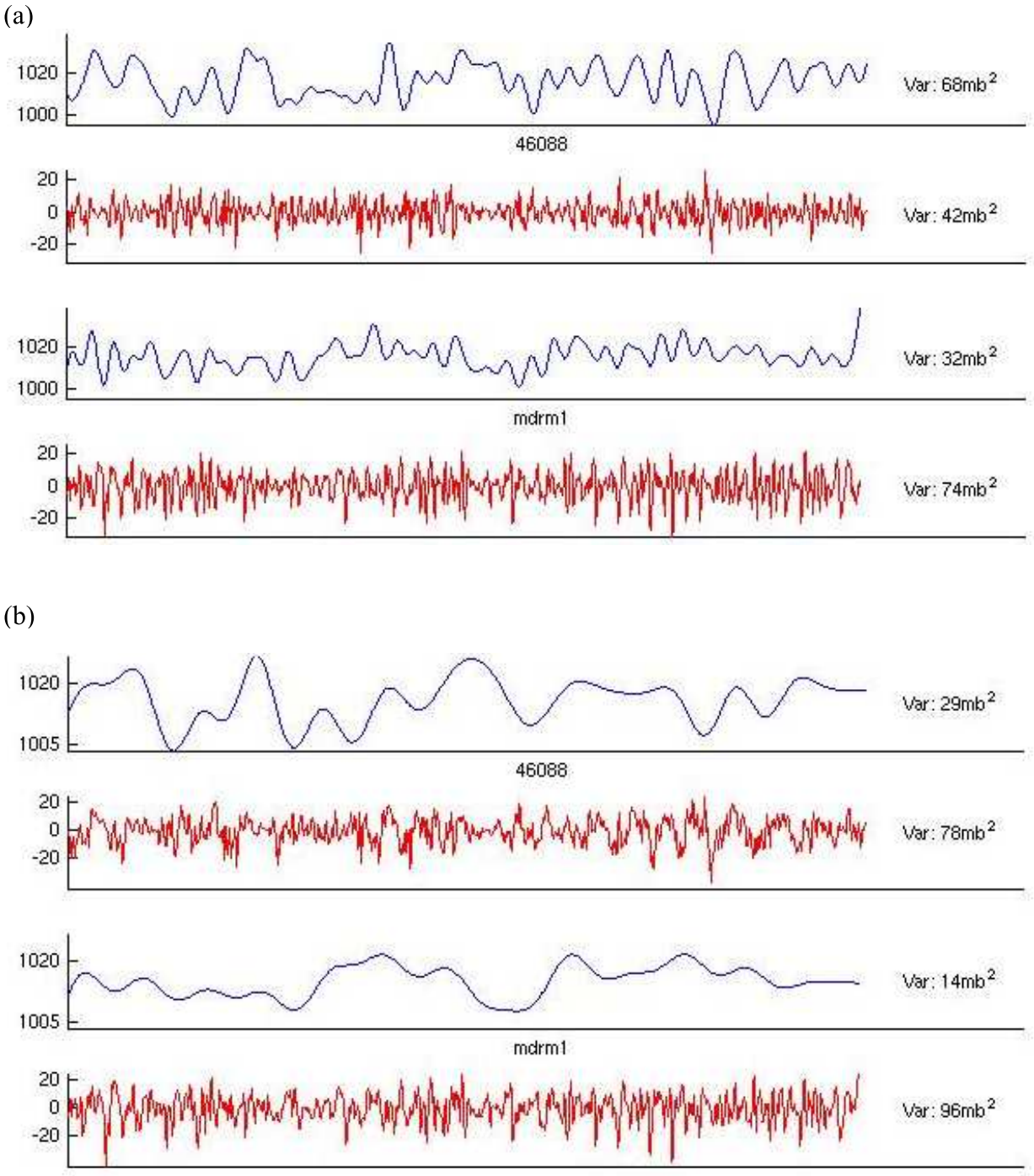


Figure 3.3 (a) Filtered time series with low-passed (blue) and high-passed (red) data for West Coast station 46088 and East Coast station mdrm1 for about 900 forecasts made twice-daily. The pass was made using a Butterworth Filter with a cutoff period of 10 days. The variance contribution from each time series is to the right. (b) Same as (a) except using a cutoff period of 30 days.

4. COMPARISON OF MODEL FORECAST SKILL OF SEA LEVEL PRESSURE ALONG THE UNITED STATES EAST AND WEST COASTS

4.1 INTRODUCTION

In this chapter, the frequency and magnitude of model forecast errors for four different operational forecast models are examined at observation sites along the East and West Coasts of the United States for three five-month cold seasons. The station-matching method designed in Chapter 3 is employed to compare error statistics at stations on the two coasts that experience equal high-passed sea level pressure variance. The average sizes of sea level pressure errors on each coast are compared as well as the frequency with which certain error thresholds are exceeded. Error statistics for the four models are compared to determine their relative forecast skill, and indications of model improvement during the study period are examined.

4.2 METHODS

In this study, model forecasts of sea level pressure are compared to pressure observations at coastal buoys and Coastal Marine Automated Network (CMAN) 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 ECMWF, the CMC-GEM, NCEP GFS and NAM models. The major characteristics of each model are given in Table 4.1. Each model was upgraded during the period of

study. For example, the NAM model switched from the Eta to the Weather Research and Forecasting Non-hydrostatic Mesoscale Model (WRF-NMM) on June 20, 2006. Table 4.2 provides a summary of model updates; further details are posted on the model websites[1].

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 grids to the observation location. Typical instrument accuracy is on the order of 1 hPa[2]. For several of the West Coast sites, the model terrain heights differ from the true terrain height. The largest difference is 276 m, and most differences are 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). Therefore, no additional corrections were made to the model estimates of sea level pressure. Assuming the pressure errors are unbiased and Gaussian, they can be neglected when considering statistics averaged across many realizations.

To select observation sites for this study, buoys and CMAN stations along the East and West Coasts were considered. At each site, the observed sea level pressure total variance and high-pass filtered variance (time-scales shorter than 30-days) were calculated over the three cold seasons. The 30-day high-pass filter was chosen to eliminate variance on monthly to seasonal time-scales (see Chapter 3). The decision

---

[1] CMC (http://www.msc.ec.gc.ca/cmc/op_systems/recent_e.html), ECMWF (http://www.ecmwf.int/products/data/operational_system/evolution/index.html), GFS (http://wwwt.emc.ncep.noaa.gov/gmb/STATS/html/model_changes.html) and NAM (http://www.emc.ncep.noaa.gov/mmb/mmbpll/eric.html#TAB4).

[2] Instrument accuracy is given by the National Data Buoy Center at http://www.ndbc.noaa.gov/rsa.shtml

to use a 30-day cutoff period is discussed in detail in Chapter 3. Eleven observation pairs (one site from the East Coast and one from the West) were chosen with nearly equal high-pass variance for each paired member. We also paired stations using the total variance, producing nearly identical results (not shown). The chosen 22 stations had nearly complete data, with less than one quarter of one percent of the observations missing. A list of stations and their latitudes and longitudes is given in Table 4.3. The differing coastal variance distributions can be seen in Fig. 4.1, which presents the location of the observation sites used in this study along with their observed high-pass variance in sea level pressure. The relative contributions to the total observed variance by different spectral frequencies varies significantly between the East and West Coasts. This finding is examined in Chapter 3. However, the observed sea level pressure distributions of the paired sites are remarkably alike, with a similar number of extreme values and a similar kurtosis about the mean sea level pressure value. Two representative pairs of East and West Coast sea level pressure distributions are shown in Fig. 4.2.

At each observation station, an absolute error is calculated for each model, for 24- 48- and 72-h forecasts, and for the 0000 and 1200 UTC forecast cycles. A coastal absolute error (CAE) for a given model, forecast hour and forecast cycle is defined as the mean of the absolute errors for all 11 buoys on a coast. A monthly or seasonal CAE is defined as the mean of all the CAE's calculated over the time period of interest (month, cold-season, or three-cold-season study period). 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 eleven large errors on a coast if all observation sites exceed the large error criterion.

4.3 RESULTS

The monthly coastal absolute errors (CAEs) for three forecast hours, all four models, and both coasts are given in Fig. 4.3 for the three cold seasons. It is clear that the monthly CAE's are larger on the West Coast than the East Coast for most months, forecast lead-times and models. This is most evident in Fig. 4.3c where the difference between the East Coast CAE and West Coast CAE are plotted. In Fig. 4.4, CAE's averaged over the three cold seasons are shown for each model and forecast lead-time. In all cases, the West Coast three-cold-season CAE's are larger than the East Coast CAE's at the 99.9% confidence level (using t-statistics). In addition, the frequency of large forecast errors is larger for the West Coast than the East Coast (Fig. 4.5). 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.[3] A comparison of Fig. 4.5a with Fig. 4.5b reveals that forecasts on the West Coast meet the error criterion at least two times more frequently than those on the East Coast.

Figs. 4.3, 4.4 and 4.5 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 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.

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

[3] These thresholds are arbitrary, but chosen to distinguish frequent small errors from infrequent large errors. Other thresholds (such as 5, 7, or 9-hPa for 24, 48, and 72-hour forecasts) were examined and the conclusions drawn from those results were similar to those discussed here (not shown).

method, the NAM still has the most errors while ECMWF has the least, and the West Coast has more errors than the East (not shown).

The coastal mean errors (CME), which can be thought of as biases, vary significantly by model, month, coast, and forecast lead-time (Fig. 4.6). 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 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.

Fig. 4.7 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 than the East Coast for both models.

Also apparent in Fig. 4.7 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.

In order to determine the temporal evolution of forecast accuracy for the individual models over the three-cold-season study period, CAE's averaged over each

cold season for all models and forecast lead-times were calculated. These CAE's were then subtracted from the three-season mean CAE for each model and forecast lead-time to define an 'anomaly' CAE (Fig. 4.8). 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 on the figure with a '**' and '*', 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 on the figure with a '++' and '+', respectively. And, lastly, 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 on the figure with a '%%' and '%', 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 48-h lead time (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. 4.9). In that figure 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 variance, whereas the third had significantly higher variance. On the East Coast, the third cold season also had the highest variance, but it wasn't 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 superior 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. The implication of this lack of improvement is that other sources of error, such as the common data assimilation system, might be important for NAM.

