

Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction

LE BAO

University of Washington, Seattle, Washington

TILMANN GNEITING*

Heidelberg University, Heidelberg, Germany

ERIC P. GRIMIT

3TIER, Inc., Seattle, Washington

PETER GUTTORP AND ADRIAN E. RAFTERY

University of Washington, Seattle, Washington

(Manuscript received 9 July 2009, in final form 14 October 2009)

ABSTRACT

Wind direction is an angular variable, as opposed to weather quantities such as temperature, quantitative precipitation, or wind speed, which are linear variables. Consequently, traditional model output statistics and ensemble postprocessing methods become ineffective, or do not apply at all. This paper proposes an effective bias correction technique for wind direction forecasts from numerical weather prediction models, which is based on a state-of-the-art circular–circular regression approach. To calibrate forecast ensembles, a Bayesian model averaging scheme for directional variables is introduced, where the component distributions are von Mises densities centered at the individually bias-corrected ensemble member forecasts. These techniques are applied to 48-h forecasts of surface wind direction over the Pacific Northwest, using the University of Washington mesoscale ensemble, where they yield consistent improvements in forecast performance.

1. Introduction

Forecasts of wind direction have varied and important uses, ranging from air pollution management to aircraft and ship routing and recreational boating. However, wind direction is an angular variable that takes values on the circle, as opposed to other weather quantities, such as temperature, quantitative precipitation, or wind speed, which are linear variables that take values on the real line. As a result, traditional postprocessing techniques for forecasts from numerical weather prediction models tend

to become ineffective or inapplicable. For example, Engel and Ebert (2007, p. 1351) note that bias correction was “found not to be beneficial to wind direction forecasts”. The purpose of this paper is to develop effective bias correction and ensemble calibration techniques that are tailored to wind direction by taking the angular nature of the variable into account.

The remainder of the paper is organized as follows. In section 2 we describe our approach to bias correction and ensemble calibration in detail. We adopt the circular–circular regression approach of Downs and Mardia (2002) and Kato et al. (2008), and develop a Bayesian model-averaging scheme for directional variables, where the component distributions are von Mises densities centered at the individually bias-corrected ensemble member forecasts. Section 3 provides a case study on 48-h forecasts of surface wind direction over the Pacific Northwest in 2003 using the University of Washington mesoscale ensemble (Grimit and Mass 2002; Eckel and

* Additional affiliation: University of Washington, Seattle, Washington.

Corresponding author address: Tilmann Gneiting, Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany.
E-mail: t.gneiting@uni-heidelberg.de

Mass 2005). In these experiments, our methods turn out to be effective and yield consistent improvement in forecast performance. The paper closes with a discussion in section 4.

2. Methods

Wind direction is an angular variable that takes values on the circle, and as such can be represented in various equivalent ways. We use degrees to describe predicted and observed wind directions, with 0° , 90° , 180° , and 270° denoting a northerly, easterly, southerly, and westerly wind, respectively. The angular distance or circular absolute error,

$$\text{AE}_{\text{circ}}(f, v) = \min(|v - f|, 360 - |v - f|), \quad (1)$$

between two directions $0 \leq f, v < 360$ then is a non-negative quantity with a maximum of 180° . Occasionally, it will be useful to identify a direction, v , with the point

$$\theta(v) = e^{i\pi v/180}$$

on the unit circle in the complex plane. Under this one-to-one mapping, directions of 0° , 90° , 180° , and 270° correspond to 1, the imaginary unit i , -1 , and $-i$, respectively (i , 1) and $(-i, -1)$.

a. Bias correction

Systematic biases are substantial in dynamic modeling systems (Atger 2003; Mass 2003), and bias correction is an essential and well-established step in weather forecasting. The predominant approach is based on regression, using model output statistics (MOS) schemes based on multiple linear regression for linear variables, such as temperature or pressure, and logistic regression for binary variables, such as precipitation occurrence or freezing (Glahn and Lowry 1972; Wilks 2006a). For wind, the traditional approach is to develop separate MOS equations for the zonal and meridional components, and derive single-valued forecasts of wind speed and wind direction from them. However, this does not take the dependencies between the wind components into account and can lead to biases (Carter 1975; Glahn and Unger 1986). Thus, we take a different approach to predicting wind direction, and propose the use of a state-of-the-art circular–circular regression technique.

Specifically, let f and v denote the predicted and observed wind direction, respectively. Let $\theta(f)$ and $\theta(v)$ denote the associated points on the unit circle in the complex plane, as described above. Downs and Mardia (2002) and Kato et al. (2008) propose a regression equation of the form

$$\theta(v) = \beta_0 \frac{\theta(f) + \beta_1}{1 + \bar{\beta}_1 \theta(f)}, \quad (2)$$

where β_0 is a complex number with modulus $|\beta_0| = 1$, β_1 is any complex number and the bar denotes complex conjugation. The mapping from $\theta(f)$ to $\theta(v)$ is a Möbius transformation in the complex plane, which is one-to-one, and maps the unit circle to itself. The regression parameters β_0 and β_1 need to be estimated from training data. While β_0 is a rotation parameter, β_1 can be interpreted as pulling a direction toward a fixed angle, namely the point $\beta_1/|\beta_1|$ on the unit circle, with the concentration about this point increasing as $|\beta_1|$ increases, except that the antipodal point remains fixed (Kato et al. 2008). Figure 1 provides an illustration. Figure 1a shows twelve equally spaced nontransformed angles. Figure 1b illustrates the Möbius transformation (2) when $\beta_0 = 1$ (i.e., no rotation) and $\beta_1 = 0.3i$ (i.e., a contraction toward the imaginary unit in the complex plane, which corresponds to a direction of 90°). In Fig. 1c, the parameter values are $\beta_0 = 1$ and $\beta_1 = 0.6i$, resulting in a stronger contraction. Figure 1d uses $\beta_0 = -i$ and $\beta_1 = 0.6i$, thereby compositing the contraction with a counterclockwise rotation by 90° .

