Hydrometeorological Accuracy Enhancement via Postprocessing of Numerical Weather Forecasts in Complex Terrain

DOUG McCOLLOR

University of British Columbia, and BC Hydro Corporation, Vancouver, British Columbia, Canada

ROLAND STULL

University of British Columbia, Vancouver, British Columbia, Canada

(Manuscript received 27 November 2006, in final form 27 May 2007)

ABSTRACT

Statistical postprocessing techniques such as model output statistics are used by national weather centers to improve the skill of numerical forecasts. However, many of these techniques require an extensive database to develop, maintain, and update the postprocessed forecasts. This paper explores alternative postprocessing techniques for temperature and precipitation based on weighted-average and recursive formulations of forecast–observation paired data that do not require extensive database management, yet provide distinct error reduction over direct model output.

For maximum and minimum daily temperatures, seven different postprocessing methods were tested based on direct model output error for forecast days 1–8. The methods were tested on a 1-yr series of daily temperature values averaged over 19 stations in complex terrain in southwestern British Columbia, Canada. For daily quantitative precipitation forecasts, three different postprocessing methods were tested over a 6-month wet season period.

The different postprocessing methods were compared using several verification metrics, including mean error (for temperature), degree of mass balance (for precipitation), mean absolute error, and threshold error. All of the postprocessing methods improved forecast skill over direct model output. The postprocessing methods for temperature forecasts require a much shorter training period (14 days) than precipitation forecasts (40 days) to accomplish error reduction over direct model output forecasts. The postprocessing methods that weight recent error estimates most heavily perform better in the short term (days 1–4) while methods that weight recent and earlier error estimates more evenly show improving relative performance in the midterm (days 5–8).

For temperature forecasts, Kalman filtering produced slightly better verification scores than the other methods. For precipitation forecasts, a 40-day moving-average weighting function and the best easy systematic estimator method produced the best degree of mass balance results, while a seasonally averaged method produced the lowest mean absolute errors and lowest threshold errors. The methods described in this paper require minimal database management or computer resources to update forecasts, and are especially viable for hydrometeorological applications that require calibrated daily temperature and precipitation forecasts.

1. Introduction

Statistical postprocessing to calibrate NWP output has a long history in operational weather forecasting. It is firmly established that statistical postprocessing of NWP output significantly improves the skill of deterministic forecasts, primarily through the reduction of systematic errors (Hamill et al. 2000). The goal of post-processing was originally to produce point forecasts of sensible weather elements that were not readily available from early low-resolution models, incorporating statistical relationships between large-scale aspects of the flow and specific point observations. More recently, postprocessing has been employed to improve the skill of point forecasts of sensible weather elements obtainable directly from higher-resolution models. The objective is to reduce the future systematic error between direct model output (DMO) forecasts and verifying observations by building and employing statistical relationships between past model output and past observations.

Previous studies (Stensrud and Skindlov 1996; Mao

Corresponding author address: Doug McCollor, Dept. of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Rd., Vancouver, BC V6T 1Z4, Canada.
E-mail: doug.mccollor@bchydro.bc.ca

et al. 1999; Eckel and Mass 2005; Stensrud and Yussouf 2005) have shown how straightforward moving-average techniques can reduce systematic error in DMO. In the study presented in this paper, moving-average and other related postprocessing techniques are applied to daily maximum and minimum temperature forecasts and daily quantitative precipitation forecasts (QPFs). These sensible weather element forecasts of temperature and precipitation are extremely important in hydrometeorological applications such as runoff forecasting. The goal of these postprocessing techniques is to reduce the error in the current forecast using estimates of past error.

To achieve this goal, a balance must be reached between a short learning period, to enable an updating system to respond quickly to changes in error patterns, and a longer training period that increases statistical stability (Woodcock and Engel 2005). Several moving-average and weighted-average methods are compared in the present paper, along with a Kalman filter technique, to determine the effect of these straightforward postprocessing techniques in reducing systematic error in DMO forecasts. Second, the techniques presented here are compared to postprocessed forecasts produced by the Canadian Meteorological Centre (CMC), in an effort to test these new techniques against an established standard.

a. Background

Both systematic error and random error exist in a forecast. In mountainous regions, the systematic component of error, often termed bias, is largely due to differences between model topography and actual orography and land use, and also deficiencies in the model representation of physical processes. Subgrid-scale processes parameterized in the model do not adequately represent local dynamic and thermodynamic effects required for detailed point-specific forecasting (Hart et al. 2004). Improvement in model parameterization and representation of topography as well as postprocessing DMO can reduce the systematic error component. Random forecast error, on the other hand, is largely associated with the inherent chaotic nature of atmospheric motion and cannot be reduced through postprocessing. Associated observations also have inherent and sometimes unknown bias and random error due to equipment limitations such as sampling rate, calibration, and site representation.

Forecast error, then, can be written $\varepsilon^f = x^f - x^t$, where x^t is the unknown truth and x^f is a forecast. Sample realizations of the real atmosphere can be taken as observations y^o, which also have error $\varepsilon^o = y^o - x^t$. If observation errors are random (not due to equipment miscalibration), then $\langle y^o \rangle = \langle x^t \rangle$, where $\langle \rangle$ is the expected or time-averaged value. A bias is then the expected value of the forecast error, $\langle \varepsilon^f \rangle = \langle x^f - x^t \rangle = \langle x^f - y^o \rangle$. Expected values can be estimated from a sufficiently large sample of errors. The expectation is approximated by the mean of a given sample of errors, where the sample can include the entire climatological record or be limited to (hence characteristic of) the recent synoptic flow regime. The best postprocessing techniques will attempt to address both sources of bias in the forecast.

Calibrated forecasts of temperature and precipitation are necessary when used as input into hydrologic streamflow forecast models. Such forecasts pose a unique challenge in areas of complex terrain where meteorological conditions exhibit dramatic spatial variability (Hart et al. 2004). Specifically for hydrometeorological applications, improvements in streamflow forecasts require more accurate local-scale forecasts of precipitation and temperature (Clark and Hay 2004).

DMO of sensible weather elements, such as temperature and precipitation amount, are often not optimal estimates of local conditions in complex terrain. One reason is the large systematic error inherent in the NWP models (Mao et al. 1999; Clark and Hay 2004). Numerous methods have been devised to adapt NWP forecasts to produce local, or point, forecasts for specific applications. The general methodology of postprocessing techniques derives a statistical regression between NWP-produced values and weather-element observations over a set period of time, and then uses this regression to adjust future NWP forecasts. If successful, the adjusted, or postprocessed, forecasts should have reduced mean error compared to the original DMO.

In complex mountainous terrain, regression adjustments to local NWP forecasts are highly flow dependent. All but extremely high-resolution NWP models will have difficulty discerning topographically induced differences in precipitation patterns, but a properly trained postprocessing technique could likely improve the individual forecasts by greatly reducing the measured forecast–observation error due to topographic effects. Only postprocessing techniques with very short averaging periods would be able to adapt rapidly changing flow-dependant conditional bias correction factors. However, short averaging periods lead to statistical instability. The best postprocessing methods must find a balanced trade-off between conditional bias correction and statistical stability.

b. Review

The earliest method of statistical postprocessing that gained widespread use was the perfect prog method

(Klein et al. 1959), which derives regression equations between analyzed fields and near-concurrent observations. However, since the perfect prog method uses analyses for the predictors, forecast model error is not taken into account in determining the regression coefficients. The method of model output statistics (MOS; Glahn and Lowry 1972) uses model forecast fields for predictors; hence, MOS forecast variance is both model and forecast-period dependant. Therefore, MOS requires forecasts from a model that remains unchanged over at least several seasons (Stensrud and Yussouf 2005) so that the operational forecast relationship matches the developmental historical relationship. Woodcock and Engel (2005) suggest that the importance of DMO will increase relative to MOS because MOS cannot easily accommodate new sites, models, and model changes; MOS often employs over a million predictive equations that require at least 2 yr of stable developmental data for their derivation (Jacks et al. 1990).