4.4 DISCUSSION

This study 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 among the modeling systems in forecast performance exist. It was noted that large sea level pressure errors occur more frequently on the West Coast than the East Coast of the U.S. 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 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 Colle and Charles (2008), 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 CAE and the frequency 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 errors compared to the other models for the 24-h East Coast forecasts. In addition, the cold-season averaged CAE for the NAM improved between 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 GSI data assimilation system, including the use of gravity wave

drag/mountain blocking and additional satellite assets in the GSI data assimilation system. 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 4.2). The ECMWF modifications included updates to 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 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 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.

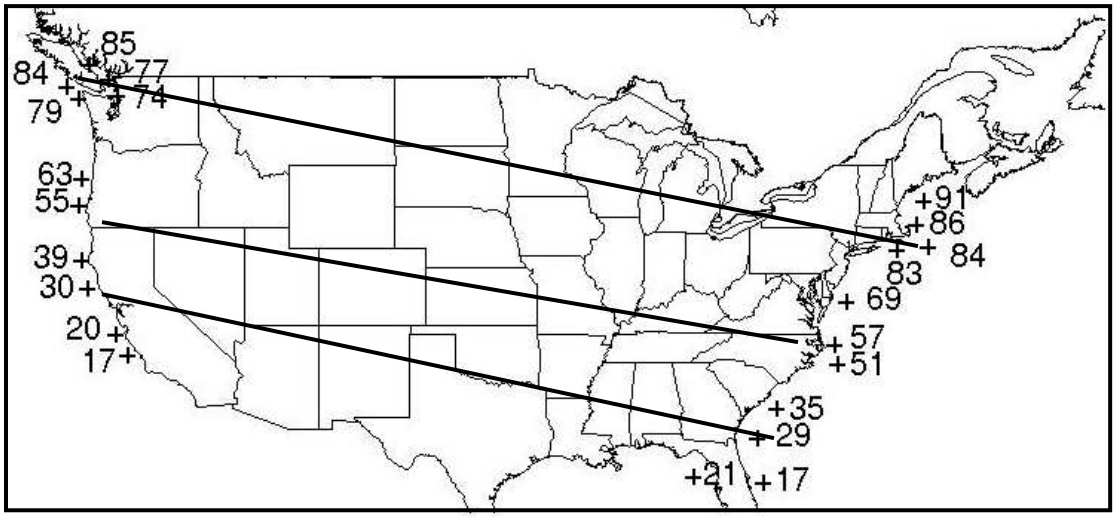


Figure 4.1 Locations of the coastal stations used in this study are indicated by the plus signs. The adjacent numbers are the 30-day high-pass variance in sea level pressure [hPa^2] at each station calculated from all winters. The solid lines match buoys on each coast with similar variance.

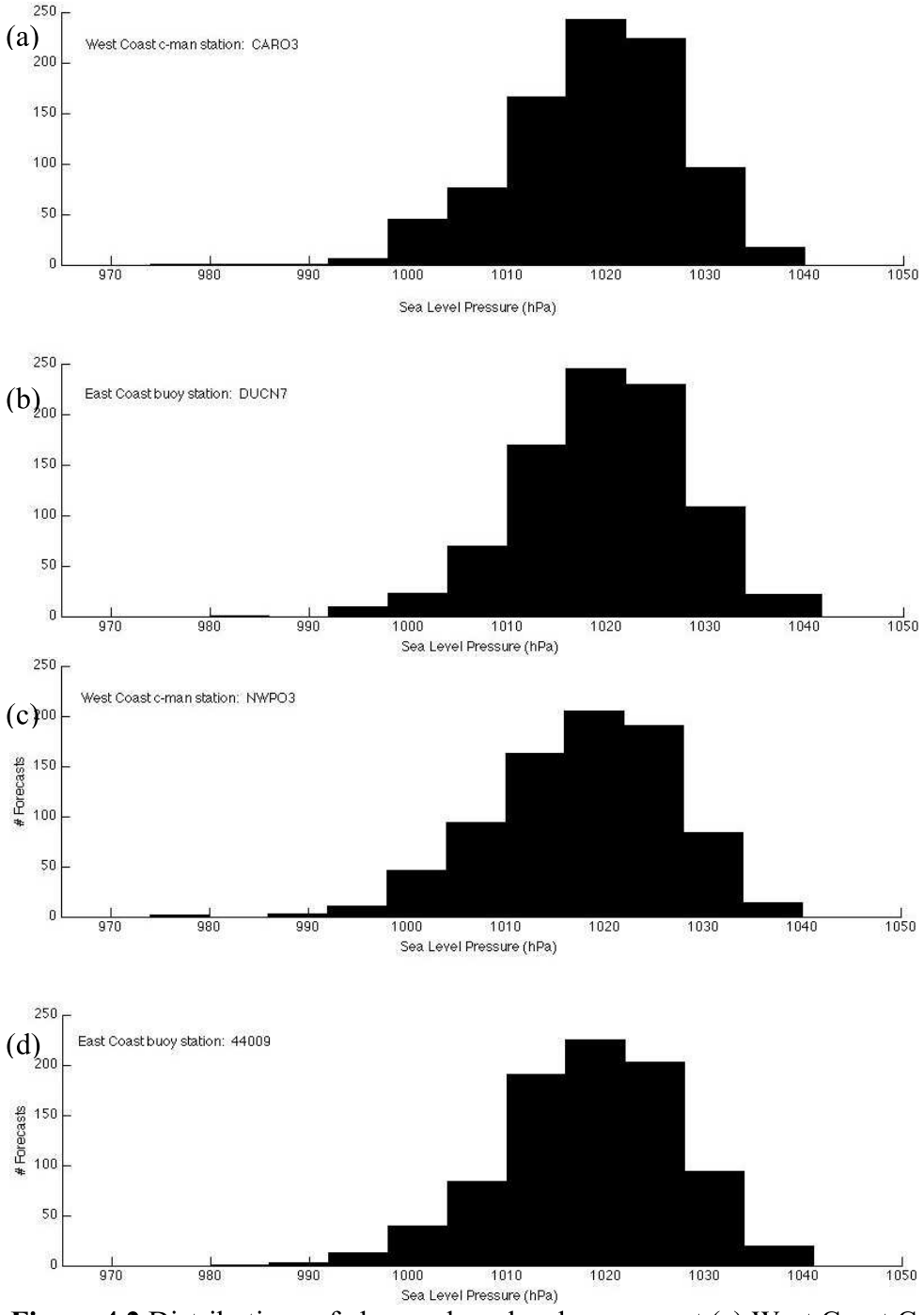


Figure 4.2 Distributions of observed sea level pressure at (a) West Coast Coastal Marine Automated Network (C-MAN) station Cape Arago, Oregon (CARO3), (b) East Coast C-MAN station Duck Pier, North Carolina (DUCN7). Both have a sea level pressure variance of about 70 hPa^2 for the three cold seasons. Distributions of observed sea level pressure at (c) West Coast C-MAN Newport, Oregon (NWPO3) and (d) East Coast buoy 44009. These stations both have a variance of about 80 hPa^2.

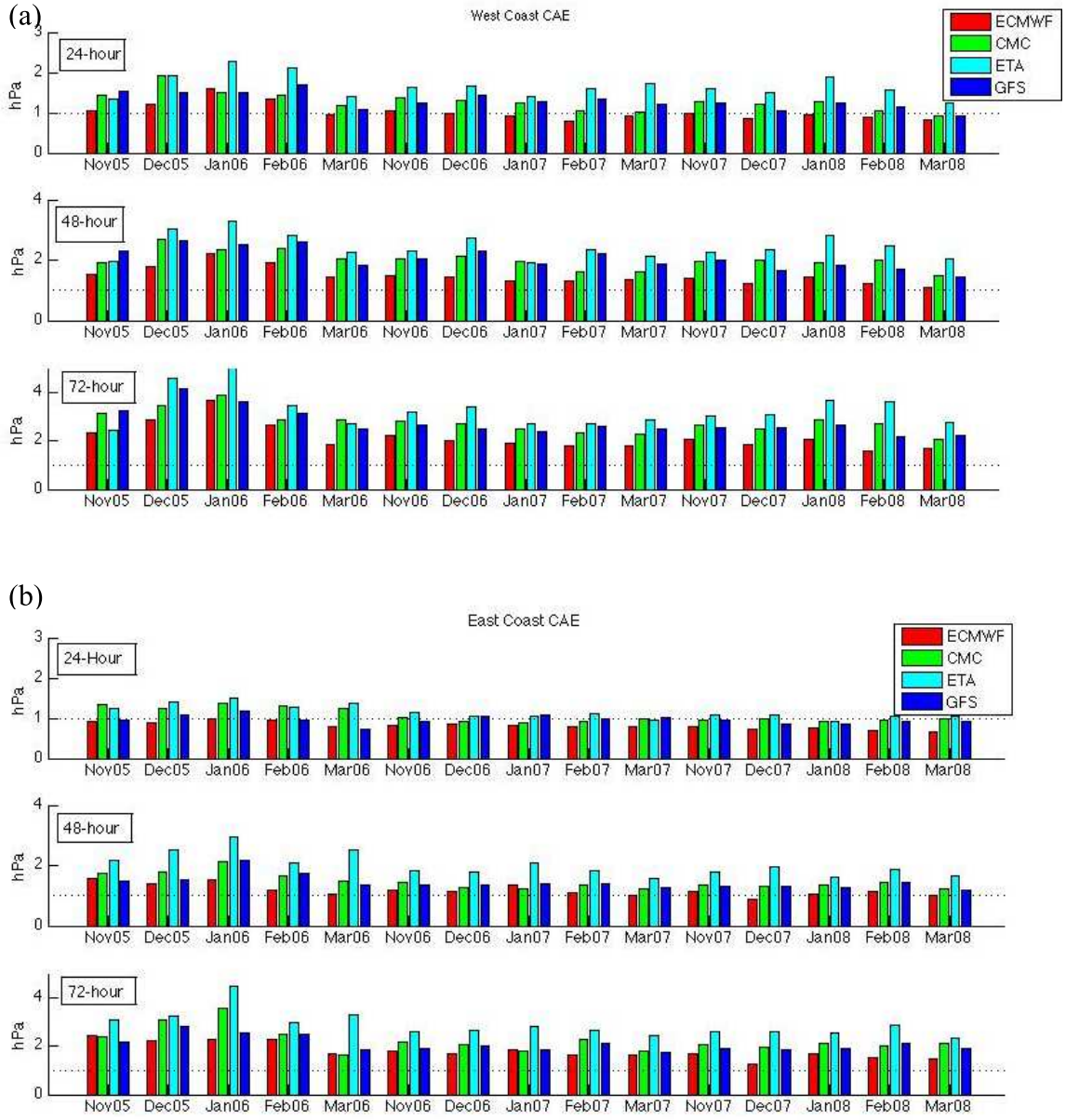


Figure 4.3 (a) West Coast monthly coastal absolute error (CAE) in hPa for 24-h (top), 48-h (middle), and 72-h (bottom) forecasts for ECMWF (red), CMC-GEM (green), NAM (cyan), and GFS (blue) models. (b) Same as (a) only for the East Coast.