In our circular–circular regression approach to bias correction for wind direction, we estimate the Möbius transformation (2) from training data by numerically minimizing the sum of the circular distances between the fitted bias-corrected forecasts and the respective verifying directions as a function of the regression parameters.

For comparison, we consider two reference techniques. The first is median-angle correction, which arises as the special case of circular–circular regression in which the parameter $\beta_1 = 0$ is fixed. Then the Möbius transformation (2) is simply a rotation. In our minimum circular distance approach to estimation, the rotation parameter β_0 becomes the circular median of the directional errors in the training data. The second reference technique is mean-angle correction; that is, a rotation by the circular mean of the directional errors in the training data.

Recall here that if the training data comprises the pairs $(f_1, v_1), \dots, (f_n, v_n)$ of predicted and observed directions, the median of the directional errors is the angle m that minimizes

$$\sum_{k=1}^n \text{AE}_{\text{circ}}(v_k - f_k, m),$$

which is equivalent to the definition in Eq. (2.32) of Fisher (1993, p. 36). The circular mean is obtained by forming the vector sum of the directional errors, each of which is represented as a unit vector in the complex plane, and rescaling to the unit circle, or equivalently, by

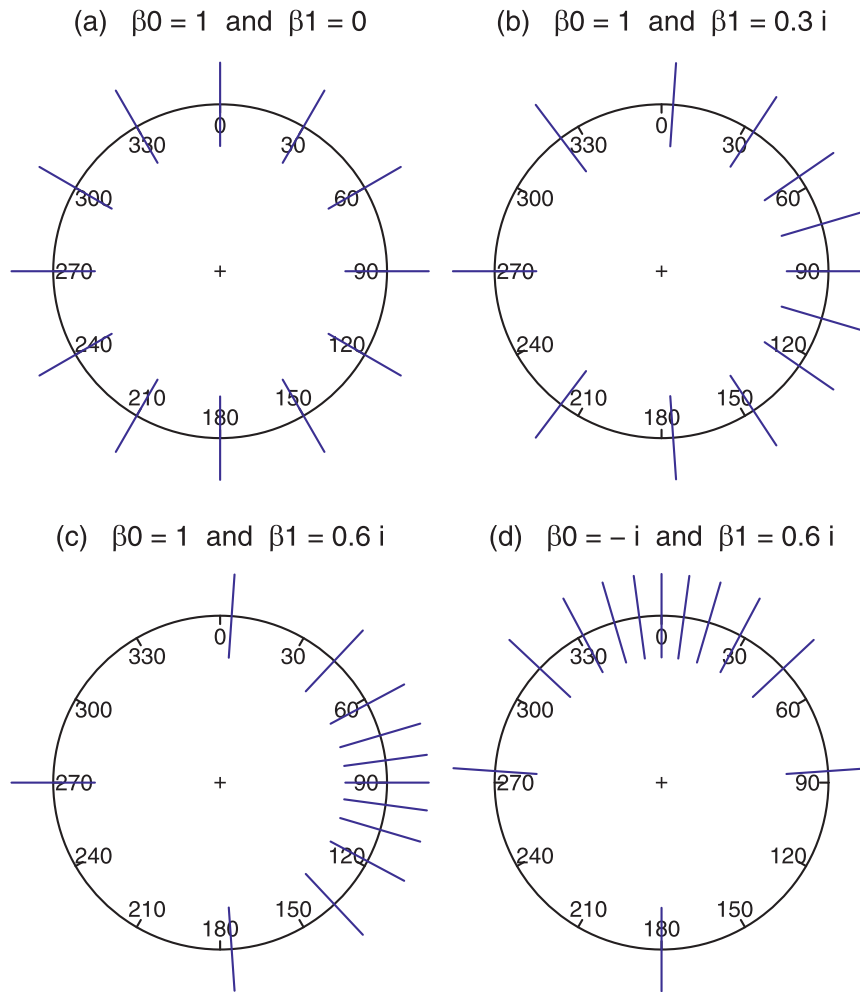


FIG. 1. Illustration of the Möbius transformation (2) in the original unit of degrees. The parameter β_0 prescribes a rotation, while β_1 can be interpreted as pulling a direction toward a fixed angle, corresponding to the point $\beta_1/|\beta_1|$ on the unit circle, with the concentration about this point increasing as $|\beta_1|$ increases, except that the antipodal point remains fixed. (a) Twelve equally spaced nontransformed angles, (b) Möbius transformation with parameters $\beta_0 = 1$ and $\beta_1 = 0.3i$, corresponding to a weak contraction toward the imaginary unit; i.e., a direction of 90 degrees. (c) Möbius transformation with parameters $\beta_0 = 1$ and $\beta_1 = 0.6i$, resulting in a stronger contraction. (d) Möbius transformation with parameters $\beta_0 = -i$ and $\beta_1 = 0.6i$, now compositing the contraction with a counterclockwise rotation by 90°.

applying Eq. (2.9) of Fisher (1993, p. 31). Lund (1999) proposes a fitting criterion of this latter type in the context of circular regression.

b. Von Mises distribution for angular data

The von Mises distribution is a natural baseline for modeling angular data, such as wind directions, and it may be viewed as a circular analog of the Gaussian distribution (Fisher 1993, p. 49). Specifically, an angular variable is said to have a von Mises distribution with mean direction μ and concentration parameter $\kappa \geq 0$ if it has density

$$g(v|\mu, \kappa) = \frac{1}{360} \frac{\exp\{\kappa \cos[(v - \mu)\pi/180]\}}{I_0(\kappa)}$$

on the circle, where I_0 is a modified Bessel function of the first kind and order zero. As the concentration parameter κ gets close to zero, the von Mises distribution becomes a uniform distribution on the circle. In the appendix, we review maximum likelihood (ML) estimation for the concentration parameter of the von Mises distribution, which can be viewed as a limiting case of Bayes estimation under weak prior information (Guttorp and Lockhart 1988).

c. Bayesian model averaging

During the past decade, the ability of ensemble systems to improve deterministic-style forecasts and to predict forecast skill has been convincingly established (Palmer 2002; Gneiting and Raftery 2005). However, forecast ensembles are typically biased and underdispersive (Hamill and Colucci 1997; Eckel and Walters 1998), and thus some form of statistical postprocessing is required. Wilks (2006b), Wilks and Hamill (2007), and Bröcker and Smith (2008) review and compare techniques for doing this.