For hydrometeorological applications that require operational forecasts for many locations and many projection times, the development and maintenance of an MOS system can prove too onerous. Even national weather centers such as the Meteorological Service of Canada (MSC) discontinued their MOS forecast system in the late 1980s because significant model changes had become too frequent and the development cost too great to maintain a statistically stable MOS system (Wilson and Vallée 2002).

Further research has indicated updateable MOS (UMOS) can lessen the disadvantage of model variance over standard MOS routines. UMOS (Wilson and Vallée 2002, 2003) can be implemented with briefer training than MOS, but requires an even more extensive database overhead to maintain and update. UMOS forecasts in western Canada are provided by MSC, but only at a limited set of valley-bottom airport sites, similar to that found by Hart et al. (2004) for MOS forecasts in the western United States.

The drawbacks to MOS-based techniques have led to other approaches to statistically postprocessing model forecast output that do not require long data archival periods (Stensrud and Yussouf 2003). The overall objective of error correction is to minimize the error of the next forecast using measurement estimates of past errors. A short learning period enables an updating system to respond quickly to factors that affect error (e.g., model changes and changes in weather regime) but increases vulnerability to missing data and/or large errors and other limitations of small sample estimates (Woodcock and Engel 2005).

Simple 7- and 12-day moving-average error calculations are shown to improve upon raw model point fore-
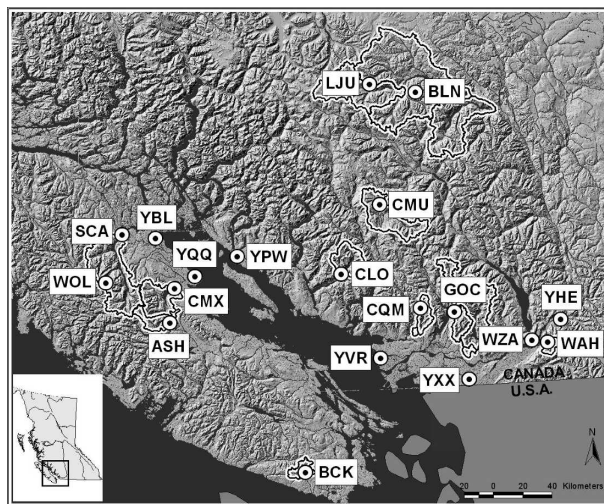

FIG. 1. The area of study in southwestern BC, Canada. Nineteen weather-station locations (dots) in 15 watersheds (black lines) are depicted in the figure. Elevations range from 2 m MSL at YVR to 1920 m MSL at BLN.

casts (Stensrud and Skindlov 1996; Stensrud and Yussouf 2003, 2005). Eckel and Mass (2005) successfully implemented a moving-average technique as a postprocessing method to reduce the systematic error in DMO, using a 14-day moving-average error calculation to reduce mean error in temperature forecasts. Jones et al. (2007) found a 14-day moving average bias correction improved temperature forecasts better than a 7- or 21-day moving average. Woodcock and Engel (2005) state that a 15–30-day bias correction window effectively removes bias in DMO forecasts using the best easy systematic estimator.

Kalman filtering (Homleid 1995; Majewski 1997; Kalnay 2003; Roeger et al. 2003; Delle Monache et al. 2006b) is adaptive to model changes, using the previous observation–forecast pair to calculate model error. It then predicts the model error to correct the next forecast. This recursive, adaptive method does not need an extensive, static database to be trained.

In summary, statistical postprocessing adds value to DMO by objectively reducing the systematic error between forecasts and observations by producing site-specific forecasts. In the following sections, several weighted-average postprocessing methods, a recursive method, and an updateable MOS method are investigated for hydrometeorological applications (specifically temperature and precipitation forecasts) in complex terrain.

2. Data

A mesonetwork of observations and forecasts from 19 locations in southwestern British Columbia is em-

ployed in this study to compare various postprocessing techniques (see Fig. 1). The mesonetwork includes 12 observation sites operated as part of a hydrometeorologic data analysis and forecast program in support of reservoir operations for 15 managed watersheds within the region. Additionally, seven sites maintained by MSC were included in this study. Station elevations range from 2 m MSL at coastal locations to 1920 m MSL in high mountainous watersheds, to capture significant orographic components of the forecasts and observations. Southwestern British Columbia is a topographically complex region of land–ocean boundaries, fjords, glaciers, and high mountains, and is subject to rapidly changing weather conditions under the influence of landfalling Pacific frontal systems.

a. Temperature

The time period 1 April 2004–31 March 2005 defined the 1-yr evaluation portion of this study. This period exhibited examples of extreme temperature. The lowest temperature recorded at any of the stations during the period was −24°C, while the highest temperature recorded at any station was +38°C. Concurrent forecasts were obtained from NWP forecasts produced by the Canadian Meteorological Centre. Point forecasts for each of the 19 sites were obtained for forecast days 1–8 from the operational Global Environmental Multiscale (GEM) model (Côté et al. 1998a,b). Each forecast day was treated separately by applying the techniques independently to a series of forecast–observation pairs valid for each particular forecast day (1–8) only.

For maximum and minimum temperature forecasts, six different postprocessing techniques were applied to the DMO on a daily basis for direct comparison during the period of study. The first method was a seasonal mean error (SNL). In SNL, the average mean error over a constant previous 6-month period was subtracted from the current day's forecast. SNL mean error was seasonally adjusted. For forecasts issued for the cool season (1 October–31 March), a mean error calculated from the previous cool season was calculated and applied. For the warm season (1 April–30 September), a separate mean error was calculated from the previous warm season and applied. The second method was a moving average of the previous error estimates (MA), with each previous value receiving equal weight. A simple moving average is the unweighted mean of the previous *n* data points, where *n* is a predetermined value. The third method was a linear-weighted average of the previous errors (LIN). For LIN forecasts, recent error estimates are weighted more heavily than previous error, in a linear fashion. The fourth method was a weighted average with a $\cos^2$ weighting function

TABLE 1. Table of postprocessing methods and descriptions for temperature and precipitation forecasts. The postprocessing methods are applied to DMO forecasts of daily maximum and minimum temperature for days 1–8. The three methods (SNL, MA, and BES) are evaluated for 24-h precipitation forecasts for days 1–8.

Postprocessing method	Description
SNL	6-month seasonal average
MA	Moving average
LIN	Linear weighting
COS^2	cos^2 weighting
BES	Best easy systematic estimator
KF	Kalman filter

(COS^2) applied. The fifth method uses the best easy systematic estimator (BES) as described in Woodcock and Engel (2005). The BES method is robust with respect to extreme values but represents the bulk of the mean error distribution because it involves quartiles:

$$BES = (Q1 + 2Q2 + Q3)/4, \qquad (1)$$

where Q1, Q2, and Q3 are the first, second, and third quartiles, respectively. The sixth method was a Kalman filter (KF) applied to the forecast error (see appendix A for a description of the KF method), which recursively weights recent errors more heavily than past errors.

The concept of using LIN and COS^2 weighting functions is also to weight recent error estimates more heavily than past error estimates in a smoothly varying form. The objective is to achieve a balance between a short weighting period to include recent changes in error due to weather regime and model changes and a longer weighting period to enhance statistical stability. The postprocessing techniques are summarized in Table 1.

The equation used to apply the weights is given by

$$\varepsilon_c^f = \frac{1}{M} \sum_{k=1}^{M} w_k \cdot \varepsilon_k^f, \qquad (2)$$

where

- ε_c^f is the error correction applied to today's DMO by subtracting this value from the current DMO forecast;
- M is the number of prior day's errors to be weighted, also known as the window length;
- w_k is the weight on the kth day prior to the current day, where $\Sigma_{k=1}^{M} w_k = 1$; and
- ε_k^f is the error estimate on the kth day prior to the current day, $x_k^f - y_k^o$.

