Strengths and Weaknesses of MOS, Running-Mean Bias Removal, and Kalman Filter Techniques for Improving Model Forecasts over the Western United States

WILLIAM Y. Y. CHENG AND W. JAMES STEENBURGH

Department of Meteorology, University of Utah, Salt Lake City, Utah

(Manuscript received 18 September 2006, in final form 5 March 2007)

ABSTRACT

Despite improvements in numerical weather prediction, model errors, particularly near the surface, are unavoidable due to imperfect model physics, initial conditions, and boundary conditions. Here, three techniques for improving the accuracy of 2-m temperature, 2-m dewpoint, and 10-m wind forecasts by the Eta/North American Meso (NAM) Model are evaluated: (i) traditional model output statistics (ETAMOS), requiring a relatively long training period; (ii) the Kalman filter (ETAKF), requiring a relatively short initial training period (~4–5 days); and (iii) 7-day running mean bias removal (ETA7DBR), requiring a 7-day training period. Forecasts based on the ETAKF and ETA7DBR methods were produced for more than 2000 MesoWest observing sites in the western United States. However, the evaluation presented in this study was based on subjective forecaster assessments and objective verification at 145 ETAMOS stations during summer 2004 and winter 2004/05. For the 145-site sample, ETAMOS produces the most accurate cumulative temperature, dewpoint, and wind speed and direction forecasts, followed by ETAKF and ETA7DBR, which have similar accuracy. Selected case studies illustrate that ETAMOS produces superior forecasts when model biases change dramatically, such as during large-scale pattern changes, but that ETAKF and ETA7DBR produce superior forecasts during quiescent cool season patterns when persistent valley and basin cold pools exist. During quiescent warm season patterns, the accuracy of all three methods is similar. Although the improved ETAKF cold pool forecasts are noteworthy, particularly since the Kalman filter can help better define cold pool structure by producing forecasts for locations without long-term records, alternative approaches are needed to improve forecasts during periods when model biases change dramatically.

1. Introduction

Numerical weather prediction (NWP) has undergone revolutionary development in the last 40 yr with advances in computational power, modeling techniques, and observations. Nevertheless, NWP model forecasts contain systematic biases due to imperfect model physics, initial conditions, and boundary conditions (Paegle et al. 1997; Mass et al. 2002; Hart et al. 2004). In addition, NWP models prognose fields at gridpoint locations that represent a volume average (Pielke 2002), yet meteorologists and end users are interested frequently in forecasts for specific locations. Unfortunately, there is no universally accepted method for downscaling model forecasts to specific locations, especially when the model elevation differs from that of the observing site. Even when the model resolution is increased, it does not necessarily improve model performance. For example, Mass et al. (2002) found limited improvement in the temperature forecasts by the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5) when they decreased the grid spacing from 12 to 4 km over the Pacific Northwest. They attributed the problem to "model deficiencies, inadequate mesoscale and synoptic initialization, and the inherent predictability limits of mesoscale flows." For these reasons, postprocessing techniques, such as model output statistics (MOS), have been developed to improve and provide location-specific forecasts from model guidance. MOS uses multiple linear regression to produce an improved forecast at specific locations by using model forecast variables and prior observations as predictors (Glahn and Lowry 1972). MOS remains a useful tool and during the 2002 Olympic Winter Games, MM5-based MOS

Corresponding author address: William Y. Y. Cheng, Dept. of Atmospheric Science, Colorado State University, Campus Delivery 1371, Fort Collins, CO 80523.
E-mail: cheng@atmos.colostate.edu

DOI: 10.1175/2007WAF2006084.1

outperformed the native forecasts produced by MM5 and was equally or more skillful than human-generated forecasts by the Olympic Forecast Team (Hart et al. 2004).

One major drawback of MOS is the need for a long training dataset. There are, however, other postprocessing methods that are much simpler to implement that do not require a long training dataset such as the Kalman filter and running-mean bias removal. For example, Roeger et al. (2003) applied the Kalman filter to adjust the surface forecast bias of the Canadian Mesoscale Compressible Community model (MC2; Benoit et al. 1997) at six stations in British Columbia. Kalman (1960) introduced the concept of the Kalman filter (KF) in his groundbreaking paper that described a recursive solution to a discretized linear filtering problem. Since Kalman's 1960 paper, the KF has found its way into many applications in science, engineering, and economics. The KF combines a model with observations to provide a better estimate of a state variable than either the model or observations can provide alone. This process is carried out iteratively. This procedure is reminiscent of data assimilation, and indeed the KF is used in atmospheric data assimilation (Houtekamer and Mitchell 2001). The technique of running-mean bias removal is quite simple and involves adjusting the model forecast based on the mean bias from forecasts in the recent past (e.g., in the past 7 days).

As far as the authors are aware, there has not been a previous study in which MOS, KF, and running-mean bias removal have been compared. Because MOS needs a long training dataset and works best with a frozen model, this paper examines how other easier to implement postprocessing methods, such as KF and running-mean bias removal, perform relative to MOS and under what conditions these alternative postprocessing methods produce better forecasts. Specifically, we evaluate three postprocessing methods for improving 2-m temperature, 2-m dewpoint, and 10-m wind speed and direction forecasts produced by the Eta/North American Meso (NAM)[1] model: (i) traditional MOS (ETAMOS), (ii) the Kalman filter (ETAKF), and (iii) a 7-day running mean bias removal (ETA7DBR). We focus on surface forecasts because of their importance to the general public, commercial activities, and forecasters. We will show that, overall, MOS produces superior forecasts. However, during quiescent large-scale patterns, particularly in the cool season when persistent cold pools develop in valleys and basins, ETAKF and ETA7DBR perform better. Further, in these persistent cold pool situations, ETAKF adjusts more rapidly than ETA7DBR, while MOS consistently exhibits a warm bias in the afternoon. In contrast, ETAKF and ETA7DBR perform poorly during periods when the Eta Model biases change abruptly.

[1] The Eta Model was renamed the NAM model in January 2005 (Dimego et al. 2005). Hereafter, we simply refer to it as the Eta Model.

2. Methodology

We evaluated the performance of the three techniques and the native Eta Model forecasts for the summer of 2003 (June–August) and the winter of 2004/05 (December–February). We considered the following variables in the evaluation: (i) 2-m temperature, (ii) 2-m dewpoint, and (iii) 10-m wind speed and direction. The surface observations required for verification and to produce the ETAKF and ETA7DBR forecasts were provided by the MesoWest cooperative networks, maintained by the Cooperative Institute for Regional Prediction at the University of Utah (Horel et al. 2002), and were assumed to be the "ground truth." ETAKF and ETA7DBR forecasts were produced for over 2000 MesoWest stations, although we selected 145 stations in the western United States for statistical analysis to facilitate comparison with ETAMOS. These 145 stations are mainly airport stations, thus biasing the sample toward valley locations.

Implemented in the Department of Meteorology at the University of Utah, the ETAKF forecasts were provided in real time to five National Weather Service Weather Forecast Offices (NWS WFOs) in the western and central regions as part of a Cooperative Program for Operational Meteorology, Education and Training (COMET) project. Forecaster feedback was used to select case studies that illustrate the strengths and weaknesses of the techniques. The appendix describes the ETAKF algorithm. ETA7DBR uses the 7-day running-mean difference between (i) the bilinearly interpolated Eta Model forecast and (ii) the MesoWest observation to adjust the Eta Model forecast. ETAMOS was obtained from the National Centers for Environmental Prediction (NCEP).

The bias (also known as, mean) error [BE; Wilks (2006), Eq. (7.30)] and mean absolute error [MAE; Wilks (2006), Eq. (7.27)] of the 2-m temperature, 2-m dewpoint, 10-m wind speed, and 10-m wind direction of the Eta Model, ETAKF, ETA7DBR, and ETAMOS were evaluated on a *3-hourly basis* from forecast hours 6 to 48. Wind speed and direction observation–forecast pairs were not considered for observed wind speeds of 2.5 m s^{-1} or less due to instrument inaccuracies at low wind speeds (Cheng and Steenburgh 2005). Only the MAE was evaluated for the 10-m wind direction.

For brevity and ease of presentation, BE and MAE from the two forecast cycles (0000 and 1200 UTC) were combined by weight averaging. For a given verification time, the weight-averaged value, $\bar{z}$, is given by

$$\bar{z} = (N_{00}\bar{z}_{00} + N_{12}\bar{z}_{12})/(N_{00} + N_{12}),$$

where $\bar{z}_{00}$ and $\bar{z}_{12}$ are the average values of z, and N_{00} and N_{12} are the numbers of forecast–observation pairs from the 0000 and 1200 UTC cycles, respectively. The BE and MAE time series begin at 0600 UTC (hour 6 of the 0000 UTC cycle) and end 54 h later at 1200 UTC (hour 48 of the 1200 UTC cycle). No weight averaging is applied during the first and last 12 h of these time series since the model runs do not overlap.

3. Results

a. 2-m temperature

For the summer season, ETAMOS had the lowest cumulative 2-m temperature MAE of 1.5°C, compared with 2.6°C for the Eta Model (Table 1). ETA7DBR and ETAKF had MAEs of 1.8° and 1.9°C, respectively, illustrating that these techniques improved upon the Eta Model, but not as much as ETAMOS. As a function of time, the summer season 2-m temperature MAE of the Eta Model exhibited a pronounced diurnal variation, with maxima around 2.8°–3°C in early morning (1200 UTC; Fig. 1). Cheng and Steenburgh (2005) report similar Eta Model performance over the western United States in the summer of 2003. All of the postprocessing methods also showed similar diurnal variations but with smaller MAEs. The ETAKF and ETA7DBR MAEs were 0.6°–1°C lower than those of the Eta Model, while ETAMOS was 0.9°–1.0°C lower.