Bayesian model averaging (BMA) was introduced by Raftery et al. (2005) as a statistical postprocessing method that generates calibrated and sharp predictive probability density functions (PDFs) from ensemble forecasts. The BMA predictive PDF of any future weather quantity of interest is a weighted average of PDFs associated with the member forecasts, where the weights reflect the members' predictive skill over a training period. The initial development was for linear weather quantities, such as surface temperature, quantitative precipitation, and wind speed (Raftery et al. 2005; Sloughter et al. 2007; Wilson et al. 2007; Sloughter et al. 2009), for which the component PDFs are probability distributions on the real line. For all variables considered and on both the synoptic scale and the mesoscale, the BMA postprocessed PDFs outperformed the unprocessed ensemble forecast and were calibrated and sharp.

Here we extend the BMA approach to accommodate wind direction, which is an angular variable and thus requires component PDFs that are probability distributions on the circle. Let f_1, \dots, f_m denote an ensemble of bias-corrected forecasts. We then take the BMA predictive PDF to be a mixture of the form

$$p(v|f_1, \dots, f_m) = \sum_{j=1}^m w_j g(v|f_j, \kappa_j),$$

where the components are von Mises distributions with mean direction f_j and concentration parameter κ_j . The BMA weights w_1, \dots, w_m are probabilities and so they are nonnegative and add up to 1; that is, $\sum_{j=1}^m w_j = 1$. Our standard BMA specification uses a common concentration parameter, so that

$$p(v|f_1, \dots, f_m) = \sum_{j=1}^m w_j g(v|f_j, \kappa). \quad (3)$$

The common concentration parameter simplifies and stabilizes estimation and, in our experience with the University of Washington Mesoscale Ensemble over the Pacific Northwest, does not deteriorate the predictive performance appreciably. However, in other types of

situations the more general formulation with distinct concentration parameters might be more adequate.

The BMA weights, w_1, \dots, w_m , and the common concentration parameter κ of the component PDFs are estimated by maximum likelihood from training data. Typically, the training set comprises a temporally and/or spatially composited collection of past, bias-corrected ensemble member forecasts, f_{1k}, \dots, f_{mk} , and the corresponding verifying direction, v_k , where $k = 1, \dots, n$, with n the number of cases in the training set. The likelihood function ℓ is then defined as the probability of the training data, viewed as a function of the w_j 's and κ ; that is,

$$\ell(w_1, \dots, w_m; \kappa) = \prod_{k=1}^n \sum_{j=1}^m w_j g(v_k|f_{jk}, \kappa),$$

where the product extends over all instances in the training set. The ML estimates are those values of the w_j 's and κ that maximize the likelihood function; that is, the values under which the verifying directions were most likely to materialize.

The likelihood function typically cannot be maximized analytically, and so it is maximized using the expectation-maximization (EM) algorithm (Dempster et al. 1977; McLachlan and Krishnan 1997). The EM algorithm is iterative and alternates between two steps, the E (or expectation) step, and the M (or maximization) step. It uses unobserved quantities z_{jk} , which can be interpreted informally as the probability of ensemble member j being the most skillful forecast for verification v_k . The z_{1k}, \dots, z_{mk} are nonnegative and sum to 1 for each instance k in the training set.

In the E step, the z_{jk} are estimated given the current values of the BMA weights and component PDFs. Specifically,

$$z_{jk}^{(l+1)} = \frac{w_j^{(l)} g(v_k|f_{jk}, \kappa^{(l)})}{\sum_{q=1}^m w_q^{(l)} g(v_k|f_{qk}, \kappa^{(l)})}, \quad (4)$$

where the superscript l refers to the l th iteration of the EM algorithm, and thus $w_q^{(l)}$ and $\kappa^{(l)}$ refer to the estimates at the l th iteration. In the M step we obtain updated estimates

$$w_j^{(l+1)} = \frac{1}{n} \sum_{k=1}^n z_{jk}^{(l+1)} \quad (5)$$

of the BMA weights. Furthermore, an updated estimate, $\kappa^{(l+1)}$, of the common concentration parameter is obtained by optimizing the expected complete-data log likelihood given the latent variables; that is, by maximizing

$$\sum_{k=1}^n \sum_{j=1}^m z_{jk}^{(l+1)} \log g(v_k | f_{jk}, \kappa) \tag{6}$$

over $\kappa \geq 0$. For implementation details, see the appendix.

An ensemble forecast can occasionally be poor, in the sense that all member forecasts turn out to be substantially different from the verifying direction. This possibility motivates a BMA specification with a uniform mixture component. Under this more general specification, to which we refer as BMA⁺, the predictive density becomes

$$p(v | f_1, \dots, f_m) = \sum_{j=1}^m w_j g(v | f_j, \kappa) + w_{m+1} u(v), \tag{7}$$

where u is the density of a uniform distribution on the circle; that is, a von Mises distribution with concentration parameter $\kappa = 0$, and where the BMA weights are non-negative and add up to 1, so that $\sum_{j=1}^{m+1} w_j = 1$. The intention here is similar to that of the addition of a climatological mixture component, as proposed and implemented for linear variables by Rajagopalan et al. (2002) and Bröcker and Smith (2008). The adaption of the EM estimation algorithm from the BMA to the BMA⁺ specification is straightforward.

d. Forecast verification

Wind direction is an angular variable, and standard scoring rules for linear variables do not apply. Instead, we use circular analogs of the absolute error and the continuous ranked probability score, as introduced by Grit et al. (2006).