The next issue to address is the length of the sample error correction window, to find a balance between synoptic pattern changes and model changes (requiring a shorter window) and statistical stability (requiring a
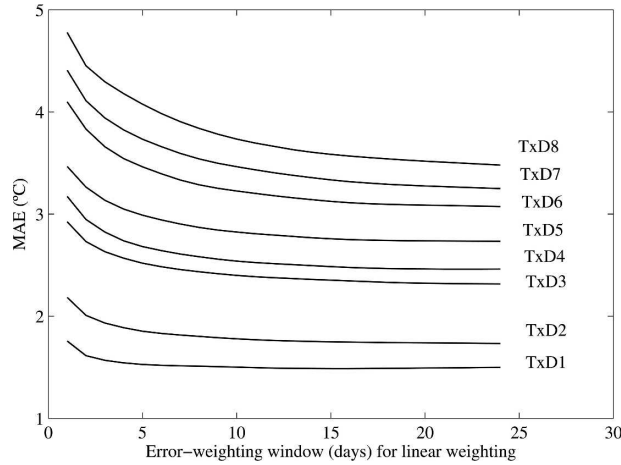
FIG. 2. Mean absolute error (°C) vs window length in days for the LIN method of postprocessing systematic error correction for daily maximum temperature. Plots are for forecast days (bottom) 1 through (top) 8 as shown in the diagram.

longer window). Published studies include windows of $M = 7$ days (Stensrud and Skindlov 1996), 12 days (Stensrud and Yussouf 2005), 21 days (Mao et al. 1999), and 15–30 days (Woodcock and Engel 2005). Woodcock and Engel (2005) test the relationship between number of days in the running error-correction window and improvement in the day 1 forecast.

A similar test was performed in the study reported here, for forecast days 1–8. It was found that all postprocessing methods tested showed a rapid decrease in mean absolute error, reaching an asymptotic value by 14 days (see Fig. 2 for the test using maximum temperature forecasts and the LIN method of mean error reduction). In Fig. 2, all forecast days show a rapid improvement in MAE initially, then reach a long-term value. Longer-range forecasts take systematically more time to reach the long-term value. By day 14 all forecast days have reached a steady value. Similar results were obtained (not shown) for the MA, COS^2, and BES methods and for minimum temperatures. Therefore, a common value of 14 days was employed for the MA, LIN, COS^2, and BES methods.

UMOS

To examine a direct comparison between the postprocessing methods described in this study and one of the standard methods described in section 1, a set of UMOS temperature forecasts was obtained from CMC for the same forecast period: 1 April 2004–31 March 2005. UMOS forecasts were available only for a subset of stations consisting of six of the seven MSC stations, and were available only for temperature. Also, only the first 2 days of the forecast cycle have UMOS forecasts available. The postprocessing techniques listed in Table

1 were recalculated for this same subset of UMOS-available stations. The UMOS forecasts are included in this study to serve as a direct comparison among the various weighted-average techniques described earlier in this section.

b. Precipitation

Forecast verification for precipitation was performed for the 6-month period from 1 October 2004 through 31 March 2005, to encompass the local wet season. The verification period exhibited examples of extreme precipitation: the highest 24-h precipitation amount reported at a single site was 152 mm.

Observations of 24-h precipitation amounts were obtained from the observation network described at the beginning of section 2. Equivalent forecasts were obtained from NWP forecasts produced by the same CMC GEM model that produced the temperature forecasts described in section 2a. Point forecasts for each of the 19 sites were obtained for forecast days 1–8. Each forecast day was treated separately by applying techniques independently to a series of forecast–observation pairs valid for each particular forecast day (1–8) only.

Accuracy in precipitation forecasts is a much more challenging goal than for temperature forecasts for reasons discussed later in section 3b. Day-to-day variability in precipitation is much higher than for temperature; therefore, a longer averaging period is needed. In addition, sample mean error correction, as applied to the temperature forecasts, is not appropriate for precipitation because the resulting bias-corrected forecast could be negative. An appropriate error measure for quantitative precipitation forecasts is the degree of mass balance (DMB) between DMO and observations. DMB describes the ratio of the predicted to the observed net water mass for a given interval (Grubišić et al. 2005) and is given by

$$\text{DMB}_N = \frac{\sum_{k=1}^{N} x_k^f}{\sum_{k=1}^{N} y_k^o},\qquad(3)$$

where DMB_N is the degree of mass balance for the interval of N days, x_k^f is the 24-h precipitation forecast for day k, and y_k^o is the associated observation for day k.

DMB cannot be applied to a single day's forecast–observation pair because the operation would result in division by zero on days with no observed precipitation. DMB must cover a period of time for which some precipitation has been observed. For the current study, it was found that a period of 21 days was sufficient to ensure precipitation had occurred and that DMB could
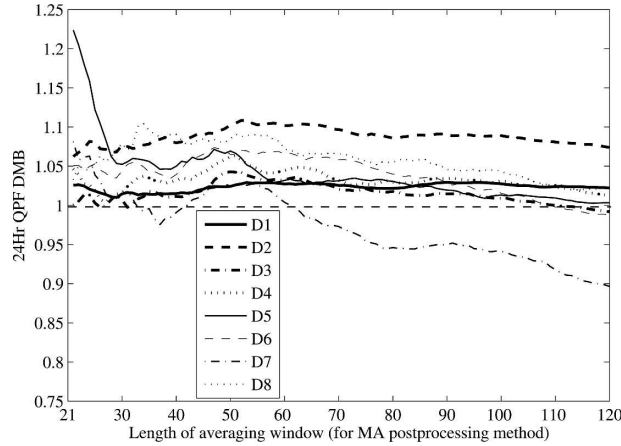
FIG. 3. The 24-h precipitation forecast degree of mass balance as a function of averaging window length for the MA postprocessing method. Calculations for each of the forecast days 1 through 8 are shown. DMB values near one are better since values greater than one indicate the degree of overforecasting precipitation, and values less than one indicate the degree of underforecasting precipitation.
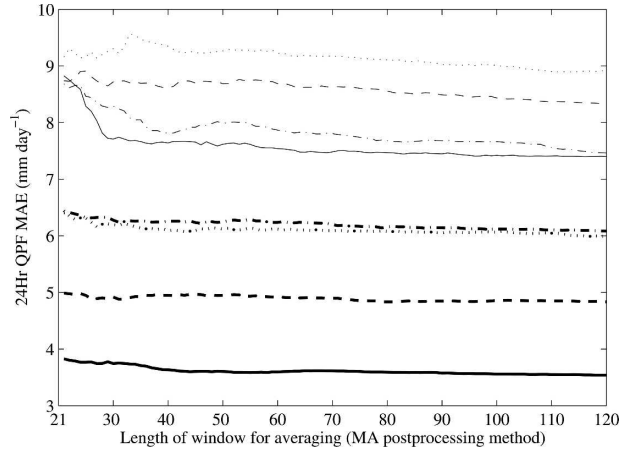

FIG. 4. The 24-h precipitation forecast MAE as a function of averaging window length for the MA postprocessing method. Calculations for each of the forecast days 1 (lowest MAE) through 8 (highest MAE) are shown. Lower values are better.

be evaluated. For the same reason, some of the bias-correction methods employed for temperature forecasts are not suitable for precipitation forecasts because they must be applied on a daily basis. The techniques suitable for DMB correction are SNL, MA, and BES.

A DMB-correction method should, ideally, correct model over- or underprediction of precipitation. DMB calculations with varying length of MA window from 21 to 120 days, for forecast days 1–8, are shown in Fig. 3. The results indicate that the MA technique corrects the DMB to within about 10% with a DMB-correction window of 30–40 days. By day 40 all forecast days have essentially reached a steady value between 1.0 and 1.1 DMB, indicating the method is producing forecasts that reflect the quantity of precipitation observed. There is an exception for forecast day 7, which drifts to lower DMB values at longer range. This drifting away from a steady value is not evident for the other forecast days and appears to be a statistical artifact for the day 7 forecast period only. Similar results (not shown) were obtained for the BES method.

It was also found that the postprocessing methods applied to precipitation showed a decrease in mean absolute error, reaching an asymptotic value by 40 days (see Fig. 4 for the test employing the MA method of DMB correction). The longer-range forecasts take longer to stabilize to a steady value than the shorter-range forecasts. By day 40 all forecast days have reached a steady MAE value. Therefore, a common value of 40 days was employed for the MA and BES methods.

DMB-corrected quantitative-precipitation forecasts, using the MA and BES methods, are calculated by

$$QPF^c = \frac{DMO}{DMB^N},\qquad(4)$$

where DMB^N is the DMB correction error applied to the current DMO forecast and QPF^c is the new DMB-corrected QPF forecast.