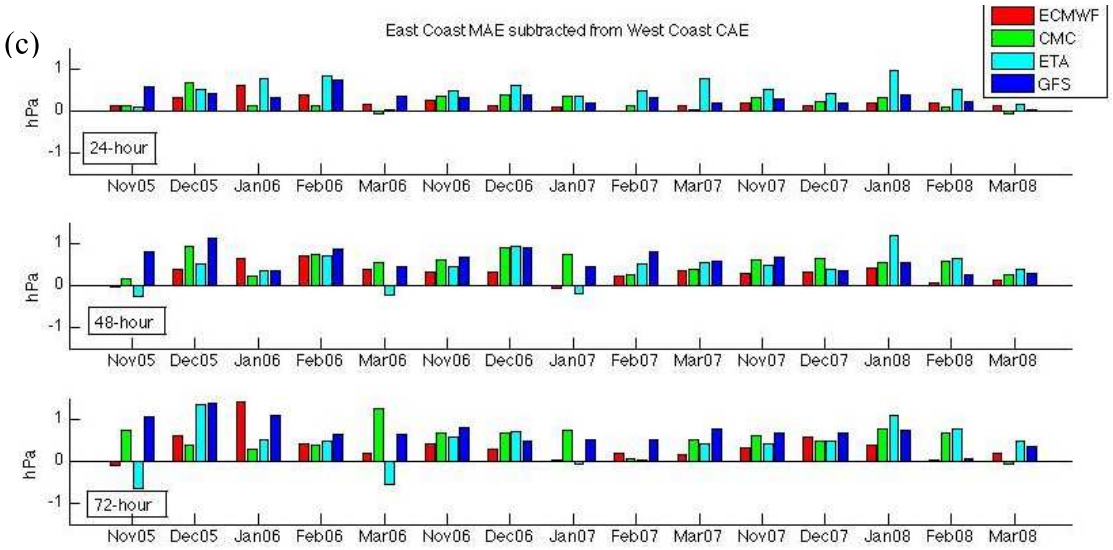


Figure 4.3 (c) The difference between (a) and (b). Positive values indicate greater West Coast CAE.

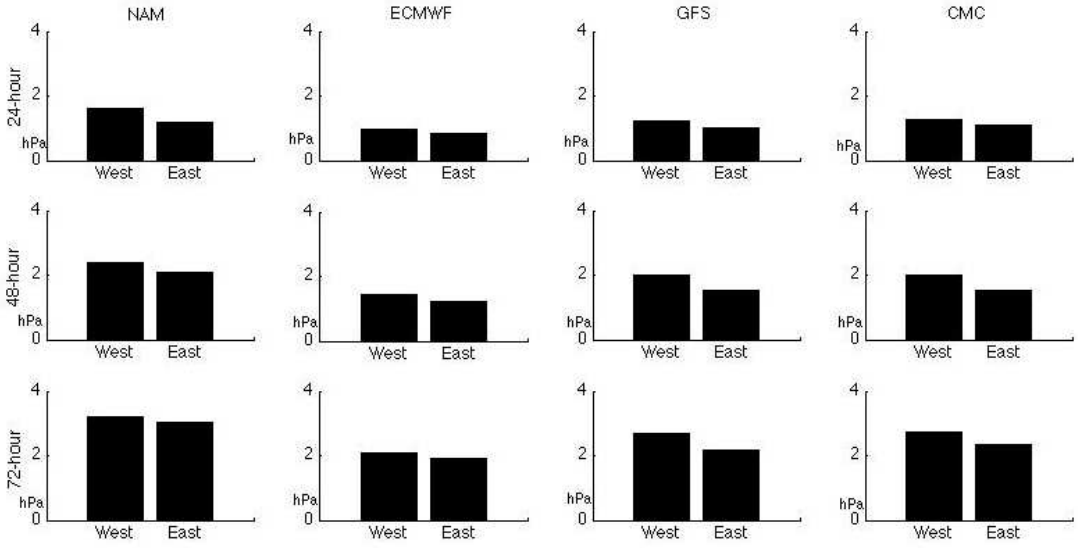


Figure 4.4 Mean absolute coastal errors [hPa] averaged over three 5-month cold seasons by model, coast and forecast lead-time.

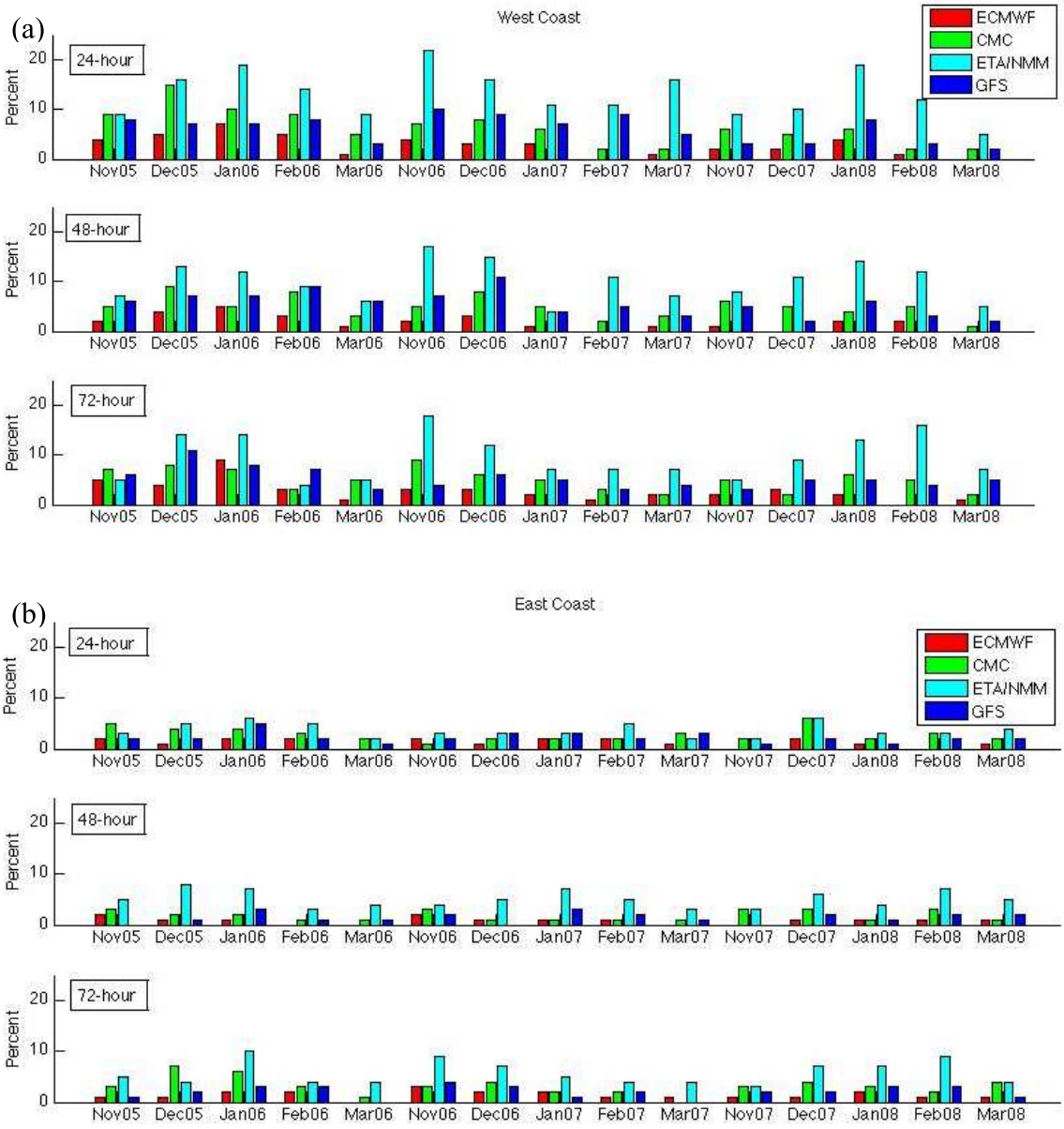


Figure 4.5 (a) The frequency of errors (in percent) at West Coast sites with magnitudes greater than 3, 5, 7 hPa for forecasts with 24-h (top), 48-h (middle), and 72-h (bottom) lead-times. Values are shown by month for the ECMWF (red), CMC-GEM (green), NAM (cyan), and GFS (blue) models. (b) Same as (a) only for the East Coast.