From a probabilistic forecast for an angular quantity, we can create a single-valued forecast by determining the circular median of the predictive distribution, as described above and by Fisher (1993, pp. 35–36). To assess the quality of this forecast, we use the mean circular distance or circular absolute error, $AE_{\text{circ}}(f, v)$, between the single-valued forecast f and the verifying direction, v , as given by (1) in the unit of degrees.

To assess probabilistic forecasts for an angular quantity, we use the angular or circular continuous ranked probability score, which is defined by

$$\begin{aligned} CRPS_{\text{circ}}(P, v) &= \mathbb{E}_P\{AE_{\text{circ}}(V, v)\} \\ &\quad - \frac{1}{2} \mathbb{E}_P\{AE_{\text{circ}}(V, V^*)\}, \end{aligned} \tag{8}$$

where P is a forecast distribution on the circle, v is the verifying direction, V and V^* are independent copies of an angular random variable with distribution P , and $\mathbb{E}\{\cdot\}$ denotes the expectation operator. Note that when P is

a uniform distribution on the circle, then $CRPS_{\text{circ}}(P, v)$ equals 45° , independently of the verifying direction. The circular continuous ranked probability score is proper and reduces to the circular absolute error when the forecast is single-valued, just as the linear continuous ranked probability score generalizes the absolute error (Grit et al. 2006). It takes the unit of degrees and allows for the direct comparison of deterministic (single-valued) forecasts, discrete ensemble forecasts, and post-processed ensemble forecasts that can take the form of a predictive density.

The general goal in probabilistic forecasting is to maximize the sharpness of the forecast distributions subject to calibration (Gneiting et al. 2007). For probabilistic forecasts of a linear variable, it is straightforward to assess calibration via rank or probability integral transform (PIT) histograms, and sharpness via the mean width of the corresponding prediction intervals. For an angular variable, such as wind direction, there are no established ways of assessing calibration and sharpness. For example, there is no direct analog of the verification rank or PIT histogram; because there is no natural ordering on the circle if a forecast distribution is uniform or multimodal. Similarly, there is no generally accepted measure of dispersion for circular distributions (Fisher and Lee 1992, p. 666).

To address these challenges, we introduce a novel measure of sharpness for a circular distribution P , namely

$$\begin{aligned} S(P) &= \frac{1}{2} \mathbb{E}_P\{AE_{\text{circ}}(V, V^*)\} \\ &= \mathbb{E}_P\{CRPS_{\text{circ}}(P, V)\}, \end{aligned} \tag{9}$$

where V and V^* are independent copies of an angular random variable with distribution P . The quantity $S(P)$ has two complementary interpretations. By the first equality in (9), it equals one-half times the expected circular distance between two directions drawn independently and at random from the forecast distribution. Thus, $S(P)$ is a natural measure of sharpness on the circle, attaining a value of zero for a point measure, and a value of 45° for a uniform distribution. The smaller $S(P)$, the sharper the forecast distribution, and the sharper, the better, subject to calibration.

The second equality in the defining (9) is immediate from (8) and shows that $S(P)$ equals the expected value of the circular continuous ranked probability score when the forecast distribution is P , and the verifying wind direction is drawn at random from this distribution. Thus, if a forecasting method is calibrated, we expect the mean sharpness measure (9) and the mean circular continuous ranked probability score to be roughly equal.

TABLE 1. Composition of the eight-member UWME (Eckel and Mass 2005), with member acronyms, and organizational and synoptic model sources for the initial and lateral boundary conditions. The organizational sources are the United States National Centers for Environmental Prediction (NCEP), the Canadian Meteorological Centre (CMC), the Australian Bureau of Meteorology (ABM), the Japanese Meteorological Agency (JMA), the Fleet Numerical Meteorology and Oceanography Center (FNMOC), the Taiwan Central Weather Bureau (TCWB), and the Met Office (UKMO).

Member	Source	Driving synoptic model
GFS	NCEP	Global forecast system
ETA	NCEP	Limited-area mesoscale model
CMCG	CMC	Global-environment multiscale model
GASP	ABM	Global analysis and prediction model
JMA	JMA	Global spectral model
NGPS	FNMOC	Navy operational global atmospheric prediction system
TCWB	TCWB	Global forecast system
UKMO	UKMO	Unified model

Approximate equality can be checked informally or via significance tests, as proposed by Held et al. (2010).

For some predictive distributions, such as mixtures of von Mises densities, the circular continuous ranked probability score cannot be computed analytically. In such cases, we approximate it by simulating a Monte Carlo sample v_1, \dots, v_N from the predictive distribution, and computing

$$\text{CRPS}_{\text{circ}}(P, v) = \frac{1}{N} \sum_{i=1}^N \text{AE}_{\text{circ}}(v_i, v) - \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N \text{AE}_{\text{circ}}(v_i, v_j), \quad (10)$$

which agrees with (8) when the predictive distribution, P , assigns mass $1/N$ to each of v_1, \dots, v_N . Similarly, the sharpness measure (9) can be approximated by the rightmost term in (10). In order for the approximation to be accurate, the sample size N needs to be large, and we generally use $N = 1\,000$.

3. Results for the University of Washington ensemble over the Pacific Northwest

a. The University of Washington mesoscale ensemble

The University of Washington ensemble system is a mesoscale, short-range ensemble based on the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) mesoscale model (MM5; Grell et al. 1995). It forms an integral part of the Pacific Northwest regional environmental prediction effort (Mass et al. 2003). The original five-member mesoscale ensemble was designed as a single-model, multianalysis system that uses MM5 with a nested, limited-area grid configuration focusing on the states of Washington and Oregon (Grimit and Mass 2002). Beginning in the autumn of 2002, the size of the mesoscale ensemble was increased to eight members, using additional global analyses and forecasts, and named the University of Washington Mesoscale Ensemble (UWME; Eckel and Mass 2005). Table 1 shows acronyms and the sources of the initial and lateral boundary conditions for the member forecasts.