For 24-h precipitation forecasts, the SNL method averaged DMB over a constant 6-month period, and this factor was applied to each day's forecast. SNL was seasonally adjusted; that is, for forecasts issued for the wet season of 1 October–31 March, a constant wet-season DMB was calculated from the same period of the previous year and applied.

3. Results

a. Daily temperature

All seven postprocessing methods tested in this analysis indicated improvement over the DMO maximum and minimum daily temperature forecasts, for all forecast periods of days 1–8. Mean error (ME) and mean absolute error (MAE) were calculated to evaluate the forecasts using the following standard equations:

$$ME = \frac{1}{N}\sum_{k=1}^{N}(x_k^f - y_k^o)\quad\text{and}\qquad(5)$$

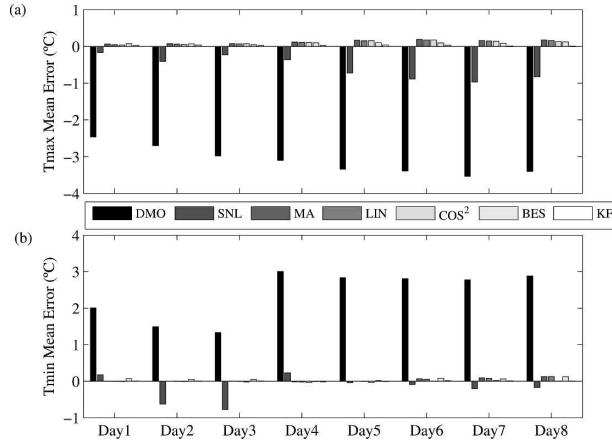$$MAE = \frac{1}{N}\sum_{k=1}^{N}|x_k^f - y_k^o|,\qquad(6)$$

FIG. 5. The daily maximum and minimum temperature mean errors from the sample for forecast days 1–8. The different postprocessing techniques, from left (black) to right (white) are DMO, SNL, MA, LIN, COS^2, BES, and KF, as shown in the legend. Zero mean error is best.

where N is the number of forecast–observation pairs, and x_k^f and y_k^o are individual forecasts and observations for day k, respectively.

In terms of mean error, the daily maximum temperature DMO forecast errors ranged from $-2.5°$ to $-3.5°C$ for days 1–8, respectively, indicating a cold bias to DMO maximum temperature forecasts (Fig. 5). Daily minimum temperature DMO forecast errors ranged from $+1.3°$ to $+3.0°C$, indicating a warm bias for DMO overnight temperature forecasts. All postprocessing methods show a significant reduction in DMO mean error for all forecast days. All methods except SNL reduce the DMO mean error to near zero for all forecast days.

The trend in forecast bias for minimum temperatures was unusual in that bias decreased from $+2.0°$ to $+1.3°C$ for days 1–3, then increased significantly to $+3.0°C$ on day 4; bias then remained nearly constant at $+2.8°C$ for days 5–8. The increase in DMO minimum temperature bias between days 3 and 4 may be due to the fact that forecast temperatures for days 1 and 2, and the first half of day 3 come from the higher-resolution regional GEM model (a 48-h forecast model), while the DMO forecast temperatures for the second half of day 3 through day 8 come from the lower-resolution version of the GEM global model (termed the global or GLB model that runs operationally out to a lead time of at least 10 days). The model grid change between days 3 and 4 may account for some of the sudden increase in minimum temperature bias after day 3, since a large part of the bias error is attributable to poor representation of topography in the model. No such change is seen in the maximum temperature forecast bias, however.
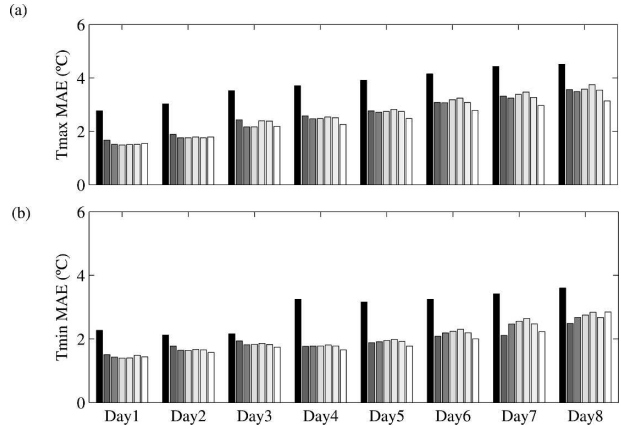

FIG. 6. The same as in Fig. 5 but for MAE. Zero MAE is best.

All postprocessing methods reduced the mean error in the DMO forecasts. The four moving-weighted-average methods (MA, LIN, COS^2, and BES) and the Kalman filter method (KF) were nearly equal in keeping mean errors under $0.2°C$ for both maximum and minimum temperature forecasts. These methods respond rapidly to changes in the model and changes in flow regime. The seasonal method of mean error reduction (SNL) maintains the same error correction factor throughout a complete 6-month period (cold or warm season). The SNL method shows gradually increasing negative mean error (for maximum temperature forecasts), with a mean error near $-0.8°C$ for days 5–8, though this method still greatly outperformed DMO. The difference in performance between SNL and the other methods, though slight, may be due to the residual impact of model changes or flow regime changes that are not readily incorporated in this method.

Mean absolute error (MAE) was evaluated for all of the postprocessing methods (Fig. 6). All postprocessing methods show significant improvement in MAE over the DMO forecasts, especially for short-range daily maximum temperature forecasts. The KF technique performs best by a slight margin. MAE was generally greater for maximum temperature forecasts than for minimum temperature forecasts, and generally increased from days 1 to 8. In terms of MAE, all six methods (SNL, MA, LIN, COS^2, BES, and KF) performed nearly equally well in keeping MAE between $1.5°$ and $3.5°C$ throughout the forecast period.

To show these results more clearly, an MAE skill score, measured with DMO as the reference forecast, is shown in Fig. 7. The MAE skill score is adapted from Wilks (1995) and Mao et al. (1999):

$$\text{MAE skill score} = 1 - \frac{\text{MAE}_{PP}}{\text{MAE}_{DMO}}. \quad (7)$$

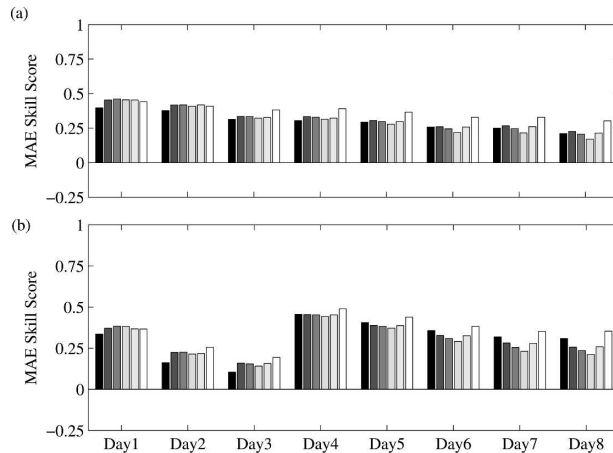FIG. 7. The same as in Fig. 6 but for MAE skill score (relative to DMO). Skill-score values closer to one are better.
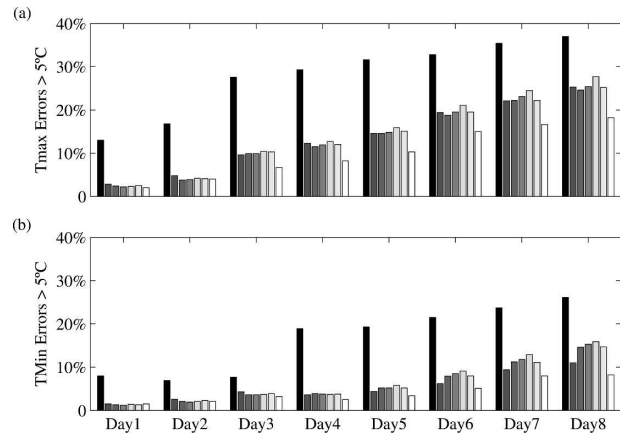

FIG. 8. The daily maximum and minimum temperature absolute errors greater than 5°C (%) from the sample for forecast days 1–8. The different postprocessing techniques, from left (black) to right (white) are DMO, SNL, MA, LIN, COS², BES, and KF. Values closer to zero are better.