The cumulative Eta 2-m temperature MAE during the winter was identical to that for the summer (2.6°C), but the improvements produced by the postprocessing

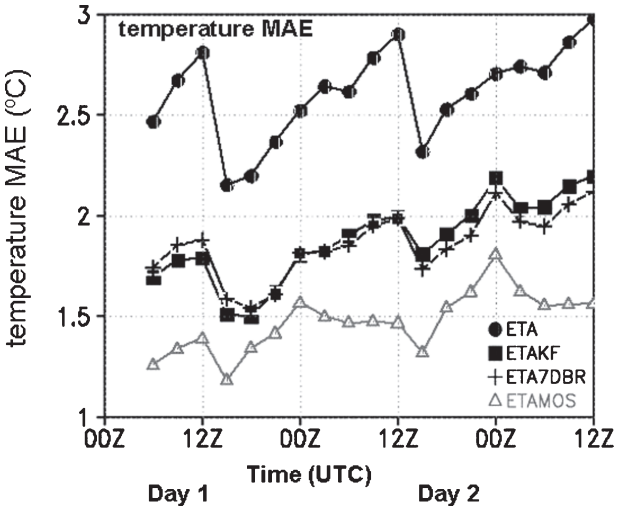

FIG. 1. Average temperature MAE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004 for Eta (solid circle), ETAKF (solid square), ETA 7DBR (plus sign), and ETAMOS (triangle).

techniques were smaller (Table 2). ETAMOS produced the best overall forecasts with a cumulative MAE of 2.0°C, 0.5°C higher than for summer, while the ETAKF and ETA7DBR cumulative 2-m temperature MAEs were both 2.4°C and only marginally better than that of the Eta Model. As a function of time, the Eta Model 2-m temperature MAE possessed a more complicated signature than its summer counterpart (Fig. 2). Unlike the summer season where the Eta Model temperature MAE was maximized in the early morning (1200 UTC), the winter temperature MAE was generally maximized at both 0000 and 1200 UTC. The ETAKF and ETA7DBR forecasts followed a similar trend, albeit with a small phase difference and with smaller MAEs. Except for a few times in the late afternoon (2100 and 0000 UTC), the ETAKF and ETA7DBR MAEs were

TABLE 1. Cumulative MAE and BE from Jun to Aug 2004.

	ETA	ETAKF	ETA7DBR	ETAMOS
Temperature MAE (°C)	2.6	1.9	1.8	1.5
Temperature BE (°C)	0.9	0.0	0.0	0.1
Dewpoint MAE (°C)	3.4	2.3	2.3	2.0
Dewpoint BE (°C)	−2.6	−0.1	−0.1	−0.4
Wind speed MAE ($m\ s^{-1}$)	1.8	1.6	1.5	1.4
Wind speed BE ($m\ s^{-1}$)	−0.7	−0.3	0.0	−0.1
Wind direction MAE (°)	53	48	53	39

TABLE 2. Cumulative MAE and BE from Dec 2004 to Feb 2005.

	ETA	ETAKF	ETA7DBR	ETAMOS
Temperature MAE (°C)	2.6	2.4	2.4	2.0
Temperature BE (°C)	0.0	0.1	0.1	0.5
Dewpoint MAE (°C)	2.3	2.2	2.1	1.8
Dewpoint BE (°C)	−0.8	−0.2	0.0	−0.1
Wind speed MAE ($m\ s^{-1}$)	1.9	1.7	1.7	1.6
Wind speed BE ($m\ s^{-1}$)	−0.4	−0.1	0.0	0.0
Wind direction MAE (°)	53	51	54	42

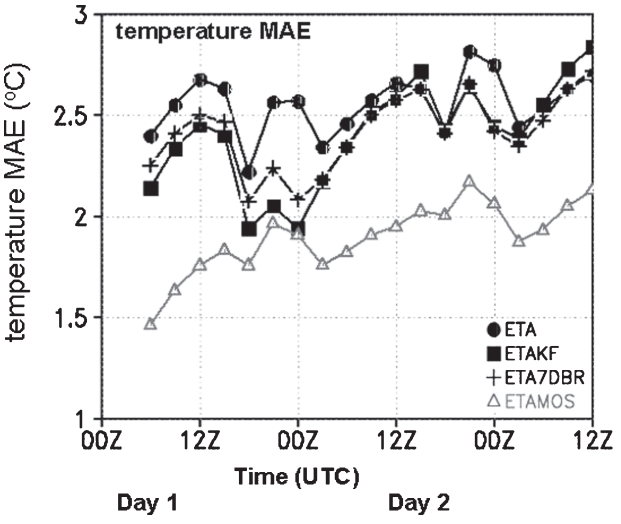

FIG. 2. Average temperature MAE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

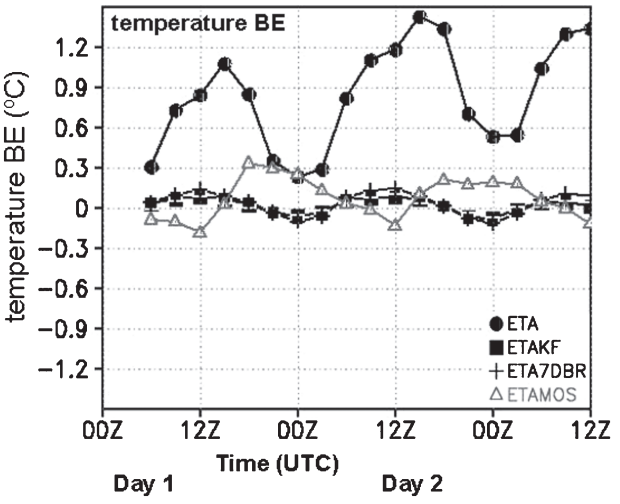

FIG. 3. Average temperature BE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

within 0.2°C of the Eta Model MAE. Although ETAMOS did not perform as well as in the summer, it generally performed better than ETAKF or ETA7DBR, with a lower MAE by 0.4°C or more at many times. Roeger et al. (2003) found that using the KF, the temperature MAE was reduced by 0.2°–0.7°C at one station on Blackcomb Mountain during the winter season, comparable to the results in this study.

For the summer, ETAKF reduced the cumulative Eta Model warm temperature bias from 0.9° to 0°C (Table 1). ETA7DBR also had negligible cumulative temperature BE. ETAMOS had a slight warm bias of 0.1°C. Examining the temperature BE as a function of time reveals that the Eta Model had a warm bias at all hours with maxima of 1°C or more in the morning hours (1500 UTC) and minima in the late afternoon (0000 UTC; Fig. 3). The ETAKF and ETA7DBR BEs were less than 0.15°C at all times. In general, the ETAMOS BE was larger but no more than 0.3°C in magnitude.

In contrast, the Eta Model had a negligible cumulative temperature BE during the winter (Table 2). ETAKF and ETA7DBR also had very small temperature BEs (both 0.1°C), whereas the ETAMOS BE was 0.5°C, larger than for the summer (0.1°C). As a function of time, the Eta Model temperature BE featured a warm bias overnight (reaching a maximum of 0.4°–0.6°C at 0900 UTC) and a cold bias during the day (reaching a minimum of −0.9°C at 0000 UTC), which reflect the Eta Model's inability to fully develop cold air pools overnight (Hart et al. 2004) and the convective boundary layer in late afternoon/early evening (Colle et al. 2003; see also Fig. 4). As for the postprocessing methods, ETAKF and ETA7DBR had negligible temperature BEs for all times, and ETAMOS had a higher cumulative temperature BE than ETAKF or ETA7DBR, with a predominant warm bias approaching 0.9°C in the late afternoon (0000 UTC).

We also evaluated the hit rate, or the total number of "correct" forecasts divided by the total number of events [also known as, probability of detection; Stephenson and Jolliffe (2003)] for accuracy thresholds of 1°, 2°, 3°, 4°, and 5°C (expressed as percent; Tables 3 and 4). Regardless of season or accuracy threshold, the rankings of the different postprocessing methods

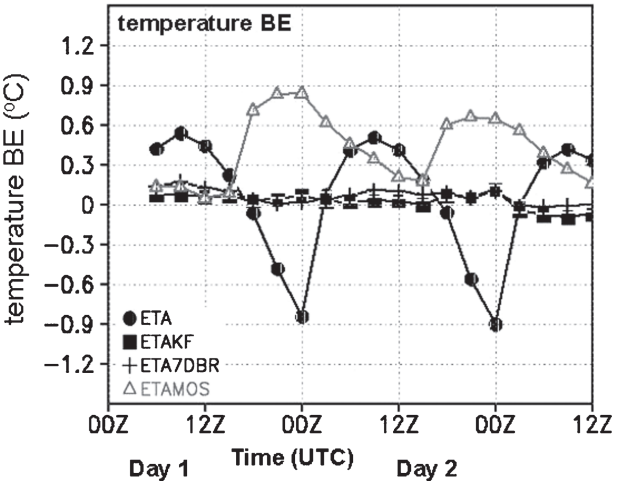

FIG. 4. Average temperature BE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

TABLE 3. Table showing the percentage of temperature forecasts achieving certain accuracy in terms of MAE for Jun–Aug 2004.