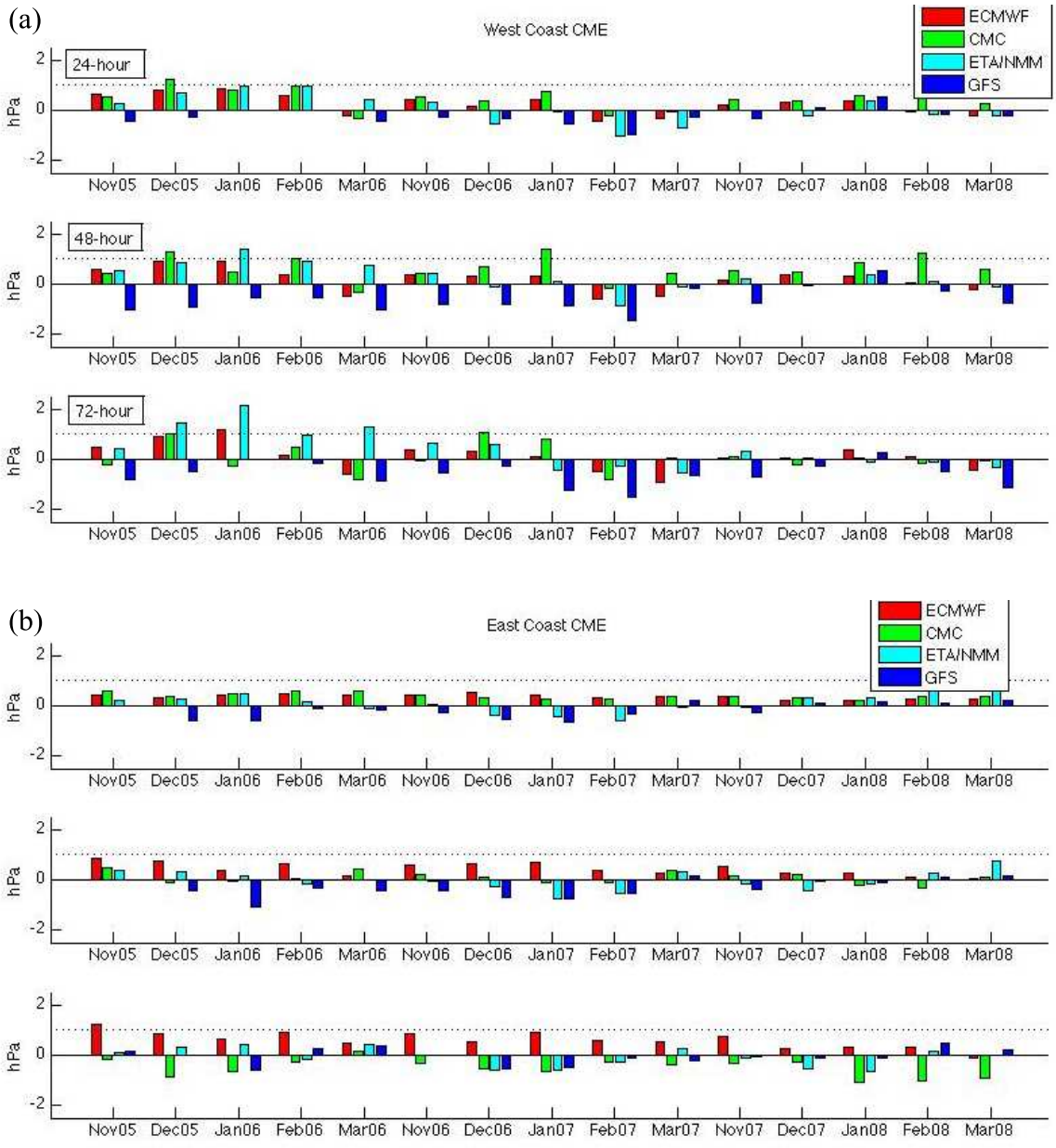


Figure 4.6 (a) West Coast monthly CME (in hPa) for forecasts with 24-h (top), 48-h (middle), and 72-h (bottom) lead-times for the ECMWF (red), CMC-GEM (green), NAM (cyan), and GFS (blue) models. (b) Same as (a) only for the East Coast.

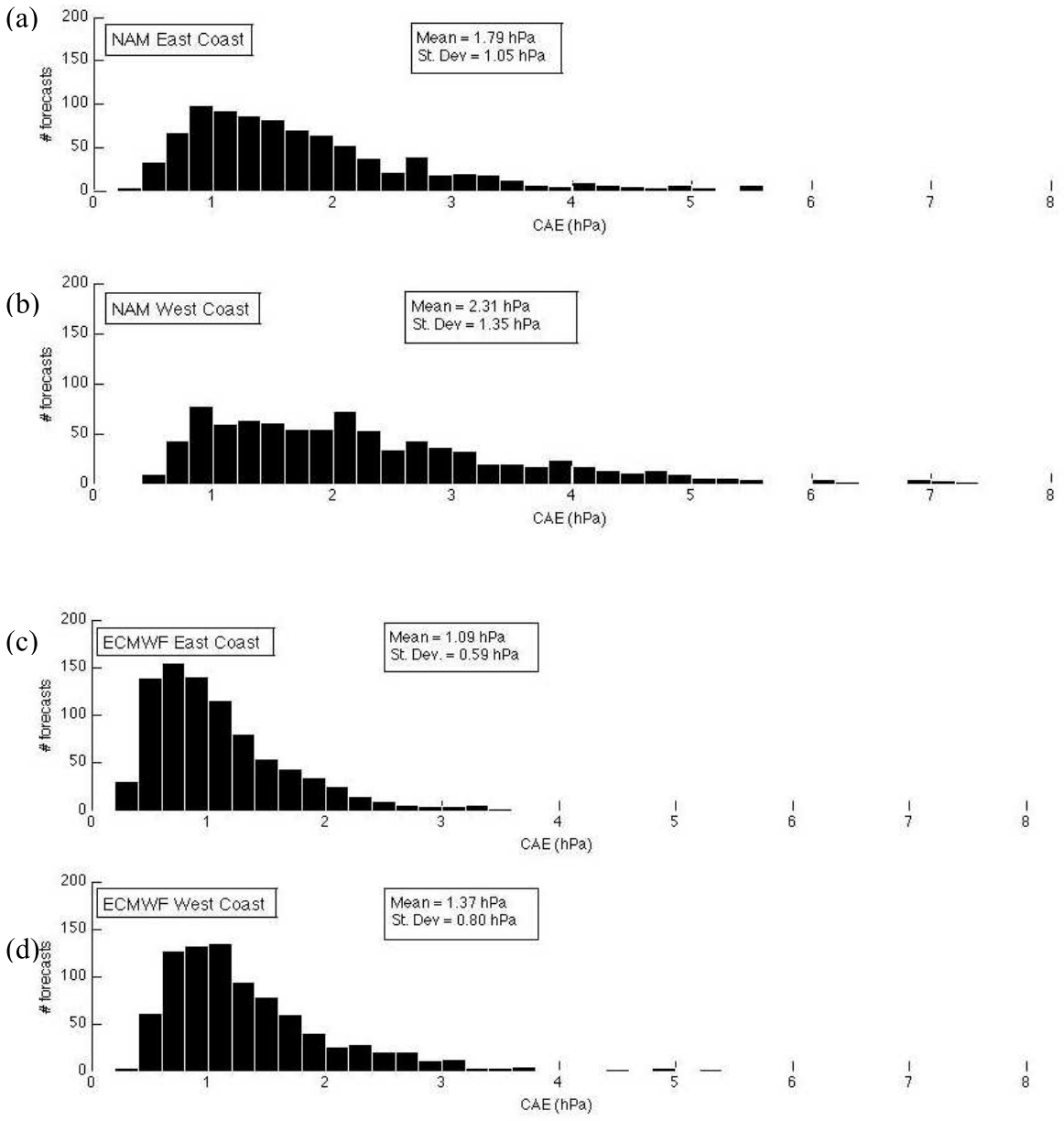


Figure 4.7 (a) Histogram of West Coast CAEs (in hPa) for the NAM model. The mean and standard deviation for each distribution is given on the figure. (b) Same as (a) only for the East Coast. (c) Same as (a) only for the ECMWF model. (d) Same as (b) only for the ECMWF model.