In this formulation MAE_{PP} represents the MAE of the particular postprocessing technique, as indicated in the text and in Fig. 7. An MAE skill score value of zero indicates no improvement in skill over the associated DMO forecast; the range from zero to one indicates improving skill, with a value of one indicating perfect forecasting skill. A negative MAE skill score indicates the postprocessing technique shows less skill than the corresponding DMO forecast.

One trend visible in Fig. 7 is that the SNL technique shows improving skill with lead time, relative to the moving-weighted-average and KF methods. As one would expect, early in the forecast period (days 1–4) the weighted-average techniques hold a slight advantage over SNL because they are weighted by recent error estimates that may be affected by weather regime changes or, infrequently, model changes that the SNL technique does not incorporate. At longer lead times (days 5–8) the inclusion of recent error becomes less effective because overall random error in the forecast increases so much at the longer lead times; the statistical stability of a 6-month seasonal error average shows its advantage at the longer lead times.

For some applications of temperature forecasting, an error threshold is important, in that the forecasts are expected to remain within a certain critical error threshold. For example, agencies or commercial providers may have guidelines or contractual stipulations that require temperature forecast errors to not exceed a particular threshold for a certain percentage of time without incurring a penalty. The postprocessing techniques evaluated here were also tested against critical error thresholds of 5° and 10°C (see Figs. 8 and 9). All postprocessing techniques show a significant reduction in errors over DMO forecasts at both thresholds, through the 8-day forecast period. The KF technique

indicates the best results, especially in the longer forecast range.

Daily maximum temperature errors are greater than minimum temperature errors, with forecast errors greater than 5°C occurring less than 10% of the time through day 3 and less than 25% of the time through day 8. Daily minimum temperature errors greater than 5°C remain below 10% through day 6.

Daily minimum temperature errors greater than 10°C are rare even for DMO forecasts, though the methods presented here still achieve improvement in these forecasts. Daily maximum temperature errors greater than 10°C occur less than 1% of the time through day 5 for all methods, increasing to occurrences about 3% of the time by day 8; these methods show much improvement over DMO forecasts, which indicate 10°C errors occur 5%–10% of the time for forecast days 5–8.
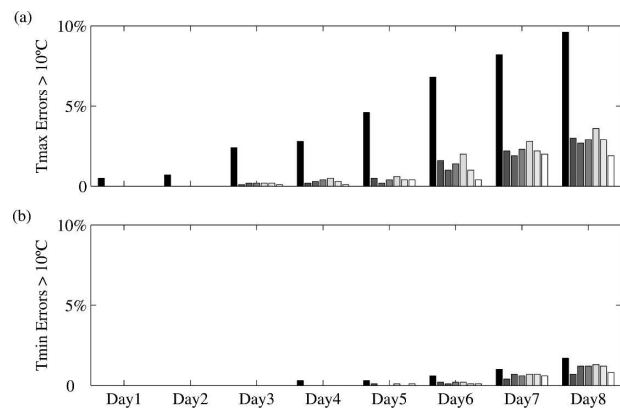

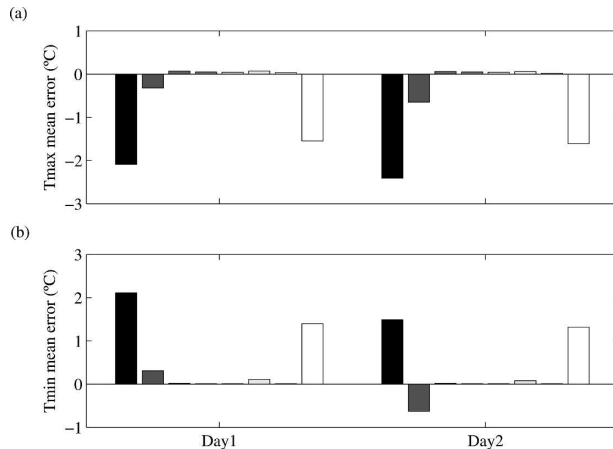FIG. 9. The same as in Fig. 6 but for temperature absolute errors greater than 10°C (%).

FIG. 10. The daily maximum and minimum temperature ME from the UMOS subsample for forecast days 1 and 2. The different postprocessing techniques, from left (black) to right (white) are DMO, SNL, MA, LIN, COS^2, BES, KF, and UMOS. Zero mean error is best.

UMOS

The UMOS forecasts were available only for a subset of six stations. The weighted-average methods (listed in Table 1) were recalculated for this subset of stations to allow for direct comparison with the UMOS forecasts. The UMOS forecasts were the poorest among all of the methods at reducing mean error, with mean error as great as 1.8°C (Fig. 10).

MAE comparisons of UMOS with weighted-average techniques (Fig. 11) indicate that UMOS forecasts fare worse than the other postprocessing methods for both maximum temperature forecasts and minimum temperature forecasts. Stensrud and Yussouf (2003) found similar results with temperature forecast errors corrected with a simple 7-day running mean that proved competitive with, or better than, MOS forecasts. Results from Stensrud and Yussouf (2005) show that tem-
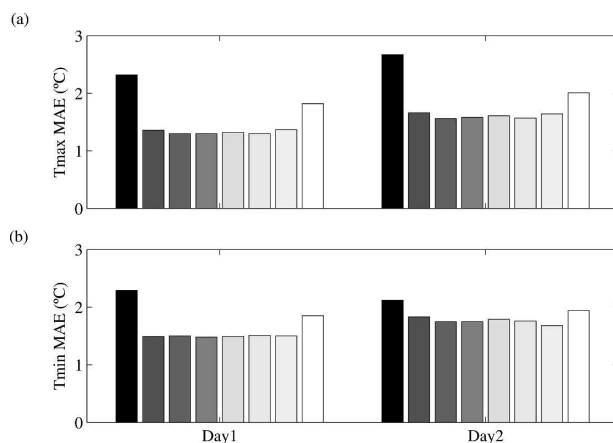

FIG. 11. The same as in Fig. 10 but for temperature MAE. Zero MAE is best.

perature forecast errors corrected with a 12-day moving average have lower error than comparable MOS forecasts.

b. Daily precipitation

One of the most long-standing and challenging problems in weather forecasting is the quantitative prediction of precipitation (Mao et al. 2000; Chien and Ben 2004; Barstad and Smith 2005; Zhong et al. 2005), and for hydrometeorological applications, precipitation is the most important factor in watershed modeling (Beven 2001). The challenge of precipitation forecasting is often exacerbated in regions with complex terrain (Westrick and Mass 2001). This difficulty in forecasting precipitation is unfortunate since forecasting for water resource planning, flash floods, and glacier mass balance in mountainous terrain depends on accurate weather forecast models. Inflow forecasting, based largely on accurate precipitation forecasts from NWP models, is critical for reservoir management (Yao and Georgakakos 2001).

There are many reasons that quantitative-precipitation forecasting is much more challenging than forecasting daily maximum or minimum temperatures, including the following.