The evaluation period of this study begins 1 January 2003 and extends through 31 December 2003, in which the UWME system provided 48-h forecasts beginning at 0000 UTC each day, with the verifying wind directions being recorded 48 h later. Model 10-m wind component forecasts at the four grid-box centers surrounding each station were bilinearly interpolated to the observation location and then rotated from grid-relative to north-relative. No adjustment was made for any vertical displacement of the model surface level from the real terrain. Station-based observations of near-surface wind were acquired in real time from 54 surface airway observation (SAO) stations in the United States and Canada. Our verification results include forecast–observation cases only when the verifying wind speed was at least 5 kt (2.57 m s^{-1}), since wind direction observations are unreliable at lower wind speeds. In view of this constraint, forecast–observation cases at the individual stations were available for a minimum of 201, median of 219, and maximum of 264 days in calendar year 2003.

TABLE 2. Raw and bias-corrected UWME ensemble forecasts of wind direction ($^{\circ}$) at Castlegar Airport, BC, valid 0000 UTC 26 Aug 2003. The member-specific circular–circular regression schemes for the bias correction and the BMA and BMA⁺ parameters were fit on a 28-day training period immediately preceding the initialization date. For this site and particular training period, the concentration parameter κ was estimated at 2.984 for BMA and 4.112 for BMA⁺.

	GFS	ETA	CMCG	GASP	JMA	NGPS	TCWB	UKMO	Unified
UWME (raw)	325.0	321.3	332.4	330.1	319.4	254.3	327.7	324.7	—
UWME (bias-corrected)	323.2	315.7	320.6	326.5	310.7	246.8	323.1	318.4	—
BMA weight	0.113	0.124	0.109	0.134	0.114	0.132	0.117	0.157	—
BMA ⁺ weight	0.098	0.110	0.099	0.119	0.105	0.115	0.110	0.147	0.097

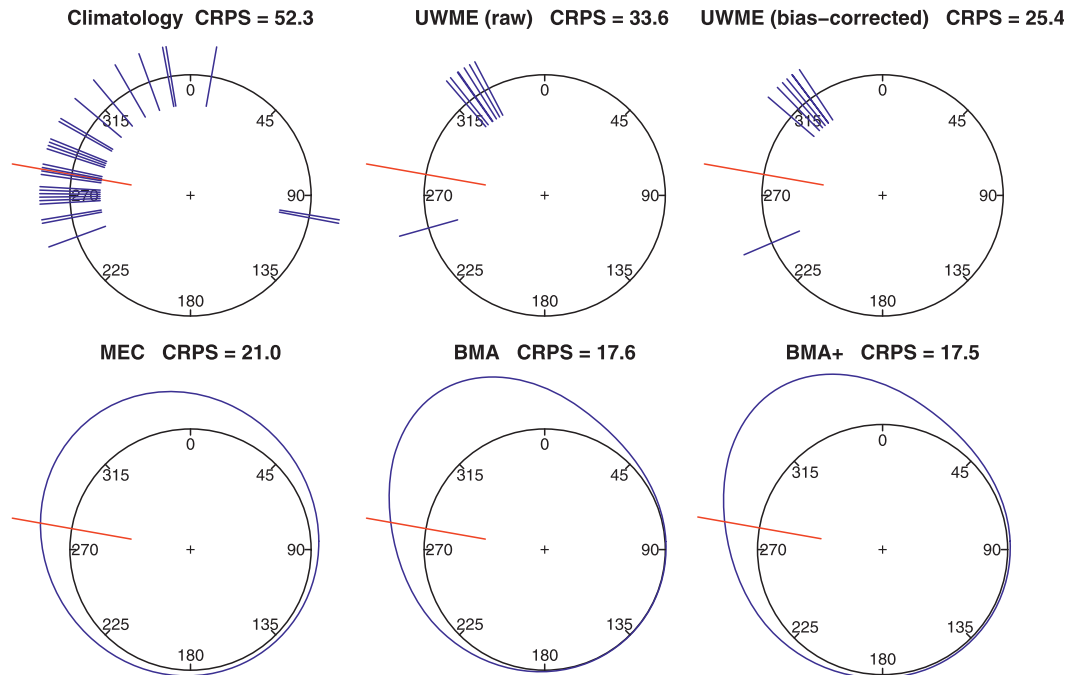


FIG. 2. Circular diagrams of forecast distributions for wind direction at Castlegar Airport, BC, valid 0000 UTC 26 Aug 2003; (top) the corresponding discrete forecast probability mass function (climatology, UWME raw, and UWME bias-corrected) or (bottom) continuous forecast probability density function (MEC, BMA, and BMA⁺). The blue lines and graphs represent the forecast distributions; the solid red line represents the verifying observation, at 280°. The circular continuous ranked probability score (CRPS) is also shown, in degrees.

Before showing composite verification results, we give a specific example of 48-h BMA and BMA⁺ forecasts of wind direction at Castlegar Airport, British Columbia (station code CYCG), valid 0000 UTC 26 August 2003. The member-specific circular–circular regression schemes for bias correction and the BMA parameters were estimated on a 28-day training period immediately preceding the initialization date, using data at Castlegar only. Table 2 shows the eight raw and bias-corrected UWME member forecasts and the respective BMA and BMA⁺ weights for this site and particular training period. The bias correction technique results in member-specific counterclockwise rotations, which range from 2° to 12°. The UKMO member received the highest BMA and BMA⁺ weights, but the weights do not differ much between the ensemble members.