	ETA	ETAKF	ETA7DBR	ETAMOS
Percentage of temperature MAE < 1°C	26	37	37	42
Percentage of temperature MAE < 2°C	49	64	64	73
Percentage of temperature MAE < 3°C	67	81	81	89
Percentage of temperature MAE < 4°C	79	90	91	95
Percentage of temperature MAE < 5°C	87	95	96	98

remained the same as in the traditional MAE analysis, with ETAMOS producing the highest hit rate followed by ETAKF and ETA7DBR. For example, at the 3°C accuracy threshold, the ETAMOS hit rate was 89% compared with 81% for both ETAKF and ETA7DBR, and 67% for the Eta Model. During the winter, the Eta Model hit rate was similar to that in the summer for all thresholds; however, the accuracy of the postprocessed forecasts dropped. For example, the 3°C hit rate for ETAKF (ETA7DBR) was 72% (70%) in the winter versus 81% (81%) in the summer, while the ETAMOS hit rate was 80% in the winter versus 89% in the summer.

Another measure of the performance of the various postprocessing methods is to calculate the percentage of the stations that had their cumulative temperature MAE lowered as compared to Eta. For the summer of 2004, 99% of the stations benefited from ETAMOS, followed by 89% for ETAKF, and 84% for ETA7DBR. For the winter of 2004/05, 94% of stations benefited from ETAMOS, followed by 60% for ETA7DBR, and 57% for ETAKF. Thus, contrary to the hit-rate method, ETA7DBR is marginally better than ETAKF when using this method of evaluation.

TABLE 4. Table showing the percentage of temperature forecasts achieving certain accuracy in terms of MAE for Dec 2004–Feb 2005.

	ETA	ETAKF	ETA7DBR	ETAMOS
Percentage of temperature MAE < 1°C	27	30	29	32
Percentage of temperature MAE < 2°C	50	55	53	62
Percentage of temperature MAE < 3°C	67	72	70	80
Percentage of temperature MAE < 4°C	79	82	82	90
Percentage of temperature MAE < 5°C	88	89	89	94

b. 2-m dewpoint

The Eta Model cumulative 2-m dewpoint MAE for the summer was 3.4°C (Table 1). ETAKF and ETA7DBR each reduced the cumulative dewpoint MAE to 2.3°C, while ETAMOS produced the lowest cumulative dewpoint MAE of 2.0°C. As a function of time, the Eta Model dewpoint MAE varied diurnally with values ranging from 2.6° to 4.1°C and an overall increase in MAE with forecast projection (Fig. 5). Dewpoint MAE minima were found at 1800 UTC with maxima at 0300–0600 UTC. The postprocessed forecasts featured a similar overall structure, but with some phase differences. Nonetheless, ETAKF and ETA7DBR dewpoint MAEs were consistently 1°–1.5°C lower than those of the Eta Model. ETAMOS generally performed better than ETAKF or ETA7DBR with a 0.5°C advantage at some times.

The cumulative winter dewpoint MAE for the Eta Model was only 2.3°C, 1.1°C lower than for summer (Table 2). The ETAKF and ETA7DBR cumulative dewpoint MAEs were only 0.1° and 0.2°C lower than that of the Eta Model, respectively. However,

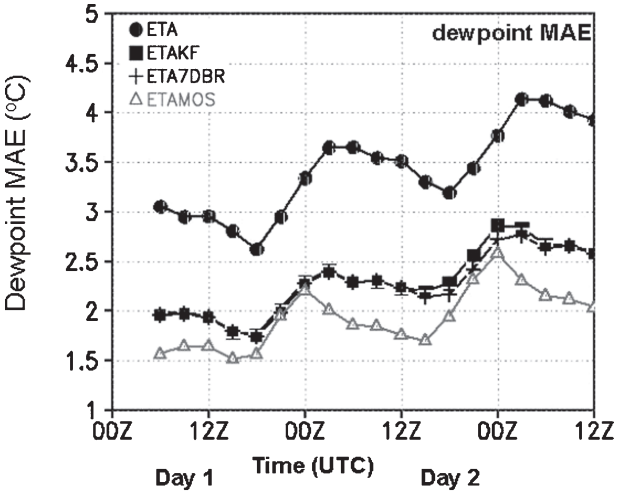

FIG. 5. Average dewpoint MAE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

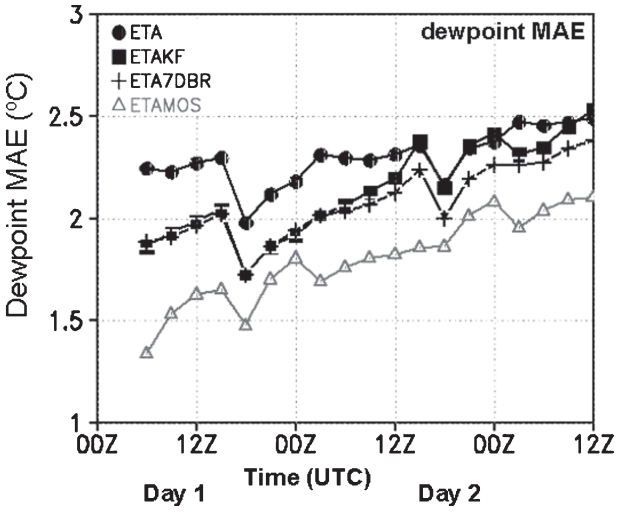

FIG. 6. Average dewpoint MAE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

ETAMOS outperformed ETAKF, ETA7DBR, and the Eta, with a cumulative MAE of 1.8°C. None of the postprocessing methods produced as much improvement as in the summer. As a function of time, growth of the Eta Model dewpoint MAE with increasing forecast projection is evident, although abrupt minima appear at 1800 UTC (Fig. 6). As expected from the cumulative MAE analysis, ETAKF and ETA7DBR only reduced the Eta dewpoint MAE by less than 0.3°C at all times and ETAMOS generally performed better.

The cumulative Eta Model dewpoint BE during the summer was −2.6°C (Table 1), indicating a dry bias similar to that found in the Eta Model during summer 2003 (Cheng and Steenburgh 2005). Both ETAKF and ETA7DBR reduced the cumulative dewpoint BE to negligible values, whereas the ETAMOS cumulative dewpoint BE was −0.4°C. As a function of time, the Eta Model dewpoint BE was below −2°C for most times with local maxima in negative BE between 0600 and 1200 UTC for all days (Fig. 7). ETAKF and ETA7DBR reduced the dewpoint BE to a magnitude of less than 0.3°C at all times. ETAMOS had a larger BE magnitude than ETAKF and ETA7DBR and was more than 0.5°C too dry for many times.

The Eta cumulative dewpoint BE was −0.8°C in the winter, much smaller in magnitude than in summer (Table 2). The ETAKF, ETA7DBR, and ETAMOS dry biases were less than 0.2°C in magnitude. As a function of time, the Eta dewpoint BE appears to decrease slightly with increasing forecast projection (Fig. 8). The smallest BE magnitudes were found at 2100 UTC (in the afternoon). All of the postprocessing methods reduced the dewpoint BE magnitude to less than −0.2°C for all times.

The rankings of the different postprocessing methods remained the same when dewpoint hit rates and the percentage of stations benefiting from the postprocessing methods were calculated (not shown).

c. 10-m wind

Wind forecasts appear to be more difficult to improve since the postprocessing methods produced only modest improvements over the Eta Model. For summer wind speed, ETAKF (ETA7DBR) had a cumulative MAE of 1.6 (1.5) m s^{-1}, which was slightly lower than the 1.8 m s^{-1} of the Eta Model (Table 1). The

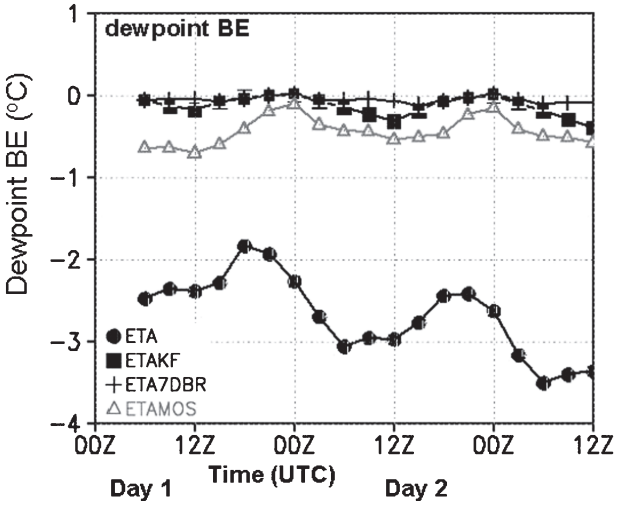

FIG. 7. Average dewpoint BE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

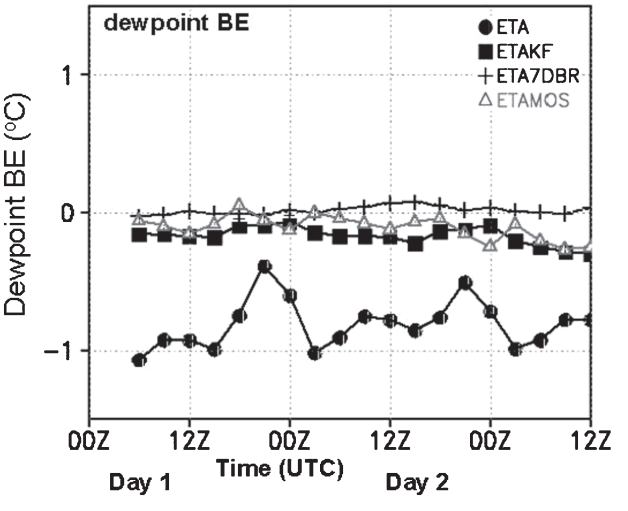

FIG. 8. Average dewpoint BE (°C) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