- The basic moisture variable in NWP models is specific humidity while temperature itself is a basic variable in primitive equation models. Instantaneous precipitation rate in NWP is a parameterized variable encapsulating many complex physical processes. QPF is an integrated compilation of precipitation rate. Precipitation requires a finite spinup time for operational NWP models that encompass a dry start.
- Precipitation is discontinuous in space and time; temperature is a continuous space–time variable.
- Precipitation exhibits a nonnormal sampling distribution; maximum and minimum temperatures tend toward a normal distribution.
- Daily precipitation amounts are highly variable from one day to the next; daily temperatures less so.
- Daily temperature falls into a natural 24-h evaluation window because of diurnal temperature trends. Cool season stratiform precipitation stems from evolving midlatitude storm systems that do not have such a natural 24-h cycle.
- Precipitation occurs in different phases and types.
- Varying rates of horizontal advection and vertical fall speed occur for different sizes and compositions of hydrometeors, which affects collection efficiency in rain gauge observations.
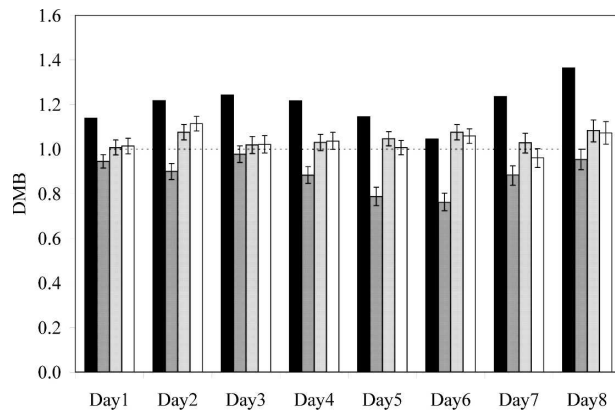- Precipitation in NWP models is calculated as a grid-square average; direct precipitation observation is by

FIG. 12. The daily 24-h precipitation DMB from the sample for forecast days 1–8. For precipitation the DMB is calculated as forecasts divided by observations so that balanced forecasts have a value of one. The different postprocessing techniques, from left (black) to right (white) are DMO, SNL, MA, and BES. Values closer to one are better. For statistical significance analysis, error bars are included to indicate the 2.5th and 97.5th percentiles of a resampled distribution, referenced to the DMO forecasts.

point (radar-derived precipitation totals are areal in nature; however, the derived values depend on subjective algorithms that are difficult to implement in mountainous terrain due to blocking of the radar beam). The difference in scale between a mesoscale model forecasting precipitation on a 10-km grid spacing and a typical precipitation gauge (1 m^2 opening) is eight orders of magnitude.

- Precipitation observation technologies are more subject to errors due to equipment limitations than temperature recording technology (e.g., tipping-bucket gauges are susceptible to upper-rain-rate limits and below-freezing temperatures; rain bucket gauges are susceptible to snow capping; all gauges are susceptible to misrepresenting hydrometeor fall rates in strong winds).

DMO for QPF shows a slight overforecasting (wet) trend throughout the 8-day forecast period, averaged over all 19 stations (see Fig. 12). The SNL method of DMB correction overcompensates by reducing the DMB to less than one for all forecast days. The MA and BES methods perform equally well in reducing DMB error in the DMO forecasts. The resulting QPF-adjusted values from the MA and BES methods result in well-balanced DMB values (near one) for all forecast days—the main objective of the postprocessing application. A nonparametric statistical significance test was performed to confirm the results. Error bars in Fig. 12 represent the 95% confidence limits for the SNL, MA, and BES methods. Details of the statistical significance test methodology are given in appendix B.

Closer inspection, however, shows that the DMO forecasts have a much wider range of error determined station to station compared to SNL, MA, and BES. For example, on forecast day 2 the DMO DMB ranges from 0.57 to 2.3, while SNL ranges from 0.36 to 1.27, MA ranges from just 0.97 to 1.27, and BES ranges from 0.92 to 1.35 (see Fig. 13). This diagram shows that the MA and BES methods are similar in correcting DMB values to near one, outperforming the SNL method in a station-by-station comparison. DMO precipitation forecast errors as shown are significant for certain stations in this region of complex terrain and require appropriate postprocessing measures to reduce topographically forced systematic errors in the forecasts.

Examination of the output for other forecast days indicates similar results. Therefore, even though the average DMB over all stations is fairly comparable for DMO compared to the other methods, either MA or BES would prove a good choice for a postprocessing method based on station-to-station DMB consideration.

Considering the MAEs of daily precipitation forecasts (Fig. 14), all of the postprocessing techniques show improvement over DMO throughout the 8-day forecast period. All of the techniques show similar improvement for days 1–4, with SNL showing slightly less MAE (better performance) on days 2–4. From forecast days 5–8 the SNL method clearly proves the best method (based on MAE), pointing to the value of a longer seasonal training period in improving midrange DMO precipitation forecasts.

An MAE skill score chart (Fig. 15) shows this trend more clearly in a different format. The MA and BES methods, relying on just the most recent 40 days for an error correction window, show very slight MAE skill relative to DMO for the day 5–8 forecasts (the methods show less than 8% improvement in MAE skill over DMO, with no skill for day 6). The long seasonal correction window employed by the SNL technique is necessary to improve the midrange precipitation forecasts (indicating about a 12% MAE skill improvement over DMO forecasts). This MAE skill score diagram shows the results evident in Fig. 14 more clearly.

Daily precipitation thresholds of 10 and 25 mm were chosen to show the improvement of the forecasts over DMO by using the stated postprocessing techniques (see Figs. 16 and 17). All techniques show slight improvement over DMO at both threshold criteria. Similar to the MAE analysis, the MA and BES methods perform best for the day 1 forecast, then the SNL method performs best for the remainder of the forecast days 2–8 (though by a very slight margin).
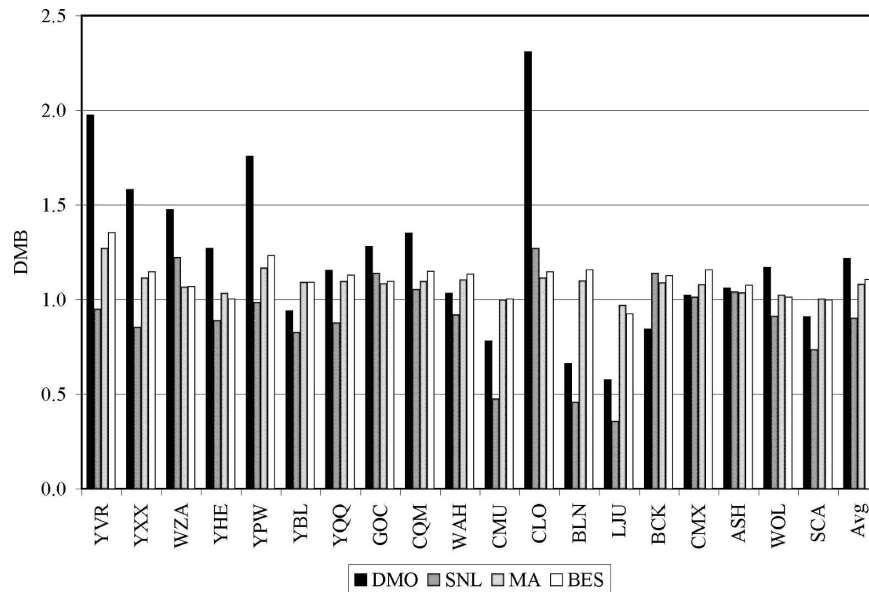
FIG. 13. A comparison of 24-h precipitation DMB values for each of the 19 stations for the day 2 forecast. The different postprocessing methods are (as labeled) DMO, SNL, MA, and BES. Values closer to one are better.

4. Summary and conclusions

Daily forecast values of maximum temperature, minimum temperature, and quantitative precipitation are the prime drivers of inflow forecasts for the 15 reservoirs within the area of study described in this paper, with precipitation being the most important factor. Seven statistical postprocessing techniques have been compared here for maximum and minimum temperature forecasts, and three methods for 24-h quantitative precipitation forecasts. For this study in mountainous western Canada, all of the techniques for temperature postprocessing showed significant improvement over direct model output. The best method was Kalman filtering, followed closely by the four 14-day moving-weighted-average methods (moving average, linear weighted average, cosine squared weighted average, and best easy systematic estimator). A method based on seasonally averaged error characteristics also showed similar positive error reduction results, especially in the longer forecast period: days 5–8. All of the methods performed better than a comparative study against updateable MOS temperature forecasts.