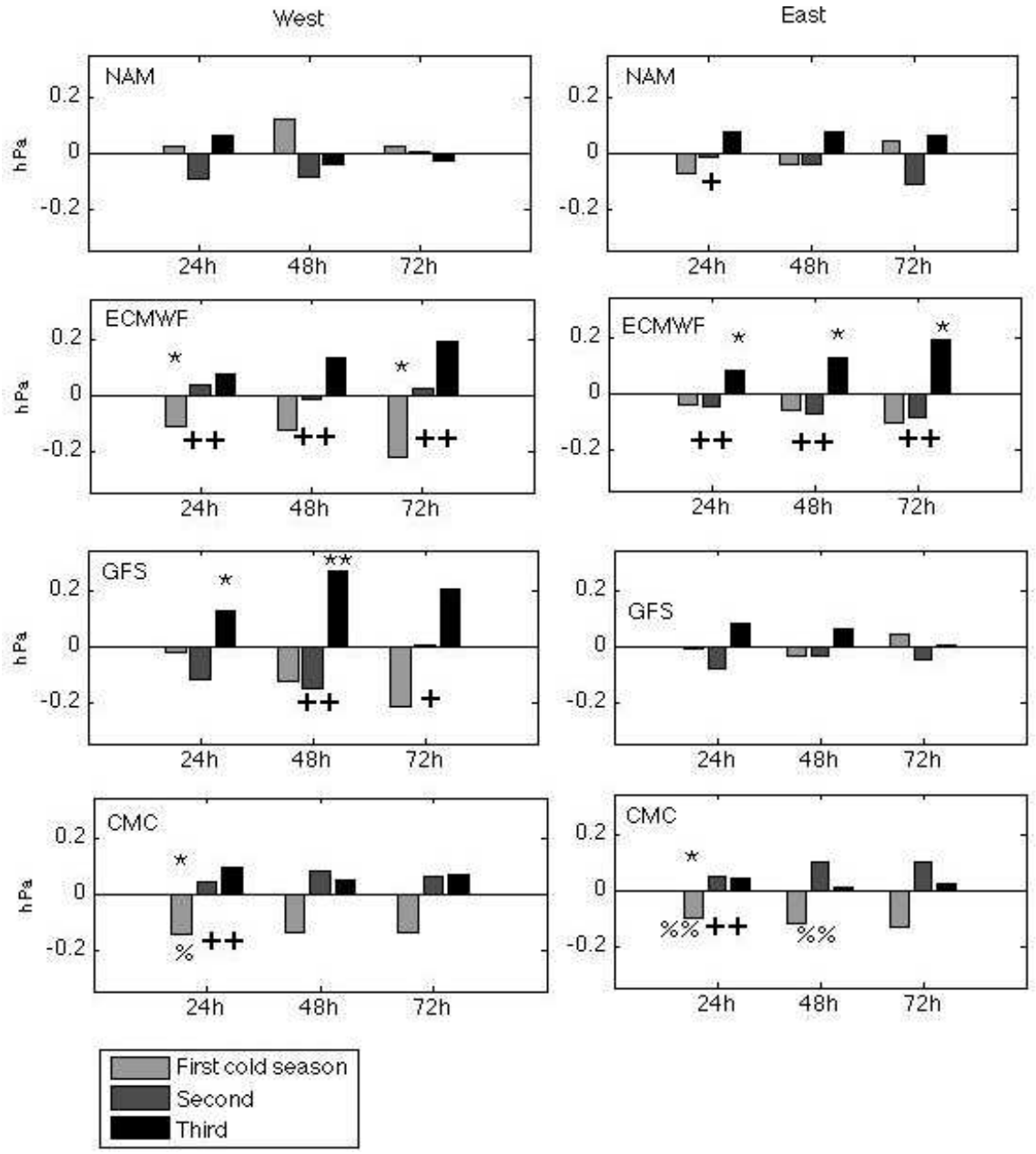


Figure 4.8 Anomaly coastal absolute error (CAE) for each model and forecast lead-time. Anomaly CAE is defined as the three-cold-season CAE minus the individual cold-season CAE. Positive values indicate the particular season had a smaller (more accurate) CAE. The stars indicate that a particular cold-season anomaly CAE was significantly different from the 3-season mean CAE at the 95% (**) or the 90% (*) level. The plus signs indicate that the 3rd season anomaly CAE was more accurate than the 1st season anomaly CAE at the 95% (++) or the 90% (+) level. The percent signs indicate that the 2nd season anomaly CAE was more accurate than the 1st season anomaly CAE at the 95% (%%) or the 90% (%) level (test performed for the NAM and CMC models only).

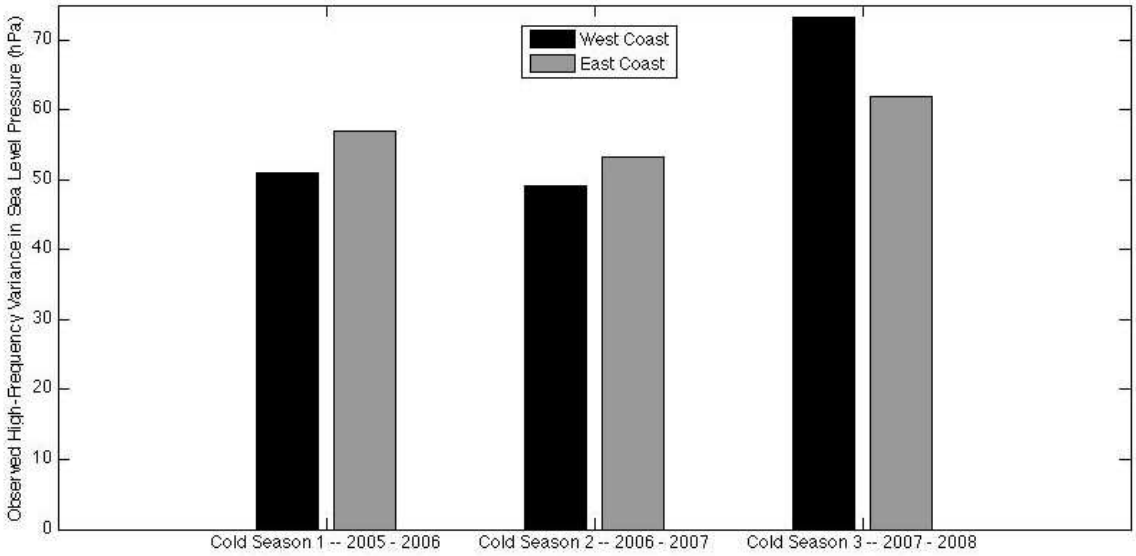


Figure 4.9 Observed 30-day high-pass sea level pressure variance [hPa^2] averaged over 11 stations on the West and East Coast shown by 5-month cold season.

TABLE 4.1: Model specifications as of 04/01/2008

Parameter	ECMWF	CMC-GEM	GFS	NAM
Domain	Global	Global	Global	Regional
Model Type	Spectral	Grid point	Spectral	Grid point
Horizontal Resolution, native	T799 (~25km)	33 km	T382 (~40km)	12 km
Horizontal Available for verification	1 deg (~111km)	1 deg	0.7 deg (~80km)	0.7 deg
Vertical levels	91	58	64	60
Data Assimilation System	4D-VAR	4D-VAR	3D-VAR	3D-VAR
Data Cutoff Time	12 hrs	6 hrs	2 hr 45 m	1 hr 15 m
Observation time window	12 hours	6 hours	6 hours	3 hours
Lateral boundary	none	none	none	6h old GFS

TABLE 4.2: Operational model changes from 11/2005 through 3/2008

Date	Model	Operational Change
12/2005	CMC-GEM	Assimilation of additional satellite data.
02/2006	ECMWF	Increased horizontal resolution to T799, increased vertical levels to 91 and raised top of model to 0.01 hPa. Ozone and grid-point humidity used in 4D-VAR and revised coefficients in ozone chemistry.
06/2006	NAM	Weather Research and Forecasting Nonhydrostatic Mesoscale Model (WRF-NMM) replaces Eta. Data assimilation switches from Eta 3D-Variational Analysis (EDAS) to Gridpoint Statistical Interpolation (GSI)
08/2006	GFS	New orography and land/sea mask, improved snow analysis and new ozone physics. Fixed error in downward longwave radiation scheme at earth's surface.
09/2006	ECMWF	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.
10/2006	CMC-GEM	Increase in horizontal and vertical resolution, improved physical parameterization, condensation and precipitation package and new surface scheme implemented.
12/2006	ECMWF	Introduced new satellite data including winds from MTSAT and GPS radio occultation data from CHAMP, GRACE and COSMIC.
12/2006	NAM	Improvements made to the convective parameterization and cloud microphysics.
5/2007	GFS	Unify the NCEP 3DVAR assimilation system under the GFS, change vertical coordinate to hybrid sigma-pressure, add new observing systems.
6/2007	ECMWF	Improved moist linear physics, revised subgrid-orography scheme, new short-wave radiation scheme and other changes.
11/2007	ECMWF	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. Introduction of four new pressure levels in the lower troposphere.

TABLE 4.3: List of Stations

West Coast		**East Coast**	
46088	48.3°N 123.2°W	44007	43.5°N 70.1°W
46131	49.9°N 125.0°W	44013	42.4°N 70.7°W
46206	48.8°N 126.0°W	BUZM3	41.1°N 71.0°W
TTIW1	48.4°N 124.7°W	44018	41.3°N 69.3°W
SISW1	48.3°N 122.8°W	44009	38.5°N 74.7°W
NWPO3	44.6°N 124.1°W	DUCN7	36.1°N 75.8°W
CARO3	43.3°N 124.4°W	41025	35.2°N 75.3°W
46022	40.7°N 124.5°W	FBIS1	32.6°N 79.9°W
46014	39.2°N 124.0°W	41008	36.4°N 80.9°W
46042	36.8°N 122.4°W	42039	28.8°N 84.8°W
46028	35.7°N 121.9°W	41009	28.5°N 80.2°W

5. SENSITIVITY TO INITIAL CONDITIONS

5.1 INTRODUCTION

Large forecast errors can be a result of initialization errors or insufficient "realism" of the model. Initializations can have large errors that lead to large forecast errors, or they can have small errors that grow rapidly into large errors in regions of high sensitivity. Knowing the most common causes of errors that lead to high-impact forecast failures can help determine how future resources can be best invested. Sensitivity is defined as the degree to which a change in one parameter's (such as sea level pressure's) initial conditions at a specific location effects a forecast parameter in a specified forecast region. The sensitivity of a model's forecast in one region to the accuracy of the initial conditions in another can be defined using an ensemble kalman filter (EnKF). The University of Washington's EnKF (Torn and Hakim 2007), consists of 90 ensemble members, all containing different but equally likely analyses. Regions where small changes in initial conditions result in significant changes in the forecast region are referred to in this paper as regions with *high sensitivity*. Some model forecasts have broad regions of high sensitivity, while other forecasts have comparatively very little sensitivity. Figure 5.1 shows a plot from a forecast verifying at 12 UTC 14 December 2006 made by the Weather Research and Forecasting Model that exhibits high sensitivity. A sensitivity value of 0.5 hPa means that a 1 hPa change in sea level pressure at that location will result in a 0.5 hPa change in sea level pressure in the boxed region shown in Fig. 5.1. In this figure, absolute sensitivity values exceed 1.5 hPa over broad regions. In contrast, Fig. 5.2 shows a forecast with must less sensitivity: there is a maximum value of approximately 0.5 hPa and it is limited to a small region. This chapter examines potential relationships between forecasts with different sensitivity values and forecasts that result in multiple models meeting a large-error criterion. It is arbitrary that only sensitivity of sea level pressure analysis to sea level pressure forecasts are

examined, as a study of other parameters may just as well reveal meaningful relationships.

5.2 METHODS

Forecasts where multiple models meet a large error criterion ("forecast busts") are compared to forecasts where the University of Washington EnKF exhibited particularly high model sensitivity to initial conditions to see if instances of forecast busts are more common for forecasts with high model sensitivity. Twice-daily 24-hour forecasts from January – March and November – December 2006 are included in the study, for a total of about 300 forecasts. In order to objectively rank forecasts by forecast sensitivity, domain-averaged sensitivity was calculated for each forecast. Figure 5.3 shows the distribution of the domain-averaged sensitivity values for about 300 forecasts, with a mean sensitivity value of 0.185 hPa/hPa. Results from the study in Chapter 4 are used to obtain the set of dates where forecast busts occurred: the criterion is that at least two models out of the GFS, the NAM, the ECMWF and the CMC-GEM had sea level pressure errors equal to or larger than a large error criterion at at least one of two stations within the forecast box shown in Figs. 5.1 and 5.2. The stations used for this study are C-MAN station desw1 (Destruction Island, WA) and buoy station 46088 (New Dungeness, WA). Instances of large errors are also compared to dates of forecasts with particularly low values of sensitivity – those that do not exceed a sensitivity "cap" – to see if instances of forecast busts are rare at times of low sensitivity. The effects of varying the definition of a large error and varying the threshold of "high sensitivity" are explored.