Figure 2 illustrates the raw and bias-corrected UWME ensemble forecasts and the BMA and BMA⁺ density forecasts at Castlegar along with two reference forecasts, to which we refer as climatology and median error climatology (MEC), respectively. The climatology forecast uses the 28 observed wind directions during the sliding training period, giving them equal weights in a discrete probability mass function. This is a short-term climatology, which can adapt to seasonal changes as well as to changes in atmospheric regimes. The MEC technique

takes the form of a von Mises density that is centered on the circular median of the bias-corrected ensemble members, with a concentration parameter that is estimated (using ML) on the same 28-day training period as the other methods. This resembles the mean error climatology method of Grimit et al. (2006), but the density is centered on the circular ensemble median, rather than the circular mean, and the estimation method is different. Each panel shows the respective forecast distribution, taking the form of either a discrete probability mass function or a continuous probability density function, along with the verifying wind direction, which was westerly at 280 degrees. The circular continuous ranked probability score (CRPS) is smallest (i.e., best) for the BMA⁺ forecast distribution, at 17.5°, followed by the BMA, MEC, bias-corrected UWME, raw UWME, and climatology forecasts.

b. Bias correction

We turn to composite verification results for bias correction. In section 2a we proposed three methods for bias correcting angular variables, namely circular–circular regression, which employs a state of the art regression approach tailored to circular data, and two benchmarks, median-angle correction and mean-angle correction. As

TABLE 3. Mean circular absolute error for raw and bias-corrected 48-h forecasts of wind direction ($^{\circ}$) over the Pacific Northwest. The results are averaged over the eight UWME member forecasts, the calendar year 2003, and the 54 stations we consider, for sliding training periods of length 7, 14, 21, 28, 35, and 42 days, using local data at the given station only.

Training period	7 days	14 days	21 days	28 days	35 days	42 days
UWME (raw)	45.14	45.14	45.14	45.14	45.14	45.14
Mean-angle correction	46.62	44.60	43.76	43.36	43.18	43.21
Median-angle correction	47.10	44.58	43.63	43.18	43.02	43.08
Circular-circular regression	49.42	45.37	43.60	42.88	42.69	42.80

noted before, we fit the bias correction schemes for each ensemble member individually.

There are two choices to be made here, namely about the method used and the length of the sliding training period. Table 3 shows the mean circular absolute error for each method, averaged over the eight UWME member forecasts, calendar year 2003, and the 54 stations we consider, for sliding training periods that range from 7 to 42 days. In choosing the length of the training period, there is a trade-off, and no automatic way of making it. Both weather patterns and model specifications change over time, so that there is an advantage in using a short training period to adapt to such changes. On the other hand, the longer the training period is, the less the estimation variance. The training sets are constrained to cases at the location at hand, and the periods are extended if there are missing data. For example, the 7-day training period always uses the seven most recent available forecast cases.

At a 7-day training period the simpler methods outperform the more complex method, namely circular-circular regression. However, as the training period grows, circular-circular regression becomes the method of choice. This is not surprising, and can readily be explained by the bias-variance trade-off, in that more complex statistical methods require larger training sets, to avoid overfitting. Overall, circular-circular regression with training periods of 28 days or more performs the best. On average, it reduces the circular absolute error by 2° or 3° , as compared to the raw forecast. In the subsequent ensemble postprocessing experiments, we thus use circular-circular regression to bias-correct the UWME member forecasts, where the regression parameters in (2) are fit on a member- and location-specific 28-day sliding training period.

c. Bayesian model averaging

With an effective bias-correction technique now at hand, we proceed to discuss ensemble postprocessing techniques for wind direction. All results below are based on the same 28-day sliding training period that we use for bias-correction via circular-circular regression, and are insensitive to changes in the length of the training

period. We compare the various methods using the mean circular continuous ranked probability score. Furthermore, we reduce the forecast distributions to the corresponding circular medians, and compute the mean circular absolute error for these single-valued forecasts.

Specifically, Table 4 shows the verification statistics for the discrete UWME (raw) and UWME (bias-corrected) forecast distributions, with the bias correction using circular-circular regression, the BMA and BMA⁺ forecasts (based on the bias-corrected UWME), and the two reference forecasts introduced and described in section 2a, namely climatology and MEC, where the latter is also based on the bias-corrected UWME. The results are averaged over calendar year 2003 and the 54 stations we consider. Bias-correction via circular-circular regression yields a reduction of the circular absolute error for the ensemble median forecast of 3° to 4° on average. As expected, ensemble calibration does not result in any further reduction of the circular absolute error, because MEC, BMA, and BMA⁺ address calibration errors only. However, the latter methods result in a much decreased mean circular continuous ranked probability score, with BMA⁺ performing the best, while BMA is a close competitor. Indeed, the BMA and BMA⁺ forecast distributions are sharper than those for MEC or climatology, and they are calibrated, because the corresponding mean sharpness measure (9) equals roughly the mean circular continuous ranked probability score.

TABLE 4. Mean circular absolute error, mean circular continuous ranked probability score, and the mean sharpness measure (9) for 48-h forecasts of wind direction ($^{\circ}$) over the Pacific Northwest, in degrees. The results are averaged over the calendar year 2003 and the 54 stations we consider. A 28-day sliding training period is applied, using local data at the given station only.

	AE _{circ}	CRPS _{circ}	Sharpness
Climatology	56.9	35.9	32.0
UWME (raw)	42.8	35.0	11.7
UWME (bias-corrected)	39.3	31.2	10.5
MEC	39.3	28.8	27.5
BMA	39.4	27.8	26.5
BMA ⁺	39.3	27.6	26.6