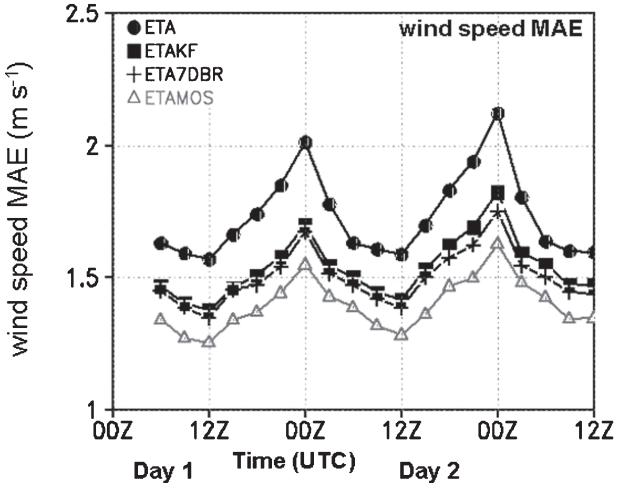

FIG. 9. Average wind speed MAE (m s^{-1}) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

ETAMOS MAE was the lowest at 1.4 m s^{-1}. As a function of time, the Eta Model wind speed MAE varied diurnally between 1.6 and 2.1 m s^{-1} with the maxima at 0000 UTC (Fig. 9). The postprocessed forecast MAEs featured a strikingly similar pattern, but with smaller magnitude (by 0.1–0.5 m s^{-1}). ETAKF was marginally better than ETA7DBR at most times while ETAMOS produced the most accurate forecasts.

In the winter, the cumulative wind speed MAE results were similar. The Eta Model cumulative wind speed MAE was 1.9 m s^{-1} (Table 2). Both ETAKF and ETA7DBR had a cumulative wind speed MAE of 1.7 m s^{-1}, only 0.2 m s^{-1} smaller than the Eta Model. ETAMOS was slightly better, with a cumulative wind speed MAE of 1.6 m s^{-1}. As a function of time, the Eta Model wind speed MAE exhibited a pattern similar to that during summer, except the maximum amplitude occurred earlier (2100 UTC; Fig. 10). The postprocessing methods showed similar variability as a function of time, but with smaller MAEs (by 0.1–0.4 m s^{-1}). Roeger et al. (2003) found that the KF reduced the wind speed MAE by up to 3 m s^{-1} at one station on Blackcomb Mountain, in British Columbia, Canada, during the winter. However, their unfiltered wind speed MAE was more than 5 m s^{-1}, much higher than that in this study, suggesting that the magnitude of their improvement may be a reflection of a consistently large model wind bias at the Blackcomb site.

The Eta Model had a cumulative summer season wind speed BE of −0.7 m s^{-1} (Table 1). ETAKF reduced the cumulative wind speed BE from −0.7 to −0.3 m s^{-1}. ETA7DBR removed the cumulative wind speed BE. ETAMOS reduced the BE by 0.6 m s^{-1} to −0.1 m s^{-1}. As a function of time, the Eta Model had a negative BE for all times and peak negative BE (about −1.2 m s^{-1}) at 0000 UTC (Fig. 11). ETAKF reduced but did not completely eliminate this BE, while ETA7DBR and ETAMOS reduced the wind speed BE at all times to less than 0.2 m s^{-1} in magnitude.

During the winter, the Eta Model cumulative wind speed BE was −0.4 m s^{-1}, smaller than that during the summer (Tables 1 and 2). All of the postprocessing methods produced negligible cumulative wind speed BEs. For the Eta Model and the postprocessing methods, the BE as a function of time was similar to that found in the summer, although the magnitude of the

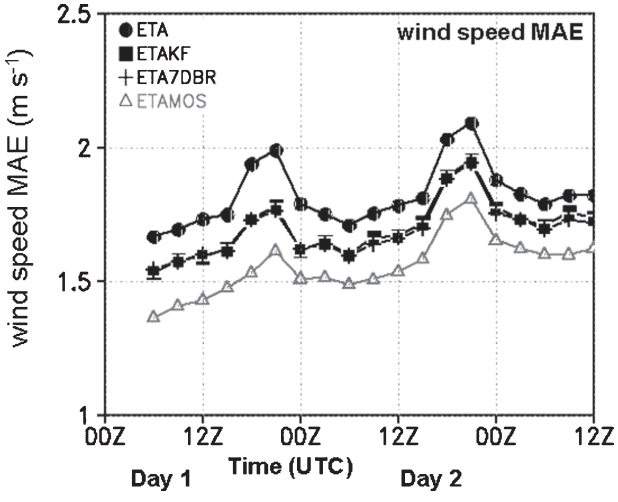

FIG. 10. Average wind speed MAE (m s^{-1}) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

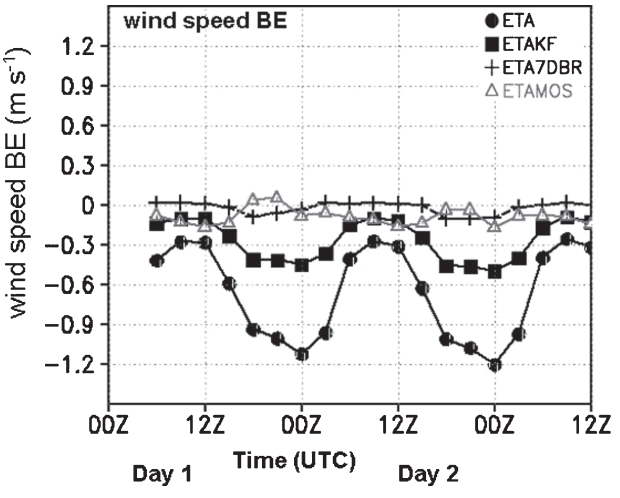

FIG. 11. Average wind speed BE (m s^{-1}) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

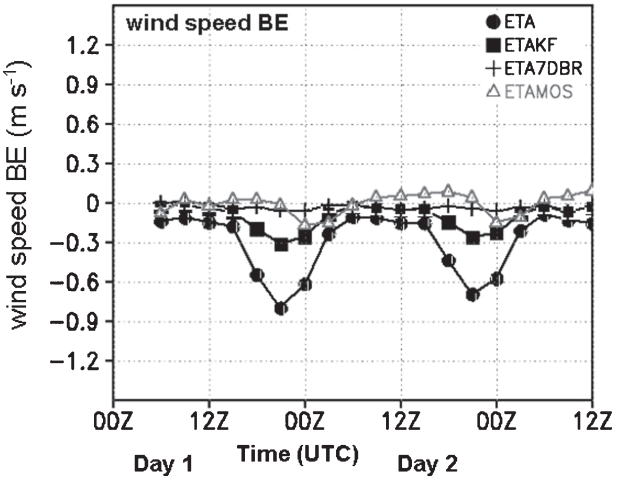

FIG. 12. Average wind speed BE ($m\ s^{-1}$) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

BE was smaller and the maximum magnitude occurred at 2100 UTC instead of 0000 UTC (Fig. 12).

The Eta Model cumulative wind direction MAE for the summer was 53° (Table 1). ETAKF reduced the cumulative MAE to 48°, whereas ETA7DBR produced no improvement. ETAMOS performed better, reducing the cumulative wind direction MAE by 14°. As in other studies for the summer season (e.g., Cheng and Steenburgh 2005), the largest Eta Model wind direction errors occurred in the morning hours (1200 UTC; Fig. 13). It is at this time that ETAKF produces the largest improvements in wind direction MAE, up to 13°. However, ETAMOS outperforms ETAKF by more than 10° for some times. As expected from the cumulative statistics, ETA7DBR was completely ineffective in reducing the wind direction MAE at any time.

During the winter, the wind direction MAE for the Eta Model was 53° (Table 2). Neither ETAKF nor ETA7DBR were effective at lowering the MAE. ETAMOS reduced the wind direction cumulative MAE by 11°. As a function of time, the Eta Model wind direction MAE was maximized between 0300 and 1500 UTC (Fig. 14). ETAKF reduced the MAE by 2°–5° and ETAMOS by 10°–15° during these hours. ETA7DBR was completely ineffective in reducing the MAE of the Eta Model at any time.

As for the other variables, hit rates and the percentage of stations benefiting from postprocessing methods were consistent with MAE (not shown).

4. Case studies

This section presents three cases from Salt Lake City, Utah (SLC), that illustrate the strengths and weaknesses of the postprocessing methods during a quiescent warm season pattern, a persistent wintertime cold pool, and a major air mass change. These cases were selected based on feedback from forecasters who subjectively identified situations in which the strengths and weaknesses of the methods were apparent. For brevity, we focus on temperature.

a. Quiescent warm season pattern

At 1200 UTC 8 July 2004 the Intermountain West was under weak southwesterly upper-level flow between a ridge centered over northern Mexico and a

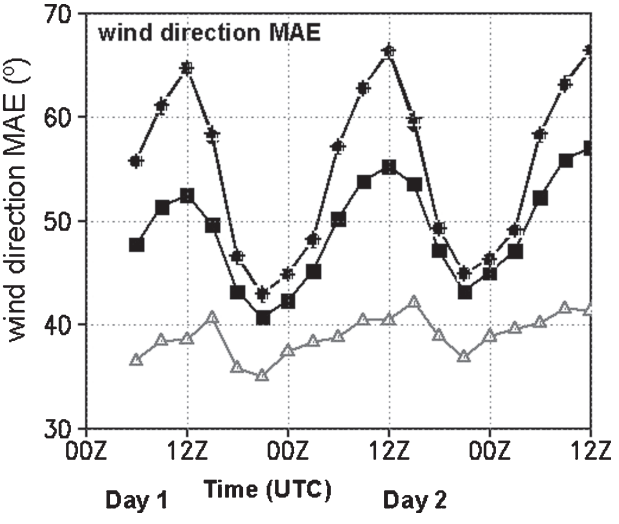

FIG. 13. Average wind direction MAE (°) as a function of time for the combined 0000 and 1200 UTC cycles from Jun to Aug 2004. The convention of the plot is the same as in Fig. 1.