All three postprocessing methods for 24-h QPFs improved error characteristics over DMO forecasts. The best method for QPF, using DMB calibration to unity as a metric, was shared by the moving-average method and the best easy systematic estimator method (both based on a 40-day averaging period). However, the sea-
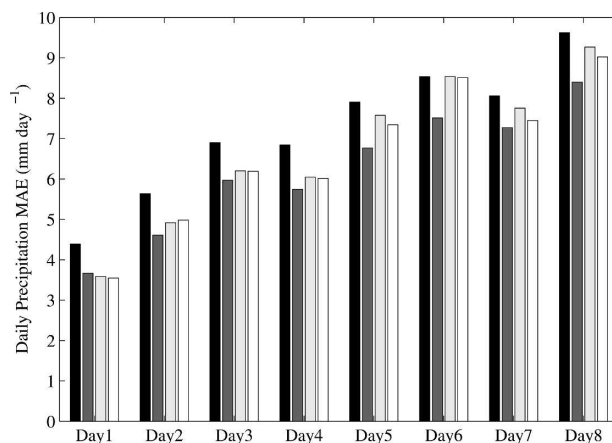

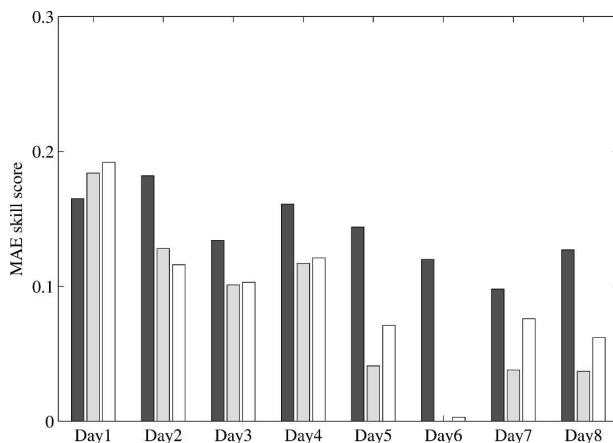FIG. 14. The same as in Fig. 12 but for precipitation MAE. Values closer to zero are better.

FIG. 15. The same as in Fig. 14 but for precipitation MAE skill score (relative to DMO). Values closer to one are better.
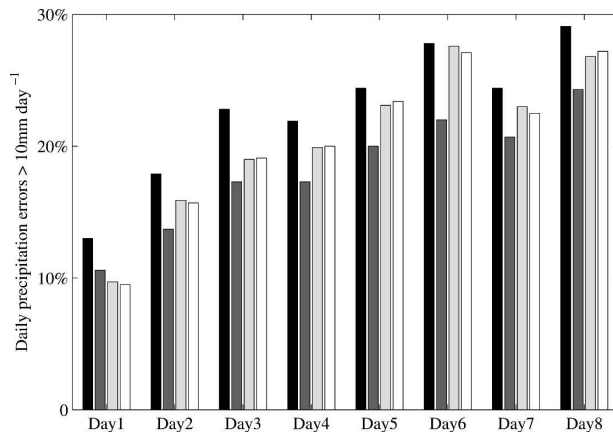
FIG. 16. The daily 24-h precipitation errors greater than 10 mm day^{-1} (%) from the sample for forecast days 1–8. The different postprocessing techniques, from left (black) to right (white) are DMO, SNL, MA, and BES.

sonal method had slightly better error reduction characteristics judging by MAE and particular error thresholds in the day 2–8 period.

The postprocessing methods tested in this study can provide requisite error reduction in DMO for local point forecasts to aid decision makers in hydrometeorological and other economic or regulatory sectors. Specifically, water resource managers rely on weather forecasts of precipitation and temperature to drive hydrologic reservoir inflow models as a major component of their decision-making process. Decision makers rely on current, value-added weather forecasts for daily reservoir operations planning, extreme high-inflow events, and long-term water resource management. Such forecasts present added challenges in regions of complex topography where steep mountains and land–ocean boundaries increase the variability of local weather.
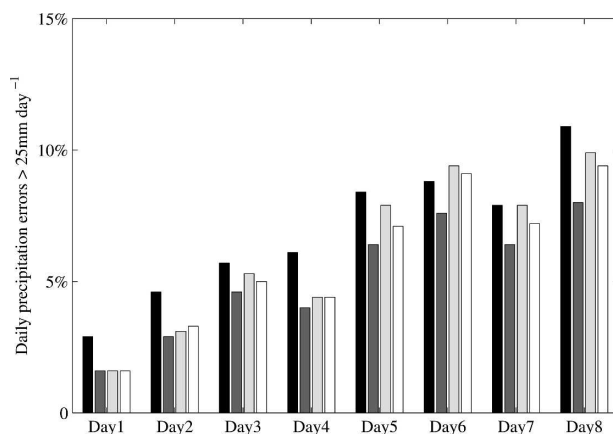

FIG. 17. The same as in Fig. 16 but for precipitation errors greater than 25 mm day^{-1} (%).

Future work

Future work to improve the techniques described here would include a multivariate approach to weighting error estimates. Including variables such as integrated water vapor and flow direction in a multivariate approach would likely prove beneficial in a postprocessing technique to improve precipitation forecasts. However, much more training data is needed for multivariate approaches that calibrate forecasts conditioned on specific events.

A multimodel approach is also a cost-effective way to reduce error in forecasts. A combination of a model-averaging approach (Ebert 2001) and the bias-correction techniques presented here may prove beneficial (Delle Monache et al. 2006a), especially for precipitation forecasting.

Another approach that may provide positive results is autoregressive-moving average (ARMA) modeling of DMO error. ARMA models of meteorological time series gained popularity in the late 1970s and early 1980s, after Box and Jenkins (1976) published a readily accessible statistical methodology for applying ARMA models to time series analysis. However, meteorological ARMA models of the time were limited to seasonal analysis applications such as drought index modeling, monthly precipitation, and annual streamflow (Katz and Skaggs 1981). ARMA techniques do provide a method to estimate future error between DMO and observations using optimized weighted past error measurements. Currently, however, major efforts in using ARMA and Box–Jenkins approaches are concentrated in the financial world. Complex ARMA schemes depend critically on developmental data (as does MOS) and hence perform poorly after changes to the underlying model.

APPENDIX A

Kalman Filter Bias Correction

The Kalman filter (KF) is a recursive, adaptive technique for estimating a signal from noisy measurements (Bozic 1994; Zarchan and Musoff 2005). Kalman filter theory provides recursive equations for continuously updating error estimates via observations of a system involving inherently unknown processes. The KF has been employed as a predictor bias-correction method

during postprocessing of short-term NWP forecasts. Thorough descriptions of KF theory, equations, and applications to weather forecast models can be found in the meteorological literature (Homleid 1995; Roeger et al. 2003; Delle Monache et al. 2006b). The KF approach is adapted here for daily maximum and minimum temperature forecasts.

Recent bias values are used as input to the KF. The filter estimates the bias in the current forecast. As in the other postprocessing methods examined in this study, the expected bias in the current forecast as estimated by the KF is removed from the current DMO forecast. The new corrected forecast should have improved error characteristics over the original DMO forecast.

Let x_k be the bias between the forecast and the verifying observation valid for time step k. The bias x_k is the signal we would like to predict for the next forecast period (at $k + 1$). The future bias is determined by a persistence of the current bias plus a Gaussian-distributed random term w_k of variance σ_w^2: $x_{k+1} = x_k + w_k$.

Similarly, the input observations y_k are assumed to be noisy, with a random error term of v_k^2 (of variance σ_v^2): $y_k = x_k + v_k$. The objective is to get the best estimate of x_k, which is termed $\hat{x}_k$, by minimizing the expected mean-square error: $p = E[(x - \hat{x})^2]$.

The recursive nature of the technique is characterized by a continuous update–predict cycle. The measurement update, or "corrector" portion of the cycle, is determined by the following equations.

Compute the Kalman gain β:

$$\beta_k = \frac{p_{k-1} + \sigma_w^2}{p_{k-1} + \sigma_w^2 + \sigma_v^2}. \tag{A1}$$

Update the estimate $\hat{x}_k$:

$$\hat{x}_k = \hat{x}_{k-1} + \beta_k[y_k - \hat{x}_{k-1}]. \tag{A2}$$

Finally, update the error covariance term p_k:

$$p_k = (p_{k-1} + \sigma_w^2)(1 - \beta_k). \tag{A3}$$

The ratio r is defined as (σ_w^2/σ_v^2). A value of $r = 0.01$ is used in this study, as suggested from previous studies (Roeger et al. 2003; Delle Monache et al. 2006b).