5.3 RESULTS AND DISCUSSION

When the threshold for high sensitivity is 0.2 hPa/hPa, the probability of a 3-hPa forecast bust occurring increases by a factor of 1.8 – or is 80% more likely compared to the chance that a random forecast becomes a forecast bust. When the

threshold for sensitivity is incrementally increased from 0.20 to 0.33 hPa/hPa (the latter is not shown), this factor increases up to 3.5, clearly indicating a relationship between forecast sensitivity and the occurrence of large forecast errors. Figure 5.4 shows that the value of this factor is greater for higher sensitivity thresholds (different thresholds are shown in different colors). This figure also shows how the relationship is stronger for more stringent "large error" criteria: for a sensitivity threshold of 0.2 hPa/hPa, the likelihood of a forecast bust of 2.5 hPa (or greater) occurring is greater by a factor of 1.6 over a random forecast, but the likelihood of a forecast bust of 4.5 hPa increases to a factor of 3.1. To put the "factors" in perspective, consider that about 20% of all forecasts reach the "forecast bust" threshold of 3.0 hPa. For a sensitivity threshold of 0.28 hPa/hPa, the likelihood of such a bust increases by a factor of 3.6. This means that over 70% of the cases with such high sensitivity result in errors over 3.0 hPa. Increasing the thresholds to require very high sensitivity or forecast bust criteria result in too few cases to draw valuable conclusions – for example, only 6 forecasts had forecast busts of 5 hPa or greater – but the trend observed in Fig. 5.4 suggests that the factor would continue to increase with increasing thresholds. Indeed, all 6 of the aforementioned forecast busts occurred with sensitivity greater than 0.28.

When a sensitivity cap of 0.185 hPa/hPa – the mean sensitivity of all forecasts – is applied, the likelihood of such a forecast qualifying for a 2.5 hPa forecast bust is 0.80 (or 80%) of the likelihood that a random forecast qualifies (Fig. 5.5). As the forecast bust threshold is increased to 3.0 and 3.5 hPa, the likelihood is just 0.60 and 0.48, respectively. Not surprisingly, when only the least sensitive cases are examined as the sensitivity cap decreases to 0.15 and 0.13 hPa/hPa, the likelihood of such forecasts reaching forecast bust criteria decreases as low as 0.2 for a forecast bust threshold of 3.5 hPa.

A benefit of knowing that such a strong relationship exists between the presence of high (or low) model sensitivity to initial conditions and the likelihood of resulting large forecast errors is that sensitivity can be calculated when the forecast is made, so it can function as a means of forecasting forecast skill or uncertainty. These

results show that the presence of anomalously high sensitivity to sea level pressure analysis bodes ill for 24-hour model sea level pressure forecast confidence, with perhaps up to a five-fold increase in the likelihood – resulting in upwards of 70% chance – that multiple model forecasts will verify with large errors. On the other hand, instances of anomalously low sensitivity can significantly boost confidence in model forecast skill, with the likelihood that a large error in sea level pressure occurs reduced to just a few percent.

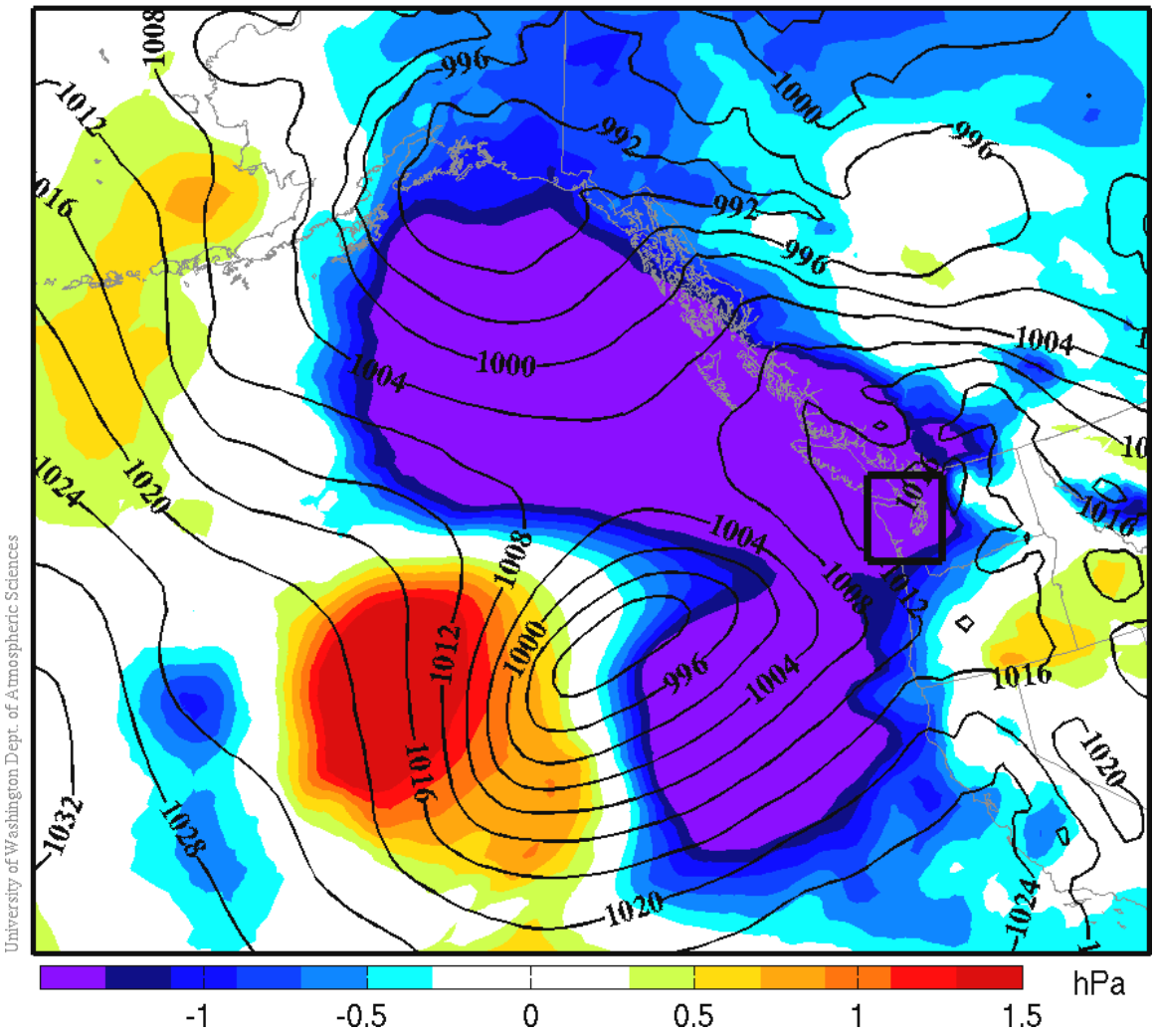


Figure 5.1 Weather Research and Forecasting Model sensitivity of the 24-h forecast for the boxed region verifying at 12 UTC 14 December 2006 to the initial conditions. Sensitivity values are defined as the result that a sea level pressure change of 1-hPa in the analysis has on the forecast sea level pressure for the boxed region. Forecast sea level pressure (black contours) is shown. Graphic courtesy of Ryan Torn and the University of Washington Ensemble Kalman Filter System.