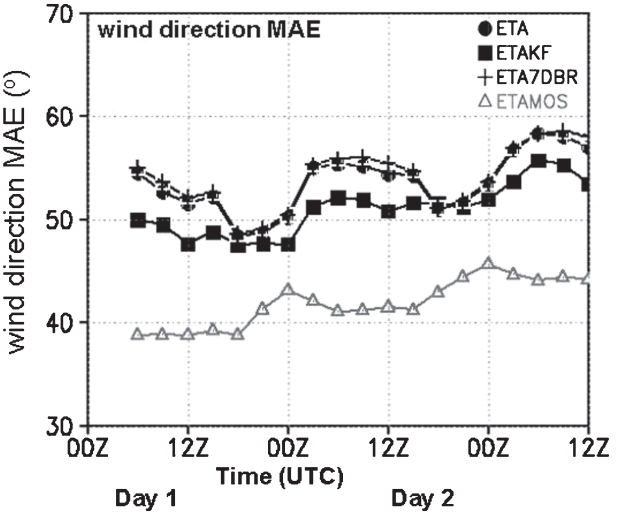

FIG. 14. Average wind direction MAE (°) as a function of time for the combined 0000 and 1200 UTC cycles from Dec 2004 to Feb 2005. The convention of the plot is the same as in Fig. 1.

trough centered over southwest Canada (not shown). Through 12 July the ridge amplified poleward and westward while the trough retrogressed and deepened off the Pacific Northwest coast. Salt Lake City observed periods of mid- and upper-level clouds associated with high-based convection. Afternoon winds were gusty due in part to microbursts, particularly on 10 July. No major large-scale airmass changes occurred.

SLC observed a large diurnal temperature cycle (10°C or greater) during this period with some variability in the quasiperiodic diurnal temperature trace due to convective cloud cover and outflow. The observed temperature for Salt Lake City fits this description in the three 48-h forecast cycles presented in Fig. 15 and initialized at 1200 UTC 8 July, 1200 UTC 9 July, and 1200 UTC 10 July, respectively. In the three forecasts, the Eta Model produced a cumulative temperature MAE of 4.7°C and consistently underpredicted the 2-m temperature, with the underprediction approaching 6°–7°C each afternoon (0000 UTC). Because of the quasi-steady nature of the large-scale pattern, the model bias remained relatively steady in successive forecast cycles and, as a result, the ETAKF and ETA7DBR forecasts were quite accurate with cumulative MAEs of 1.6° and 1.4°C, respectively. ETAMOS produced a cumulative temperature MAE of 1.5°C and no technique produced a forecast that was more than 3°C from the observed throughout the period. This example illustrates that all three techniques produce forecasts of similar quality during quasi-steady summertime conditions when model biases do not change abruptly.

b. Persistent cold pool event

Under cool season anticyclonic conditions, topographically confined cold pools develop in the western United States, resulting in cold temperatures (compared to surrounding high topography) and limited vertical mixing within valleys and basins (Whiteman et al. 2001). Such a cold pool was present over Salt Lake City from 22 to 27 January 2005 when Utah was under the influence of a persistent upper-level ridge (not shown). Throughout this period, the Salt Lake City International Airport (SLC) observed persistent low clouds and fog with temperatures between −2° and 1°C. Figure 16 shows the SLC 48-h temperature forecast from three forecast cycles initialized at (i) 1200 UTC 23 January, (ii) 1200 UTC 24 January, and (iii) 1200 UTC 25 January. For most of the 1200 UTC 23 January forecast cycle, the Eta Model produced a warm bias that reached 3°C at 2100 UTC (in the afternoon; Fig. 16a). ETAKF produced the best overall forecast and was within 2°C of the observed throughout the period (Fig. 16a). ETA7DBR was generally worse than ETAKF be-

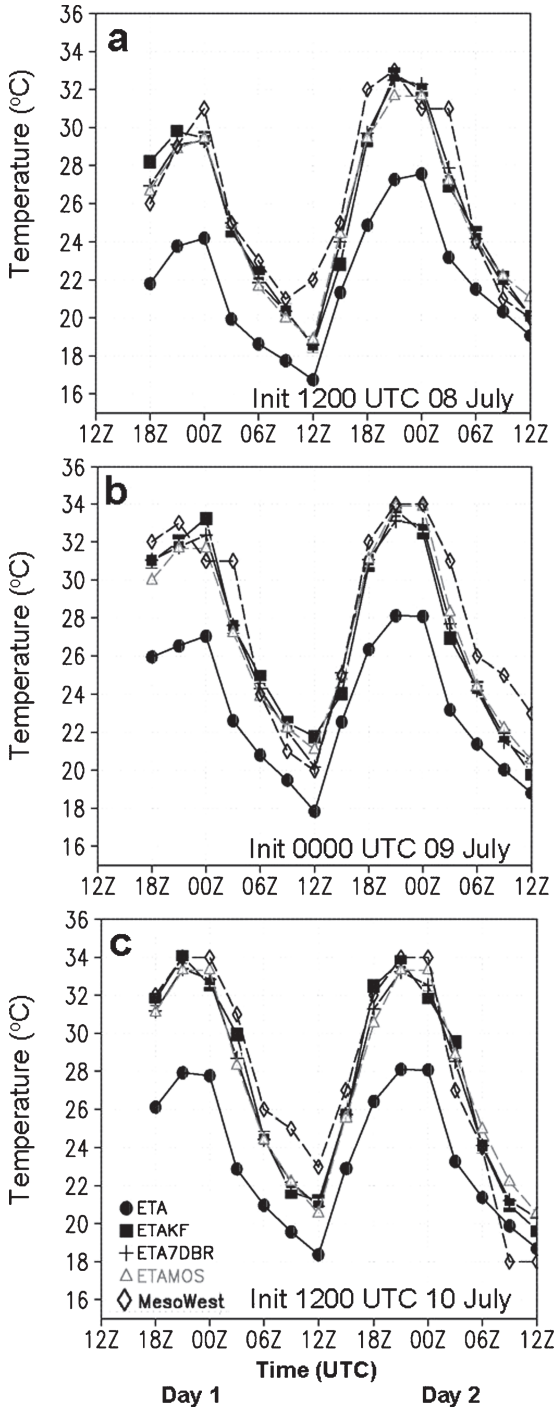

FIG. 15. Salt Lake City 48-h temperature forecast time series (°C) initialized on (a) 1200 UTC 8 Jul, (b) 0000 UTC 9 Jul, and (c) 1200 UTC 10 Jul 2004 for Eta (solid circle), ETAKF (solid square), ETA7DBR (plus sign), ETAMOS (triangle), and MesoWest (diamond). This is an example of a case of quasiperiodic observed temperature.

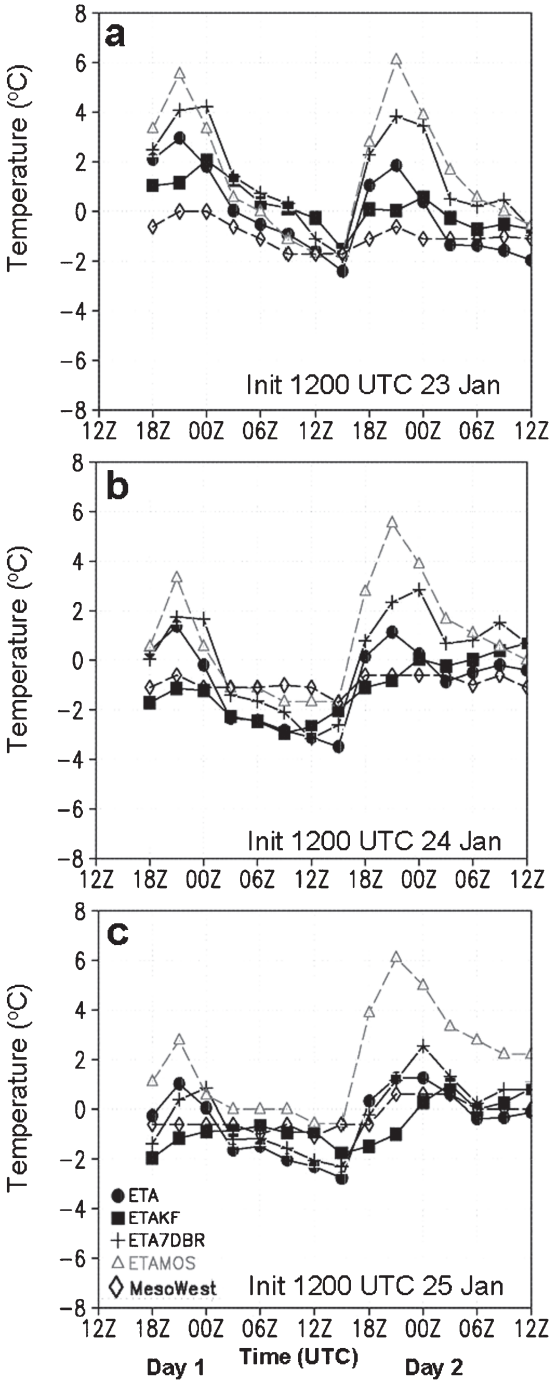

FIG. 16. Salt Lake City 48-h temperature forecast time series (°C) initialized on (a) 1200 UTC 23 Jan, (b) 1200 UTC 24 Jan, and (c) 1200 UTC 25 Jan 2005. The convention of the plot is the same as in Fig. 15. This is an example of a persistent winter cold pool case.

cause the 7-day running-mean bias has still not caught on to the developing cold pool regime. ETAMOS performed the worst of all the postprocessing methods, exhibiting a large warm bias reaching 6°C at 2100 UTC (afternoon), consistent with previous studies of MOS performance during persistent cold pools (Hart et al. 2004).