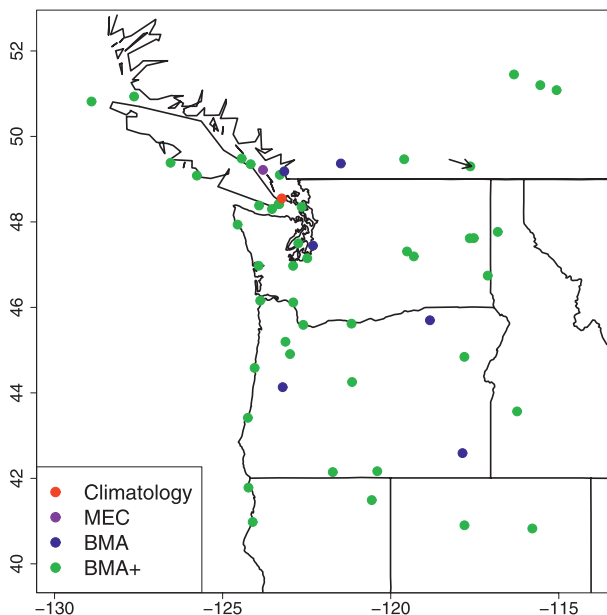


FIG. 3. Pacific Northwest domain for the UWME system, with the locations of the SAO stations considered in this study. The color at each station location indicates which of the forecast methods in Table 4 performed best in terms of mean CRPS_{circ} over calendar year 2003: green stands for BMA⁺, blue for BMA, purple for MEC, and red for climatology. The station at Castlegar Airport, BC, is marked by an arrow.

If the forecasts are stratified by verifying wind speeds (results not shown), we see differing error characteristics, in that the mean circular absolute error and the mean circular continuous ranked probability score are lower at the stronger winds when the wind direction observations tend to be more stable. However, the relative ranking of the forecast methods remains unchanged from that in Table 4.

When averaged over calendar year 2003 and the 54 stations, the mean BMA and BMA⁺ weights for the UWME members show little difference. For BMA, the smallest mean weight is 0.122 and the highest is 0.129; for BMA⁺ the corresponding range is from 0.108 to 0.116, with the uniform component contributing a weight of 0.108 on average. The individual weight estimates are quite stable, with a median absolute change in a weight from one training period to the next of 0.002.

Turning now to results at individual stations, Fig. 3 shows the Pacific Northwest domain for the UWME system, along with the locations of the 54 SAO stations considered in this study. The color at each station location indicates what forecast method had the lowest mean circular continuous ranked probability score in calendar year 2003. At 46 of these stations the BMA⁺ method performed best, and at six stations the BMA forecasts

showed the lowest score. MEC and climatology performed best at one station each.

4. Discussion

We have shown how to perform bias correction and ensemble calibration for wind direction, which is an angular variable. For bias correction, we use the state of the art circular–circular regression approach of Downs and Mardia (2002) and Kato et al. (2008), which employs a Möbius transformation to regress the verifying wind direction on the model wind direction. A possible extension is via the regression model of Lund (1999), which allows for additional linear predictor variables, such as the model wind speed. To estimate the regression parameters, we use a minimum circular absolute error criterion. Bayesian estimation methods, such as those proposed by George and Ghosh (2006) and Bhattacharya and Sengupta (2009), offer attractive alternatives.

For ensemble calibration, our preferred choice is the BMA⁺ technique, which applies Bayesian model averaging, where the von Mises components are centered on the bias-corrected ensemble member forecasts. When compared to the standard BMA approach, the BMA⁺ specification uses an additional uniform component, which can protect against gross forecast errors. A potential extension might replace the uniform component by a seasonally adaptive climatological component, which could be estimated from multiyear records of wind observations. In the semiparametric Bayes framework of Bhattacharya and Sengupta (2009), data-driven decisions regarding the equality of the concentration parameters for the von Mises components could be implemented in Dirichlet process settings.

Our methods have been developed for the UWME system (Grimit and Mass 2002; Eckel and Mass 2005), which has eight individually distinguishable members. They can easily be adapted to accommodate situations in which the ensemble member forecasts are exchangeable (i.e., statistically indistinguishable), as in most bred, singular vector or ensemble Kalman filter systems (Buizza et al. 2005; Torn and Hakim 2008). In these cases, the circular–circular regression approach to bias correction continues to apply, but the regression (2) uses a single set of parameters across ensemble members. Similarly, the BMA or BMA⁺ weights for the von Mises components in (3) and (7) need to be constrained to be equal. These modifications result in physically principled bias correction and BMA specifications, while simplifying the postprocessing. Similar adaptations allow for bias correction and ensemble calibration in multimodel systems with groups of exchangeable and/or missing members, in ways analogous to those described by Fraley et al. (2010)

for linear variables. For example, The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) system comprises ten groups, with 11 to 51 members each, which typically are exchangeable (Park et al. 2008; Bougeault et al. 2010), and thus will share common bias correction parameters as well as common BMA or BMA⁺ weights.

Our work should not be viewed as an endorsement of vector wind calibration techniques in which wind speed and wind direction are treated independently. Rather, vector wind postprocessing ought to proceed jointly on the zonal and meridional wind components. In this light, new work is underway, in which we develop bias correction and Bayesian model averaging techniques for vector wind. If the focus is on wind direction by itself, it remains to be determined whether or not vector wind postprocessing with a subsequent reduction to the directional part, is preferable to direct postprocessing of the wind direction forecasts, as proposed and studied in this paper, or separate postprocessing of the zonal and meridional components and a subsequent reduction to wind direction. Future work is called for in which these approaches are to be compared and an authoritative recommendation for an operational implementation be made.

Acknowledgments. We are indebted to Cliff Mass and Jeff Baars for sharing their insights and providing data, and to three anonymous reviewers for a wealth of constructive and helpful feedback. This research was sponsored by the National Science Foundation under Joint Ensemble Forecasting System (JEFS) Subaward S06-47225 with the University Corporation for Atmospheric Research (UCAR), as well as Grants ATM-0724721 and DMS-0706745.