The time update, or "predictor" portion of the cycle, is determined by

$$\hat{x}_{k+1} = \hat{x}_k \quad \text{and} \tag{A4}$$

$$p_{k+1} = p_k + \sigma_w^2. \tag{A5}$$

Initial starting values of $\hat{x}(0)$ and $p(0)$ are chosen to start the process. The equations converge quickly so that the results are not sensitive to the particular initial values that begin the process.

APPENDIX B

Statistical Significance Test for Precipitation Forecasts

Statistical significance tests are formulated to determine the likelihood that an observed outcome could have arisen by chance instead of design. For the 24-h QPFs analyzed in this study, a hypothesis test methodology designed for precipitation forecasts [see Hamill (1999) for details of the method] was employed.

The technique is based on resampling. The null hypothesis for the resampling test is that the difference in DMB between each of the SNL, MA, and BES forecasts and the baseline DMO forecast is zero,

$$H_0: \text{DMB}_{\text{PP}} - \text{DMB}_{\text{DMO}} = 0.0, \tag{B1}$$

and the alternative hypothesis,

$$H_A: \text{DMB}_{\text{PP}} - \text{DMB}_{\text{DMO}} \neq 0.0, \tag{B2}$$

where DMB_{PP} is the DMB of the postprocessing method SNL, MA, or BES, and DMB_{DMO} is the DMB of the direct model output. Assume a nonsymmetric two-sided test with significant level $\alpha = 0.05$. Create a virtual time series by resampling the existing set of forecasts to form a test statistic consistent with the null hypothesis above. The test statistic,

$$(\widehat{\text{DMB}_{\text{PP}}} - \widehat{\text{DMB}_{\text{DMO}}}), \tag{B3}$$

is calculated using Eq. (3) in section 2b. The virtual ($^*_1$) resampled test statistic time series consistent with the null hypothesis is generated by randomly selecting either the postprocessed forecast or the DMO baseline forecast for each day and calculating the DMB for this time series. A second virtual ($^*_2$) randomly selected time series is formed using the alternate selection of postprocessed and DMO forecasts. The resampled test statistic,

$$(\widehat{\text{DMB}_1^*} - \widehat{\text{DMB}_2^*}), \tag{B4}$$

is repeated 1000 times to build a null distribution.

The final step is to determine whether $(\widehat{\text{DMB}_{\text{PP}}} - \widehat{\text{DMB}_{\text{DMO}}})$ can be considered to fall within the distribution of $(\widehat{\text{DMB}_1^*} - \widehat{\text{DMB}_2^*})$ values (the null hypothesis), or whether the null hypothesis can be rejected. The virtual time series resampled distributions are utilized to compute the locations $\widehat{t_L}$ and $\widehat{t_U}$ such that

$$\text{Pr}^*[(\widehat{\text{DMB}_1^*} - \widehat{\text{DMB}_2^*}) - \widehat{t_L}] = \frac{\alpha}{2}, \quad \text{and}$$

$$\text{Pr}^*[(\widehat{\text{DMB}_1^*} - \widehat{\text{DMB}_2^*}) - \widehat{t_U}] = 1 - \frac{\alpha}{2}, \tag{B5}$$

where Pr* represents the probabilities numerically calculated from this distribution. Then, H_0 is rejected if

$$(\widehat{DMB}_{PP} - \widehat{DMB}_{DMO}) < \widehat{t_L} \quad \text{or} \quad (\widehat{DMB}_{PP} - \widehat{DMB}_{DMO}) > \widehat{t_U}.$$

In graphical displays of the postprocessed forecast results, the values of t_U and t_L are represented by statistical significance error bars. Forecast differences outside the interval expressed by the error bars may be considered statistically significant for the particular α.

REFERENCES

Barstad, I., and R. B. Smith, 2005: Evaluation of an orographic precipitation model. *J. Hydrometeor.,* **6,** 85–99.

Beven, K. J., 2001: *Rainfall-Runoff Modelling: The Primer.* John Wiley, 360 pp.

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis: Forecasting and Control* (rev.). Holden-Day, 575 pp.

Bozic, S. M., 1994: *Digital and Kalman Filtering: An Introduction to Discrete-Time Filtering and Optimum Linear Estimation.* 2nd ed. John Wiley, 160 pp.

Chien, F., and J. J. Ben, 2004: MM5 ensemble mean precipitation forecasts in the Taiwan area for three early summer convective (mei-yu) seasons. *Wea. Forecasting,* **19,** 735–750.

Clark, M. P., and L. E. Hay, 2004: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J. Hydrometeor.,* **5,** 15–32.

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.,* **126,** 1373–1395.

——, ——, ——, ——, ——, ——, and ——, 1998b: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.,* **126,** 1397–1418.

Delle Monache, L., X. Deng, Y. Zhou, and R. Stull, 2006a: Ozone ensemble forecasts: 1. A new ensemble design. *J. Geophys. Res.,* **111,** D05307, doi:10.1029/2005JD006310.

——, T. Nipen, X. Deng, Y. Zhou, and R. Stull, 2006b: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *J. Geophys. Res.,* **111,** D05308, doi:10.1029/2005JD006311.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.,* **129,** 2461–2480.

Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting,* **20,** 328–350.

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

Grubišić, V., R. K. Vellore, and A. W. Huggins, 2005: Quantitative precipitation forecasting of wintertime storms in the Sierra Nevada: Sensitivity to the microphysical parameterization and horizontal resolution. *Mon. Wea. Rev.,* **133,** 2834–2859.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting,* **14,** 155–167.

——, S. L. Mullen, C. Snyder, Z. Toth, and D. P. Baumhefner, 2000: Ensemble forecasting in the short to medium range: Report from a workshop. *Bull. Amer. Meteor. Soc.,* **81,** 2653–2664.

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

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting,* **10,** 689–707.

Jacks, E., J. B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson, and J. C. Su, 1990: New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting,* **5,** 128–138.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Katz, R. W., and R. H. Skaggs, 1981: On the use of autoregressive-moving average processes to model meteorological time series. *Mon. Wea. Rev.,* **109,** 479–484.

Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Atmos. Sci.,* **16,** 672–682.

Majewski, D., 1997: Operational regional prediction. *Meteor. Atmos. Phys.,* **63,** 89–104.

Mao, Q., R. T. McNider, S. F. Mueller, and H. H. Juang, 1999: An optimal model output calibration algorithm suitable for objective temperature forecasting. *Wea. Forecasting,* **14,** 190–202.

——, S. F. Mueller, and H. H. Juang, 2000: Quantitative precipitation forecasting for the Tennessee and Cumberland River watersheds using the NCEP Regional Spectral Model. *Wea. Forecasting,* **15,** 29–45.

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.

Stensrud, D. J., and J. A. Skindlov, 1996: Gridpoint predictions of high temperature from a mesoscale model. *Wea. Forecasting,* **11,** 103–110.

——, and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England. *Mon. Wea. Rev.,* **131,** 2510–2524.

——, and ——, 2005: Bias-corrected short-range ensemble forecasts of near surface variables. *Meteor. Appl.,* **12,** 217–230.

Westrick, K. J., and C. F. Mass, 2001: An evaluation of a high-resolution hydrometeorological modeling system for prediction of a cool-season flood event in a coastal mountainous watershed. *J. Hydrometeor.,* **2,** 161–180.

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

Wilson, L. J., and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) system: Design and development tests. *Wea. Forecasting,* **17,** 206–222.

——, ——, 2003: The Canadian Updateable Model Output Statistics (UMOS) system: Validation against perfect prog. *Wea. Forecasting,* **18,** 288–302.

Woodcock, F., and C. Engel, 2005: Operational consensus forecasts. *Wea. Forecasting,* **20,** 101–111.

Yao, H., and A. Georgakakos, 2001: Assessment of Folsom Lake response to historical and potential future climate scenarios 2. Reservoir management. *J. Hydrol.,* **249,** 176–196.

Zarchan, P., and H. Musoff, 2005: *Fundamentals of Kalman Filtering: A Practical Approach.* 2nd ed. American Institute of Aeronautics and Astronautics, 764 pp.

Zhong, S., H.-J. In, X. Bian, J. Charney, W. Heilman, and B. Potter, 2005: Evaluation of real-time high-resolution MM5 predictions over the Great Lakes region. *Wea. Forecasting,* **20,** 63–81.