In the 1200 UTC 24 January forecast cycle, the Eta Model forecast was better than the previous cycle with small bias errors except between 1800 and 0000 UTC when a ~1°–3°C warm bias was evident (Fig. 16b). ETAKF continued to show good agreement with observations (within 2°C). ETA7DBR was not as good as ETAKF, but was better than the previous forecast cycle with an additional day of cold pool conditions entering into the bias removal. ETAMOS again had a general warm bias, especially at 2100 UTC (5°–6°C).

For the entire 1200 UTC 25 January forecast cycle, the Eta Model, ETAKF, and ETA7DBR were all within 2°C of the observed, whereas ETAMOS continued to produce the least accurate forecasts, particularly in the afternoon, with a warm bias as high as 5°C at 2100 UTC 26 January (Fig. 16c).

The cumulative temperature MAE for these three forecast cycles was 1.1°C for the Eta Model, 0.9°C for ETAKF, 1.6°C for ETA7DBR, and 2.2°C for ETAMOS. The comparatively poor performance of MOS during persistent cold pools is well known among NWS forecasters and documented by Hart et al. (2004). Although the Eta Model forecast for SLC was generally good, its poor resolution is unable to capture the spatial structure of the cold pool throughout the lowlands of northern Utah.[2] This is particularly evident if one examines the cumulative Eta BEs at observing sites in the Salt Lake and Utah Valleys from the three forecast cycles noted above. A large warm bias is observed in the Utah Valley and in the western Salt Lake Valley (Fig. 17a). Not surprisingly, Eta Model MAEs were also high at these sites compared to the central and eastern Salt Lake Valley (Fig. 17c). In fact, at some sites the MAE and BE are equal, indicating that the warm bias was present at all times. Applying a Kalman filter resulted in generally smaller BEs and MAEs throughout both valleys (Figs. 17b and 17d). Such accurate high-density point data can be used to better forecast cold pool strength and coverage in gridded forecasts, such as those produced by the National Weather Service Interactive Forecast Preparation System (IFPS).

c. *Transient case*

When the large-scale pattern or air mass changes, such as occurs with a frontal passage, ETAKF and ETA7DBR exhibit their worst performance. The week of 4–10 December 2004 began with mostly clear skies

[2] The problem of underprediction in the cold pool strength also occurs in other forecast models (Myrick and Horel 2006).

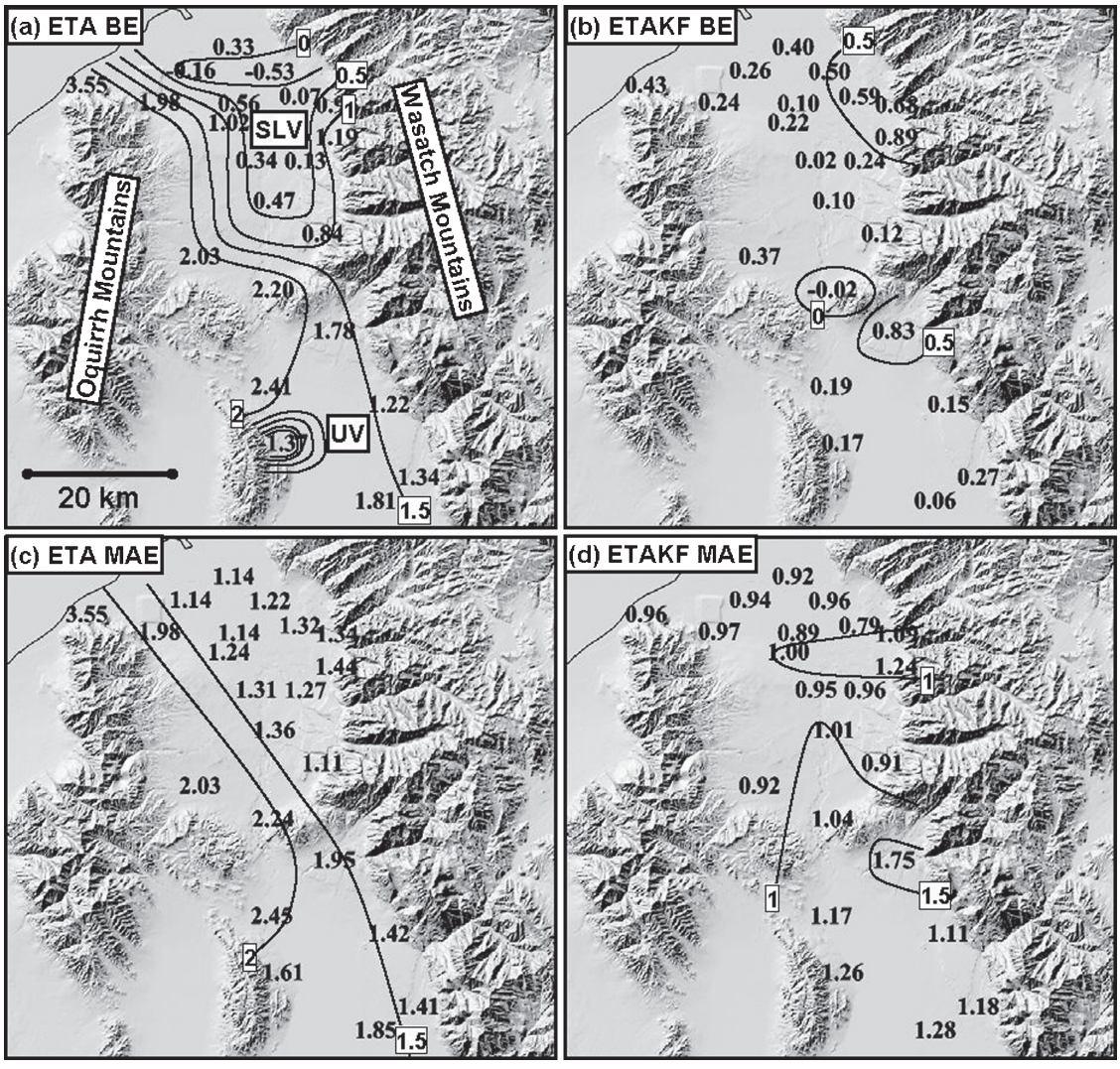

FIG. 17. Cumulative Eta Model and ETAKF verification scores at Salt Lake Valley (SLV) and Utah Valley (UV) MesoWest observing sites from the 1200 UTC 23 Jan, 1200 UTC 24 Jan, and 1200 UTC 25 Jan forecast cycles. Contours every 0.5°C up to 2°C. For (a) the Eta Model BE, (b) ETAKF BE, (c) Eta Model MAE, and (d) ETAKF MAE. Terrain-shaded background with thin contours indicates the south shore of the Great Salt Lake.

on 4 December, followed by fog, and haze on 5 and 6 December. Temperatures ranged from minima of −6° to −7°C to maxima of 0°C. Gradual warming occurred from 7 to 9 December with overcast skies and periods of snow or rain. Temperatures ranged from 0° to 3°C on 7 December to 3° to 7°C by 9 December. Temperatures increased dramatically from 5° to 13°C on the afternoon of 10 December (Fig. 18a). With mostly clear skies over the region, a larger diurnal cycle was observed on 11 (0° to 11°C) and 12 December (−2° to 11°C) compared to the previous several days.

Shown in this example are three 48-h temperature forecasts for Salt Lake City initialized at (i) 0000 UTC 9 December, (ii) 0000 UTC 10 December, and (iii) 0000 UTC 11 December (Fig. 18). Although none of the postprocessing methods were able to handle the warming trend adequately, ETAMOS had the best overall performance and was within 4°C of the observed except during the first warm afternoon on 10 December. There was considerable variability in ETAKF and ETA7DBR performance during the period, but their overall MAEs (3.2°C for both ETAKF and ETA7DBR) were worse that of ETAMOS (2.4°C). Nonetheless, all the postprocessing methods performed better than the Eta Model, which had a cumulative temperature MAE of 4.0°C.