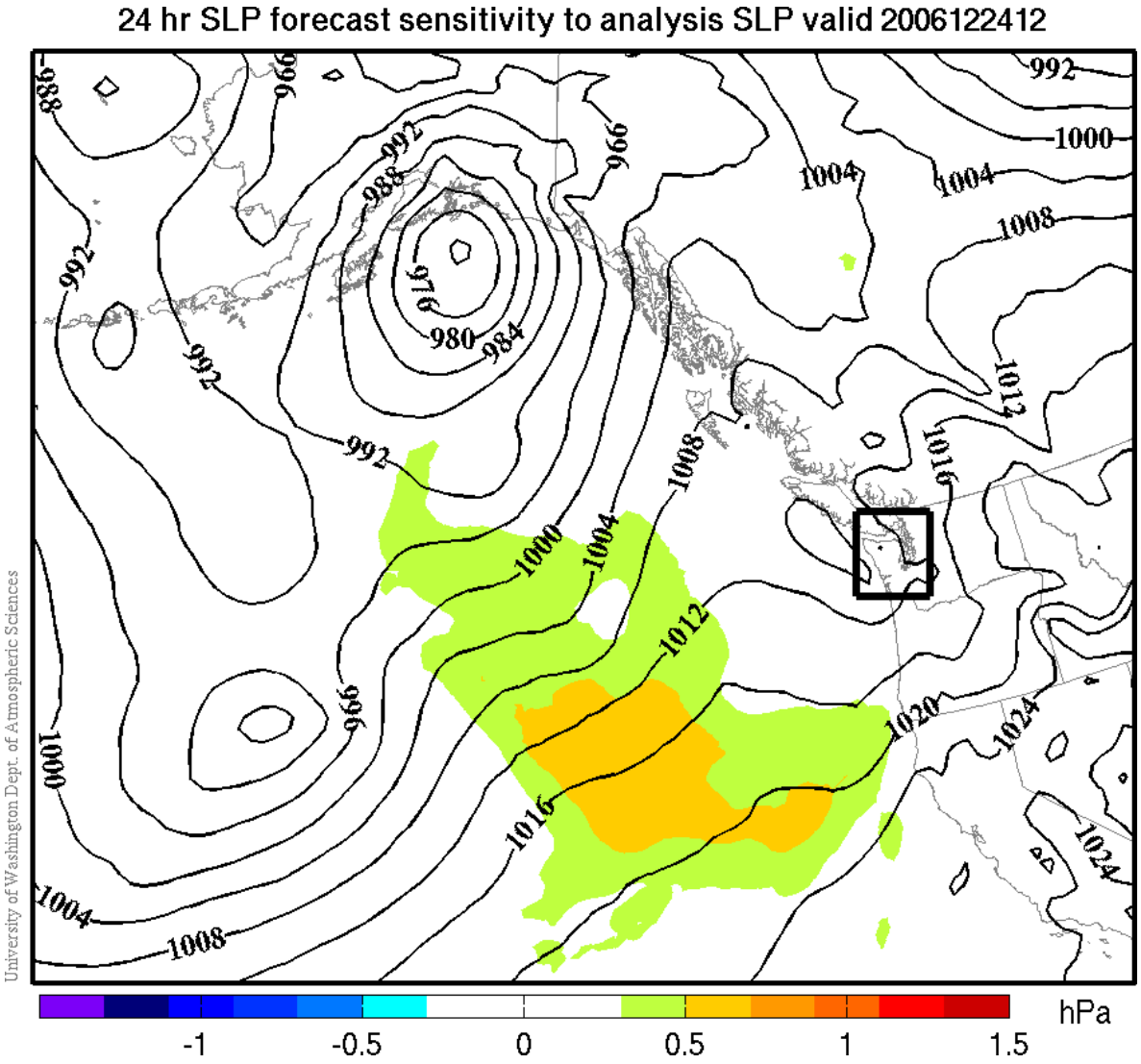


Figure 5.2 Weather Research and Forecasting Model sensitivity of the 24-h forecast for the boxed region verifying at 12 UTC 14 December 2006 to the initial conditions. Sensitivity values are defined as the result that a sea level pressure change of 1-hPa in the analysis has on the forecast sea level pressure for the boxed region. Forecast sea level pressure (black contours) is shown. Graphic courtesy of Ryan Torn and the University of Washington Ensemble Kalman Filter System.

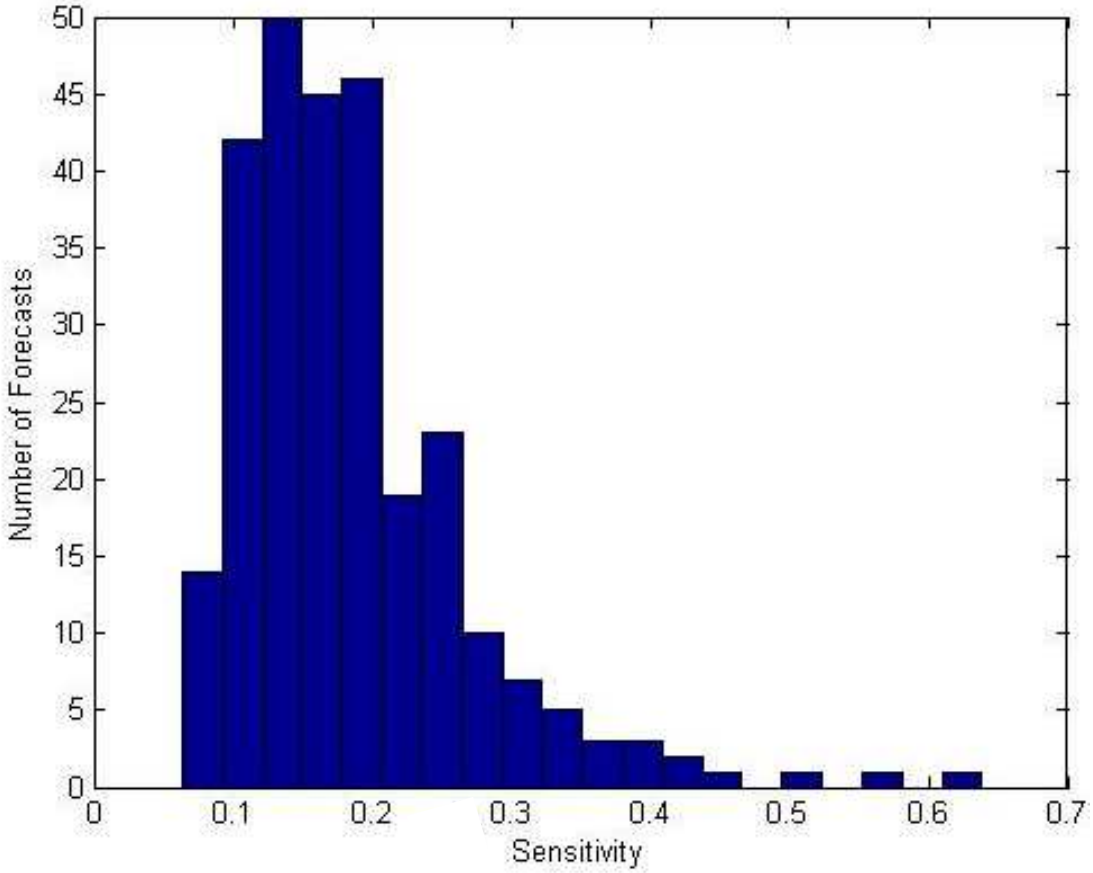


Figure 5.3 Histogram of the range of domain-averaged sensitivity (hPa / hPa).

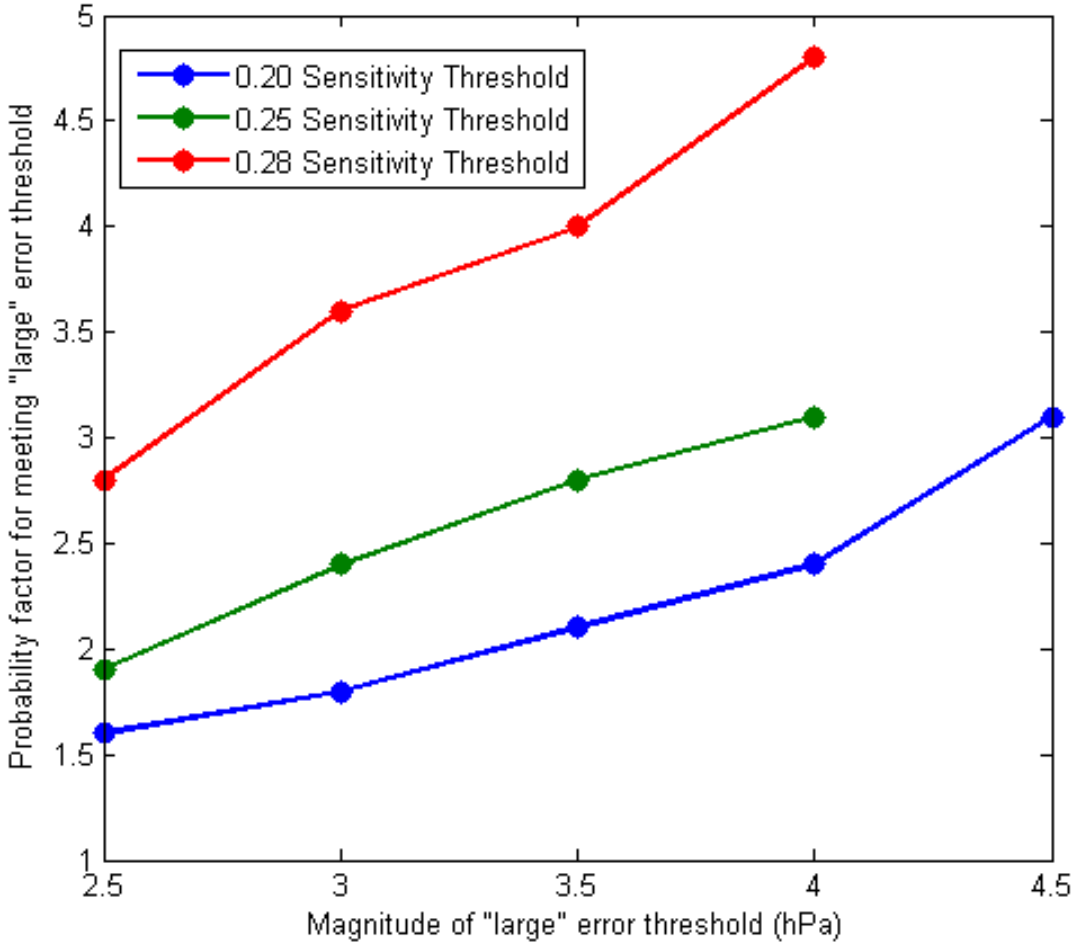


Figure 5.4 Probability factor for reaching the "large" error threshold as a function of that threshold for forecasts that meet a sensitivity threshold of 0.20 hPa/hPa (blue), 0.25 hPa/hPa (green) and 0.28 hPa/hPa (red). A factor of 1.0 indicates that the probability that a forecast meeting the sensitivity threshold will also meet the large error threshold is equal to the probability that a random forecast will meet the large error threshold.