APPENDIX

Details for the EM Algorithm

Before discussing the details of our EM implementation, we review maximum likelihood (ML) estimation for the concentration parameter of a von Mises population. Suppose that we have a sample, v_1, \dots, v_n , from von Mises distributions with known mean directions, μ_1, \dots, μ_n , and unknown common concentration parameter, κ . The corresponding log-likelihood function is

$$-n \log I_0(\kappa) + \kappa \sum_{k=1}^n \cos(v_k - \mu_k), \quad (\text{A1})$$

up to an additive constant. Therefore, the ML estimate for κ equals the unique root, $\hat{\kappa}$, of the equation

$$\frac{I'_0(\kappa)}{I_0(\kappa)} = \bar{C} \quad \text{where} \quad \bar{C} = \frac{1}{n} \sum_{k=1}^n \cos(v_k - \mu_k). \quad (\text{A2})$$

This equation can be solved numerically or using the tables in Mardia (1972). Furthermore, there are accurate analytic approximations, in that

$$\hat{\kappa} \doteq \frac{n+2}{2n(1-\bar{C})} \quad (\text{A3})$$

if \bar{C} is small, and

$$\frac{1}{\hat{\kappa}} \doteq 2(1-\bar{C}) + \frac{1}{\bar{C}}(1-\bar{C})^2 \times (0.48794 - 0.82905\bar{C} - 1.3915\bar{C}^2) \quad (\text{A4})$$

if \bar{C} is large, by Eqs. (2.8) and (2.10) of Lenth (1981), where the symbol \doteq denotes an approximate equality.

Turning now to the EM algorithm for estimating the BMA model (3) of section 2c, the M step requires an updated estimate $\kappa^{(l+1)}$ of the common concentration parameter. The update is obtained by optimizing the expected complete-data log likelihood given the latent variables; that is, by maximizing

$$\sum_{k=1}^n \sum_{j=1}^m z_{jk}^{(l+1)} \log g(v_k | f_{jk}, \kappa) \quad (\text{A5})$$

over $\kappa \geq 0$, where $g(\cdot | f_{jk}, \kappa)$ is a von Mises density with mean f_{jk} and concentration parameter κ . It is easily seen that (A5) takes essentially the same form as the log-likelihood function (A1), and thus we can apply the above methods and approximations. Specifically, putting now

$$\bar{C} = \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^m z_{jk}^{(l+1)} \cos(v_k - f_{jk}),$$

we find $\kappa^{(l+1)} = \hat{\kappa}$ from the approximation (A3) if $\bar{C} \leq 0.1$, and from the approximation (A4) if $\bar{C} \geq 0.9$. Otherwise we solve (A2) numerically.

REFERENCES

- Atger, F., 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. *Mon. Wea. Rev.*, **131**, 1509–1523.
- Bhattacharya, S., and A. Sengupta, 2009: Bayesian analysis of semiparametric linear-circular models. *J. Agric. Biol. Environ. Stat.*, **14**, 33–65.
- Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Meteor. Soc.*, in press.
- Bröcker, J., and L. A. Smith, 2008: From ensemble forecasts to predictive distribution functions. *Tellus*, **60A**, 663–678.

- Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Carter, G. M., 1975: Automated prediction of surface wind from numerical model output. *Mon. Wea. Rev.*, **103**, 866–873.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood for incomplete data via the EM algorithm. *J. Roy. Stat. Soc. Ser. B*, **39**, 1–38.
- Downs, T. D., and K. V. Mardia, 2002: Circular regression. *Biometrika*, **89**, 683–697.
- Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting*, **13**, 1132–1147.
- , and C. F. Mass, 2005: Aspects of effective short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.
- Engel, C., and E. Ebert, 2007: Performance of hourly operational consensus forecasts (OCFs) in the Australian region. *Wea. Forecasting*, **22**, 1345–1359.
- Fisher, N. I., 1993: *Statistical Analysis of Circular Data*. Cambridge University Press, 277 pp.
- , and A. J. Lee, 1992: Regression models for an angular response. *Biometrics*, **48**, 665–677.
- Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, **138**, 190–202.
- George, B. J., and K. Ghosh, 2006: A semiparametric Bayesian model for circular-linear regression. *Comm. Stat. Simul. Comput.*, **35**, 911–923.
- 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.
- , and D. A. Unger, 1986: A local AFOS MOS program (LAMP) and its application to wind prediction. *Mon. Wea. Rev.*, **114**, 1313–1329.
- Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249.
- , F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. Ser. B*, **69**, 243–268.
- Grell, G. A., J. Dudhia, and D. R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). National Center for Atmospheric Research Tech. Note NCAR/TN-398+STR, 121 pp.
- Grimit, E. P., and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting*, **17**, 192–205.
- , T. Gneiting, V. J. Berrocal, and N. A. Johnson, 2006: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quart. J. Roy. Meteor. Soc.*, **132**, 2925–2942.
- Guttorp, P., and R. A. Lockhart, 1988: Finding the location of a signal: A Bayesian analysis. *J. Amer. Stat. Assoc.*, **83**, 322–330.
- Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327.
- Held, L., K. Rufibach, and F. Balabdaoui, 2010: A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics*, in press.
- Kato, S., K. Shimizu, and G. S. Shieh, 2008: A circular-circular regression model. *Stat. Sin.*, **18**, 633–645.
- Lenth, R. V., 1981: On finding the source of a signal. *Technometrics*, **23**, 149–154.
- Lund, U., 1999: Least circular distance regression for directional data. *J. Appl. Stat.*, **26**, 723–733.
- Mardia, K. V., 1972: *Statistics of Directional Data*. Academic Press, 357 pp.
- Mass, C. F., 2003: IFPS and the future of the National Weather Service. *Wea. Forecasting*, **18**, 75–79.
- , and Coauthors, 2003: Regional environmental prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **84**, 1353–1366.
- McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, 274 pp.
- Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–774.
- Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.
- Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130**, 1792–1811.
- Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220.
- , T. Gneiting, and A. E. Raftery, 2009: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, in press.
- Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Mon. Wea. Rev.*, **136**, 3947–3963.
- Wilks, D. S., 2006a: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Elsevier Academic Press, 627 pp.
- , 2006b: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteor. Appl.*, **13**, 243–256.
- , and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 1364–1385.
- Wilson, L. J., S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 1364–1385.