The 0000 UTC 9 December initialized Eta forecast had difficulty capturing the rapid rise in temperature from 1500 to 2100 UTC 21 December, with a cold bias

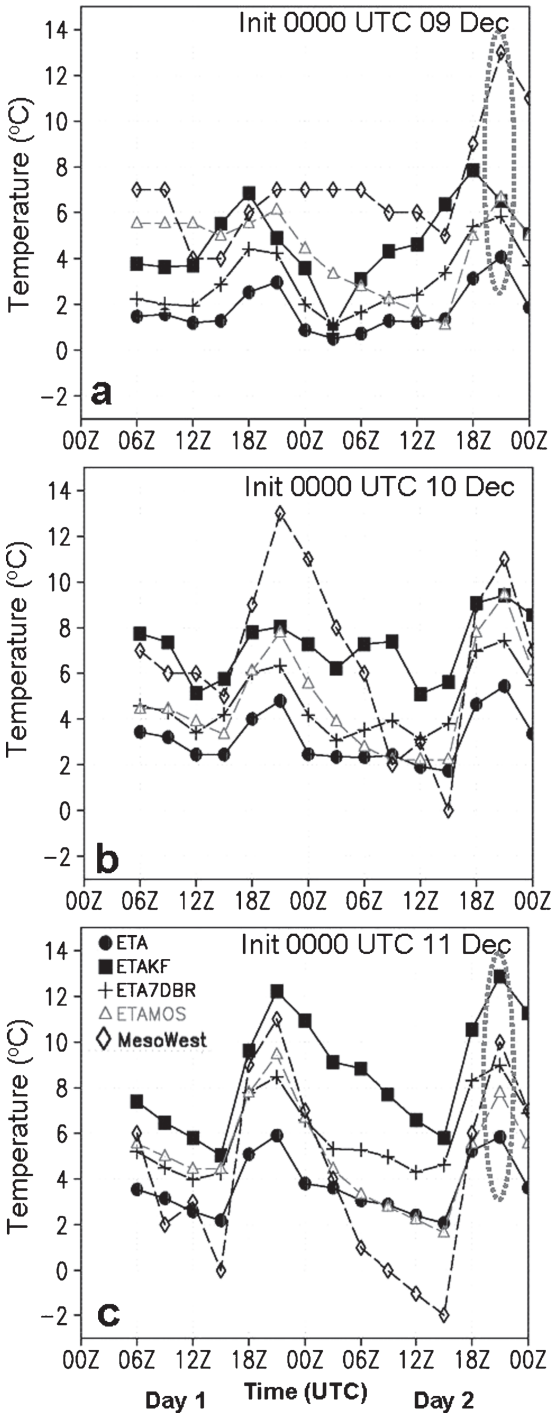

FIG. 18. Salt Lake City 48-h temperature forecast time series (°C) initialized on (a) 0000 UTC 9 Dec, (b) 0000 UTC 10 Dec, and (c) 0000 UTC 11 Dec 2004. The convention of the plot is the same as in Fig. 15. This is an example of a transient case.

as high as 9°C at 2100 UTC 21 December (day 2; Fig. 18a). The postprocessing methods performed better than Eta but still had a large cold bias at 2100 UTC 21 December. At 2100 UTC 21 December, ETAKF and ETAMOS had a cold bias of 6.5°C, and ETA7DBR had a cold bias of 7°C.

In the 0000 UTC 10 December forecast cycle, the Eta Model *again* had difficulty capturing the rapid rise in temperature from 1500 to 2100 UTC 21 December, with a cold bias as large as 8°C at 2100 UTC 21 December (day 1; Fig. 18b). The postprocessing method did not sufficiently adjust this cold bias at 2100 UTC 21 December. ETAKF and ETAMOS had a cold bias of 5°C, while ETA7DBR had a cold bias of 7°C.

The 0000 UTC 11 December initialized forecast period covered 2 days that were clear and featured diurnal cycles that were larger than the observed during the previous several days (Fig. 18c). The Eta Model produced an underdeveloped boundary layer in the afternoon and was too warm at night with a 3°–5°C afternoon cold bias and a 2°–4°C early morning warm bias. ETAKF had particular difficulties from 0000 UTC 12 December to 0000 UTC 13 December, performing worse than the Eta Model, as the forecast bias estimate at this time relied on the observed and estimated temperature biases from 0000 UTC 10 December to 0000 UTC 11 December of the 0000 UTC 9 December forecast cycle (see Figs. 18a, 18c, and 19, and appendix A), which is obviously inappropriate for the clear-sky conditions. As a result, ETAKF had a large warm bias (4°–8°C) from 0000 UTC 12 December to 1500 UTC 12 December, when the observed temperature cooled rapidly after sunset. ETA7DBR also had a similar problem from 0000 UTC 12 December to 0000 UTC 13 December as ETAKF as it used an inappropriate bias based on the recent past. However, the ETA7DBR adjustment was smaller than that of ETAKF, leading to a smaller warm bias (1.5°–6.5°C) from 0000 UTC 12 December to 0000 UTC 13 December. ETAMOS performed the best among all of the postprocessing methods with errors ranging from 0° to 4°C throughout the entire forecast cycle.

5. Summary remarks

We have evaluated the performance of three postprocessing methods for improving 2-m temperature, 2-m dewpoint, and 10-m wind speed and direction forecasts produced by the Eta/NAM model: (i) ETAKF, (ii) ETA7DBR, and (iii) ETAMOS. The following are our primary conclusions.

- Overall, ETAMOS is the most accurate technique (in terms of MAE, BE, and hit rate) for all variables.
- ETAKF and ETA7DBR have comparable accuracy for 2-m temperature, 2-m dewpoint, and 10-m wind speed. However, ETAKF produced more accurate wind direction forecasts.

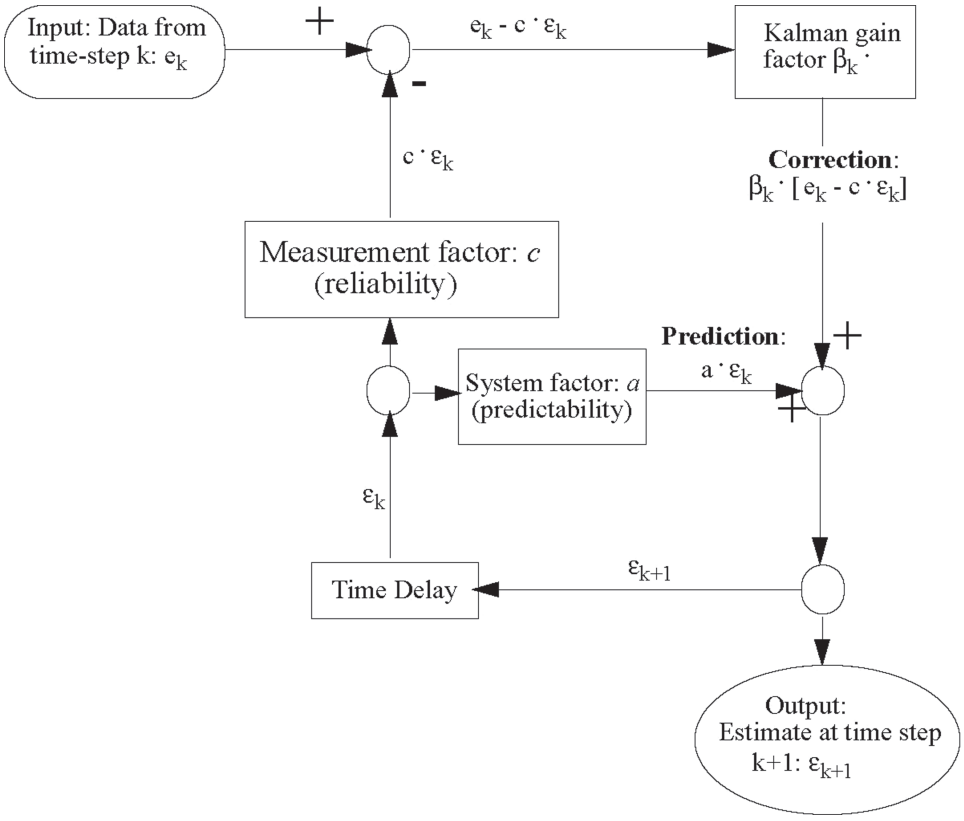

FIG. 19. Flowchart for the Kalman filter algorithm (adapted from Roeger et al. 2003).

- Although the overall performance is similar, ETAKF adjusts more quickly than ETA7DBR when model biases change, such as occurs during large-scale pattern changes. Nevertheless, the performance of ETAKF and ETA7DBR suffers when the model bias changes dramatically from one forecast cycle to the next.
- ETAKF and ETA7DBR forecasts are superior to ETAMOS during persistent wintertime cold pools and similar to ETAMOS during quiescent warm season patterns.

Despite the good performance of MOS, it requires a rather long training dataset and therefore can be difficult to apply to modeling systems that undergo major changes and to observing networks and sites that lack a long and complete historical record. The KF can be used, however, with new or recently changed modeling systems and to produce forecasts for all locations for which recent data are available. This is a characteristic that can be highly desirable for taking full advantage of observations provided by integrated cooperative mesonets such as MesoWest, which as of October 2007 included data from over 100 providers and 6000 stations in the western United States. Many of these sites are in regions or locations where MOS is unavailable and the KF can provide a superior forecast (relative to the native model output) at these locations. Further, KF-based forecasts provide superior guidance during persistent cold pool events, which are common over the western United States during the cool season. The KF cold pool forecasts are not only more accurate at station locations, but can also be used to improve graphical forecasts since they include more stations that can be used to better define cold pool structure and strength. NCEP recently announced the discontinuation of MOS for the Eta Model effective in early 2007 due to the replacement of the Eta Model with the nonhydrostatic version of the Weather Research and Forecasting (WRF) model. Meanwhile, forecasters still need numerical guidance from WRF. The KF can provide numerical guidance for WRF in the interim while WRF MOS is under development. In fact, work is under way to implement the KF at several WFOs in the western and central regions.

Even with the attractive properties of the KF, its poor performance when model biases change dramatically represents a major drawback. This is especially true in the winter season with the ubiqituous synoptic-scale weather systems. Thus, postprocessing errors

are large when the KF is applied to longer-range forecasts since the estimates of prior bias are from model runs that can be several days old. We tested the idea that the Kalman filter correction should not be applied during periods when weather pattern changes, specifically for strong cold frontal passage in the Intermountain West. In these cases, Kalman filter correction actually increases the forecast errors. The procedure is as follows. If the 24-h temperature tendency exceeds 10°C, then Kalman filter correction was not applied for the past 12 h. This procedure works well if the model forecast is accurate and is not generally applicable for all weather phenomena. More work is needed to determine how to detect different weather phenomena objectively.