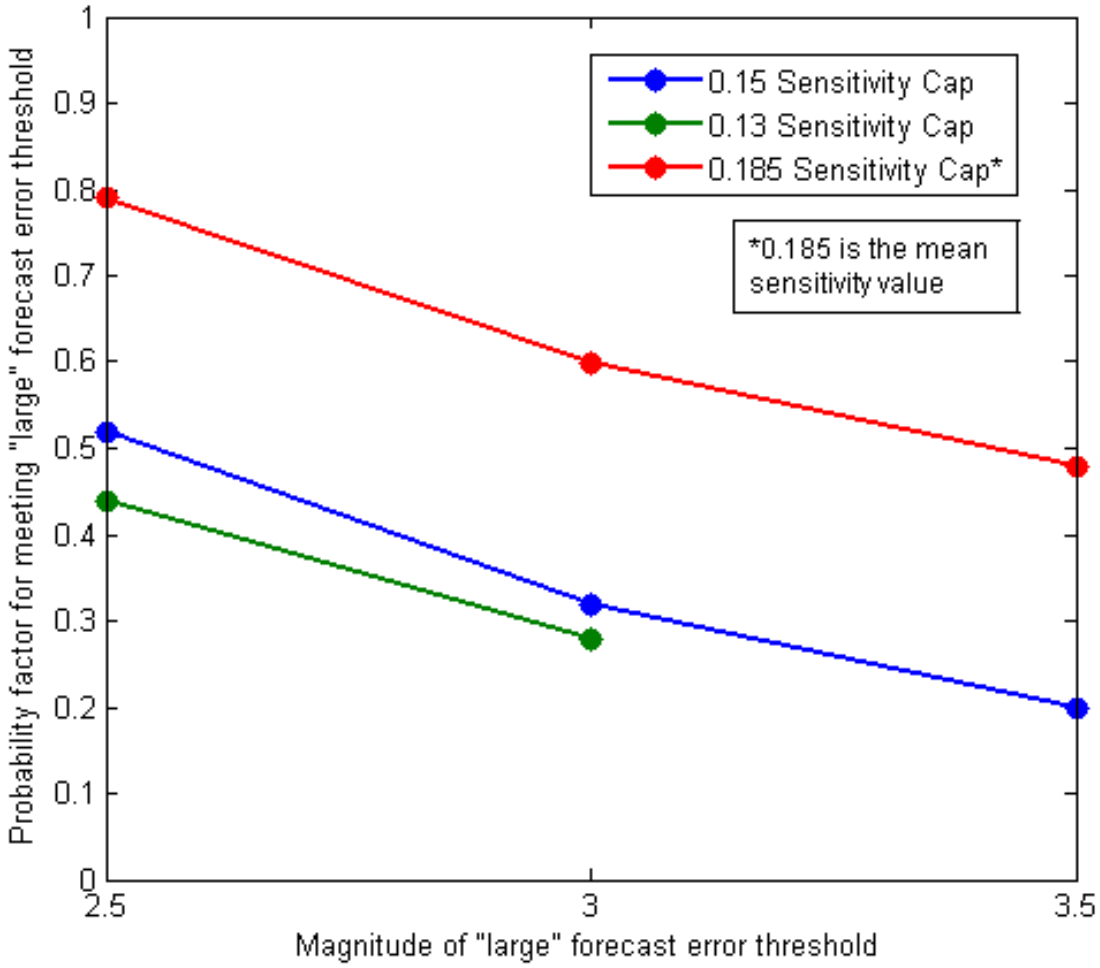


Figure 5.5 Probability factor for reaching the "large" error threshold as a function of that threshold for forecasts that fall under a sensitivity cap of 0.13 hPa/hPa (green), 0.15 hPa/hPa (blue) and 0.185 hPa/hPa (red). A factor of 1.0 indicates that the probability that a forecast meeting the sensitivity threshold will also meet the large error threshold is equal to the probability that a random forecast will meet the large error threshold.

6. SUMMARY

This study examines differences in observed sea level pressure characteristics in order to document model forecast errors in sea level pressure at observation sites along the East and West Coasts of the U.S. for the five-month cold seasons of 2005-2006, 2006-2007 and 2007-2008 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. Relationships between large forecast errors among multiple models and model sensitivity to initial conditions are investigated. Major findings include:

- A greater contribution of sea level pressure variance on the East Coast is from higher frequency structures, and a larger portion of the West Coast variance is due to seasonal and interannual variability.
- 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 experienced improvement over the study period for 24-h forecasts on both coasts.
- Large forecast errors are much more likely for forecasts with high sensitivity to initial conditions and much less likely for forecasts with low sensitivity. This relationship is stronger for stricter thresholds.
- A benefit of knowing such a strong relationship exists between the existence of high (or low) sensitivity and the occurrence (or rarity) of large forecast

errors is that sensitivity can be calculated when the forecast is made. It can therefore function as a means of forecasting forecast skill or uncertainty.

These results provide forecasters and model developers with specific information on the ability of four operational numerical models to forecast sea level pressure as well as a tool for estimating forecast quality using model sensitivity. 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 these 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.

REFERENCES

Anthes R. A., 1986: The general question of predictability. *Mesoscale Meteorology and Forecasting,* P. S. Ray, Ed., Amer. Meteor. Soc., 636–656.

Anthes R. A., Y.-H. Kuo, and J. R. Gyakum, 1983: Numerical simulations of a case of explosive cyclogenesis. *Mon. Wea. Rev.*, **111**, 1174–1188.

Bengtsson, L. and A. J. Simmons, 1983: Medium range weather prediction – operational experience at ECMWF. In *Large Scale Dynamical Process in the Atmosphere*. Eds. B. J. Hoskins and R. D. Pearce. Academic Press, London, U.K., 337 – 363.

Bjerknes, V., 1904: Das Problem von der Wettervorhersage, betrachtet vom Standpunkt der Mechanik und der Physik. *Meteor. Z.,* **21,** 1–7.

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

Black, T., M. Pyle, H.-Y. Huang, 2005: The Operational WRF-NMM at NCEP. AMS 21st Conference on Weather Analysis and Forecasting/17th Numerical Weather Prediction, 21, [np]. American Meteorological Society, 45 Beacon Street, Boston, MA 02108. [Available at: http//www.ams.confex.com/ams/htsearch.cgi].

Blackmon M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33**, 1607–1623.

Blackman, R. B. and Tukey, J. W. "Particular Pairs of Windows." *The Measurement of Power Spectra, From the Point of View of Communications Engineering.* New York: Dover, 1959, pp. 98-99.

Blackmon M. L., J. M. Wallace, N.-C. Lau, and S. L. Mullen, 1977: An observational study of the Northern Hemisphere wintertime circulation. *J. Atmos. Sci.*, **34**, 1040–1053.

Colle, B. A., C. F. Mass, and D. Ovens, 2001: Evaluation of the timing and strength of MM5 and Eta surface trough passages over the Eastern Pacific. *Wea. Forecasting*, **16**, 553-572.

Colle, B. A., and Charles, 2008: Verification of extratropical cyclones using an automated tracking algorithm, part I: short-term NAM and GFS forecasts. To be published in *Wea. Forecasting*.

Cote, J., S. Gravel, A. Methot, A. Patoine, M. Roch and A. Staniforth, 1998: The operational CMC-MRB global environmental multiscale (GEM) model. Part1: Design considerations and formulation. *Mon. Wea. Rev*., **126**, 1373-1396.

Frank W. M., 1983: The cumulus parameterization problem. *Mon. Wea. Rev.*, **111**, 1859–1871.

Grell G. A., Y. Kuo, and R. J. Pasch, 1991: Semiprognostic tests of cumulus parameterization schemes in the middle latitudes. *Mon. Wea. Rev.*, **119**, 5–31.

Guillemin E. A., 1957: *Synthesis of Passive Networks*, John Wiley, New York, 571 pp.

Gyakum J. R., 1983a: On the evolution of the *QE II* storm. I: synoptic aspects. *Mon. Wea. Rev.*, **111**, 1137–1155.

Gyakum J. R., 1983b: On the evolution of the *QE II* storm. II: Dynamic and thermodynamic structure. *Mon. Wea. Rev.*, **111**, 1156–1173.

Harper, K., L. W. Uccellini, E. Kalnay, K. Carey, and L. Morone, 2007: 50th anniversary of numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **88**, 639-650.

Kalnay, E., M. Kanamitsu, and W. E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Amer. Meteor. Soc*., **71**, 1410 – 1428.

Kalnay, E., S. J. Lord, and R.D. McPherson, 1998: Maturity of operational numerical weather prediction: medium range. *Bull. Amer. Meteor. Soc.*, **79**, 2753-2769.

Kuo Y.-H., and R. J. Reed, 1988: Numerical simulation of an explosively deepening cyclone in the eastern Pacific. *Mon. Wea. Rev.*, **116**, 2081–2105.

Lorenz E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

McMurdie, L., and C. Mass, 2004: Major numerical forecast failures over the Northeast Pacific. *Wea. Forecasting*, **19**, 338-356.

Murakami M., 1979: Large-scale aspects of deep convective activity over the GATE area. *Mon. Weather Rev*, **107**, 994–1013.

National Weather Service, 2001: Winter storms, the deceptive killers. [Available at: www.nws.noaa.gov/om/winterstorm/winterstorms.pdf.]

Parsons, D., I. Szunyogh and P. Harr, 2007: Scientific Program Overview THORPEX Pacific-Asian Regional Campaign (T-PARC). [Available at: www.ucar.edu/na-thorpex/tparc/SPO_PARC_revised.pdf.]

Pauley, P. M, 1998: An Example of uncertainty in sea level pressure reduction. *Wea. Forecasting*, **13**, 833-850.

Reed R. J., and M. D. Albright, 1986: A case study of explosive cyclogenesis in the eastern Pacific. *Mon. Wea. Rev.*, **114**, 2297–2319.

Richardson, L. F., 1922: *Weather Prediction by Numerical Process.* Cambridge University Press, 236 pp..

Roebber, P. J., 1984: Statistical analysis and updated climatology of explosive cyclones. *Mon. Wea. Rev.,* **112,** 1577–1589.

Sanders F., and J. R. Gyakum, 1980: Synoptic–dynamic climatology of the "bomb". *Mon. Wea. Rev.*, **108**, 1589–1606.

Sawyer, J. S., 1970: Observational characteristics of atmospheric fluctuations with a time scale of a month. *Quart. J. Roy. Meteor. Soc.*, **96**, 610-625.

Shapiro, M. A., and A. J. Thorpe, 2004: THORPEX International Science Plan, Version 3, 2 November 2004 WMO/TD No. 1246 WWRP/THORPEX No. 2. [Available at: www.wmo.int/pges/prog/arep/documents/CD_ROM_international_science_plan_v3_002.pdf].

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart J. of the Roy. Meteorol. Soc.*, **128**, 647-677.

Shanks J. L., 1967: Recursion filters for digital processing. *Geophysica*, **32**, 33–51.

Sinclair, M. R., J. A. Renwick, and J. W. Kidson, 1997: Low-frequency variability of Southern Hemisphere sea level pressure and weather system activity. *Mon. Wea. Rev.,* **127,** 2531–2542.

Torn, R. D., G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble kalman filter. *Mon. Wea. Rev.,* **136**, 3947–3963.

Uccellini L. W., 1986: The possible influence of upstream upper- level baroclinic processes on the development of the *QEII* storm. *Mon. Wea. Rev.*, **114**, 1019–1027.

Wedam G. B., L. A. McMurdie and C. F. Mass, 2008: Comparison of model forecast skill of sea-level pressure along the East and West Coasts of the United States. To be published in *Wea. Forecas*ting.

Zishka K. M., and P. J. Smith, 1980: The climatology of cyclones and anticyclones over North American and surrounding ocean environs for January and July, 1950–77. *Mon. Wea. Rev.*, **108**, 387–401.