Except for wind direction, generally, the 7-day running-mean bias removal method performed similarly to the Kalman filter correction. However, the Kalman filter adapts faster to weather change patterns, and there is certainly room for improvement on the version of the Kalman filter used in this study. Future studies should explore more complex forms of the Kalman filter, such as an ensemble Kalman filter (Gu and Oliver 2006), a dual Kalman filter (Aboy et al. 2005), and an unscented Kalman filter (Romanenko et al. 2004) to improve upon the current version of NWP model bias removal.

Acknowledgments. This project was sponsored by the National Oceanic and Atmospheric Administration through a series of grants to the Cooperative Institute for Regional Prediction (CIRP) at the University of Utah and a subaward from the University Corporation for Atmospheric Research as part of the COMET Outreach Program. The authors thank the science operations officers (SOOs) and forecasters of the Boise, Idaho; Grand Junction, Colorado; Pocatello, Idaho; Riverton, Wyoming; and Salt Lake City NWS WFOs for their collaboration (especially Dr. Tim Barker); Prof. John Horel for his insights regarding the Kalman filter; and the MesoWest team for enabling access to their observational database. Review of the manuscript by Prof. John Horel, Drs. Tim Barker and Brett MacDonald, and two anonymous reviewers is much appreciated. Special thanks go to Matt Masarik and Chad Kahler for retrieving the Eta grids and ETAMOS from the CIRP archive and to Brenda Thompson for proofreading the manuscript.

APPENDIX

The Kalman Filter Algorithm

In its simplest form, the KF can be described by a first-order Markov process (i.e., an autoregressive process). The KF is a recursive, predictor–corrector scheme in which the previous estimate and observation of the state variable from the k-1 forecast cycle are used to give a better estimate of the state variable for the k forecast cycle. We followed the KF formulation of Roeger et al. (2003). If $e_{i,k} = A_{i,k}(\text{fcst}) - A_{i,k}(\text{obs})$ is the bias for variable A for an arbitrary forecast hour i of the k forecast cycle, then the bias of A at forecast hour i of the $k + 1$ forecast cycle can be calculated from

$$e_{\mathrm{i},k+1} = a_i e_{i,k} + w_{i,k},$$

where $w_{i,k}$ is a term representing the noise of the system and has the form of a Gaussian distribution with variance $\sigma_{\iota}^{(w)2}$ and the coefficient a_i represents the relationship between the k and $k + 1$ forecast cycle biases. Input observations are assumed to be noisy (represented by a Gaussian noise term, $\upsilon_{i,k}$, with variance $\sigma_{\mathrm{i}}^{(\upsilon)2}$) so that

$$e_{i,k} = c_i \varepsilon_{i,k} + \upsilon_{i,k},$$

where c_i is a coefficient that relates the estimated ($\varepsilon_{i,k}$) and actual ($e_{i,k}$) biases. Finally, the Kalman filter recursive relations can be written as follows:

$$\varepsilon_{i,k+1} = a_i \varepsilon_{i,k} + \beta_{i,k}(e_{i,k} - \varepsilon_{i,k}),$$

where $\beta_{i,k}$, the gain or blending factor, is given by

$$\beta_{i,k} = a_i c_i p_{i,k}[c_i^2 p_{i,k} + \sigma_i^{(\upsilon)2}]^{-1},$$

and the prediction mean-square error is given by

$$p_{i,k+1} = a_i^2 p_{i,k} - a_i c_i \beta_{i,k} p_{i,k} + \sigma_i^{(w)2}.$$

Further details of the Kalman filter can be found in Kalman (1960) and Kalman and Bucy (1961). A flowchart of the above algorithm is presented in Fig. 19. A brief description of how the KF was applied at the University of Utah is presented below.

Eta forecast grids were available on a 3-hourly basis. Bilinear interpolation of the Eta forecast grid was performed spatially to MesoWest station points. Also, linear interpolation was performed temporally so that the Eta/NAM forecast grids were available at hourly intervals, although we present only 3-h interval forecasts here to facilitate comparison with ETAMOS. The KF was applied to each forecast hour independently for the following variables: 2-m temperature, 2-m dewpoint, 10-m wind speed, and 10-m zonal and meridional wind components (used to compute the wind direction). MesoWest observations were used to determine the observed bias for the previous forecast cycle through forecast hour 24. Because the longer-range forecasts of the previous forecast cycle have not yet verified, we use the day −2 cycle bias for forecast hours 25–48 (Fig. A1).

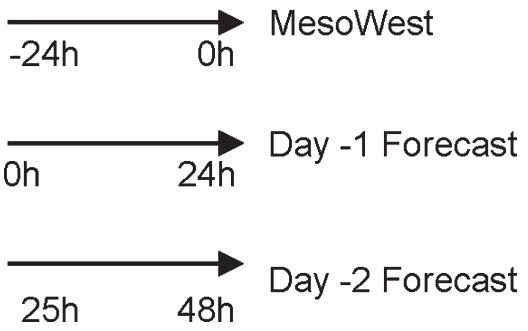

FIG. A1. Schematic of the procedure used in calculating the observed bias for the previous Kalman filter cycle for various forecast hours. Note that MesoWest observations for the previous 24 h are used. For forecast hours 0–24, the day −1 forecast is verified against the MesoWest observations in the previous 24 h. Since the observations for forecast hours 25–48 are not yet available for the day −1 forecast, we must use the day −2 forecast to compute the observed bias of forecast hours 25–48.

Note that each variable for each time has its own KF. Except for the wind vector components where a was chosen to be 0.75, a_i was chosen to be unity for all variables and times. In addition, c_i was chosen to be unity for all variables and times as well. Both $\sigma_\iota^{(v)}$ and $\sigma_\iota^{(w)}$ were chosen to be equal to each other for all variables and times. The absolute values of $\sigma_\iota^{(v)}$ and $\sigma_\iota^{(w)}$ do not matter; only their ratio is important (Bozic 1979). By trial and error, these parameter choices were found to perform the best. Also, the estimated bias and mean prediction error were set to be zero for the time step $k = -1$ at the initial time. Finally, the 4–5-day training period, mentioned in the text, refers to the time period of the KF after its cold start when the oscillatory behavior of the processing method was observed to decrease.

REFERENCES

Aboy, M., O. W. Marquez, J. McNames, R. Hornero, T. Trong, and B. Goldstein, 2005: Adaptive modeling and spectral estimation of nonstationary biomedical signals based on Kalman filtering. *IEEE Trans. Biomed. Eng.,* **52,** 1485–1489.

Benoit, R., M. Desgagne, P. Pellerin, S. Pellerin, and Y. Chartier, 1997: The Canadian MC2: A semi-Langrangian, semi-implicit wideband atmospheric model suited for finescale process studies and simulation. *Mon. Wea. Rev.,* **125,** 2383–2415.

Bozic, S. M., 1979: *Digital and Kalman Filtering*. John Wiley and Sons, 153 pp.

Cheng, W. Y. Y., and W. J. Steenburgh, 2005: Evaluation of surface forecasts by the WRF and Eta Models over the western United States. *Wea. Forecasting,* **20,** 812–821.

Colle, B. A., J. B. Olson, and J. S. Tongue, 2003: Multiseason verification of the MM5. Part I: Comparison with the Eta Model over the central and eastern United States and impact of MM5 resolution. *Wea. Forecasting,* **18,** 431–457.

Dimego, G., and Coauthors, cited 2005: Mesoscale modeling branch: Where we are and where we are going? [Available online at http://www.emc.ncep.noaa.gov/research/NCEP-EMCModelReview2005/Prodrev-EMC-mmb.pdf.]

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Gu, Y. Q., and D. S. Oliver, 2006: The ensemble Kalman filter for continuous updating of reservoir simulation models. *J. Energy Resour. Technol.—Trans. ASME,* **128,** 79–87.

Hart, K. A., W. J. Steenburgh, D. J. Onton, and A. J. Siffert, 2004: An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Wea. Forecasting,* **19,** 200–218.

Horel, J., and Coauthors, 2002: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.,* **83,** 211–225.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME—J. Basic Eng.,* **82,** 35–45.

——, and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *Trans. ASME—J. Basic Eng.,* **83,** 95–108.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

Myrick, D. T., and J. D. Horel, 2006: Verification of surface temperature forecasts from the National Digital Forecast Database over the western United States. *Wea. Forecasting,* **21,** 869–892.

Paegle, J., Q. Yang, and M. Wang, 1997: Predictability in limited area and global models. *Meteor. Atmos. Phys.,* **63,** 53–69.

Pielke, R. A., Sr., 2002: *Mesoscale Meteorological Modeling*. 2d ed. Academic Press, 676 pp.

Roeger, C., R. Stull, D. McClung, J. Hacker, X. Deng, and H. Modzelewski, 2003: Verification of mesoscale numerical weather forecasts in mountainous terrain for application to avalanche prediction. *Wea. Forecasting,* **18,** 1140–1160.

Romanenko, A., L. O. Santos, and P. Afonso, 2004: Unscented Kalman filtering of a simulated pH system. *Indust. Eng. Chem. Res.,* **43,** 7531–7538.

Stephenson, D. B., and I. T. Jolliffe, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 254 pp.

Whiteman, C. D., S. Zhong, W. J. Shaw, J. M. Hubbe, X. Bian, and J. Mittelstadt, 2001: Cold pools in the Columbia Basin. *Wea. Forecasting,* **16,** 432–447.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Academic Press, 627 pp.