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

Le Bao[1], Tilmann Gneiting[1], Eric P. Grimit[2], Peter Guttorp[1] and Adrian E. Raftery[1]

Department of Statistics, University of Washington[1]
3TIER Inc., Seattle, Washington[2]

University of Washington, Department of Statistics
Technical Report no. 557

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 post-processing methods become ineffective, or do not apply at all. We propose 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. We apply these techniques to 48-hour 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 post-processing 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-hour forecasts of surface wind direction over the Pacific Northwest in 2003 using the University of Washington Mesoscale Ensemble (Grimit and Mass 2002; Eckel and 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 degrees denoting a northerly, easterly, southerly and westerly wind. The angular distance or circular absolute error,

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

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

$$\theta(v) = e^{i\pi(90-v)/180}$$

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

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, one might develop separate MOS equations for zonal and meridional components and derive single-valued wind direction forecasts from them

(Carter 1975; Glahn and Unger 1986), but this does not take the dependencies between the wind components into account. Thus, we take a different approach 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

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

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 Moebius 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 towards a fixed angle, namely the point $\beta_1/|\beta_1|$ on the unit circle, with the concentration about $\beta_1/|\beta_1|$ increasing as $|\beta_1|$ increases (Kato et al. 2008). In our **circular-circular regression** approach to bias correction for wind direction, we estimate the regression model (2) from training data, by 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 β_0 and β_1.

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 regression equation (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 $(f_1, v_1), \ldots, (f_n, v_n)$ of predicted and observed directions, the median of the directional errors is the angle m that minimizes

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

which is equivalent to the definition in eq. (2.32) of Fisher (1993, p. 36). The respective 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 applying eq. (2.9) of Fisher (1993, p. 31).

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 analogue 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

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

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 post-processing 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 post-processing 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 post-processed 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 $f_1, \ldots, f_m$ denote an ensemble of bias-corrected forecasts. We then take the BMA predictive PDF to be a mixture of the form

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

where the components are von Mises distributions with mean direction f_i and concentration parameter κ_i. The BMA weights $w_1, \ldots, w_m$ 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

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

The common concentration parameter simplifies and stabilizes estimation and, in our experience with the University of Washington Mesoscale Ensemble, does not deteriorate predictive performance.

The BMA weights, $w_1, \ldots, w_m$, and the concentration parameter, κ, of the component PDFs are estimated by maximum likelihood (ML) from training data. Typically, the training set comprises a temporally and/or spatially composited collection of past, bias-corrected ensemble member forecasts, $f_{1k}, \ldots, f_{mk}$, and the corresponding verifying direction, y_k, where $k = 1, \ldots, n$, 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_i's and κ, that is,

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

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 as the probability of ensemble member j being the most skillful forecast for verification v_k. The $z_{1k}, \ldots, z_{mk}$ 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,

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

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

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

of the BMA weights. Furthermore, an updated estimate, $\kappa^{(l+1)}$, of the common concentration parameter is obtained by optimizing the complete data 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

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

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 nonnegative 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 analogues of the absolute error and the continuous ranked probability score, as described by Grimit 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, $\mathrm{AE}_{\mathrm{circ}}(f, v)$, between the single-valued forecast, f, and the verifying direction, v, as given by eq. (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

$$\mathrm{CRPS}_{\mathrm{circ}}(P, v) = \mathbb{E}\big\{ \mathrm{AE}_{\mathrm{circ}}(V, v) \big\} - \frac{1}{2} \, \mathbb{E}\big\{ \mathrm{AE}_{\mathrm{circ}}(V, V^*) \big\}, \tag{8}$$

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 $\mathrm{CRPS}_{\mathrm{circ}}(P, v)$ equals 45 degrees, independently of the verifying direction. The circular continuous ranked probability score is proper and reduces to the circular distance when the forecast is single-valued, just as the linear continuous ranked probability score generalizes the absolute error (Grimit 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.

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 $v_1, \ldots, v_N$ from the predictive distribution, and computing

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

which agrees with (8) when the predictive distribution assigns mass $1/N$ to each of $v_1, \ldots, v_N$. In order for the approximation to be accurate, the sample size N needs to be large, and we generally use $N = 1000$.

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

a. The University of Washington Mesoscale Ensemble (UWME)

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 Mesoscale Model (PSU-NCAR 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, multi-analysis 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-hour forecasts beginning at 0000 UTC each day, with the verifying wind directions being recorded 48 hours later. Model 10 m wind component forecasts at the four grid-box centers surrounding each station were bi-linearly 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

Table 1: Composition of the eight-member University of Washington Mesoscale Ensemble (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 United Kingdom Met Office (UKMO).

Member	Source	Driving Synoptic Model
GFS	NCEP	Global Forecast System
ETA	NCEP	Limited-Area Mesoscale Model
CMCG	CMC	Global-Environment Multi-Scale 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

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, as illustrated in Figure 2 below. Our verification results include forecast-observation cases only when the verifying wind speed was at least 5 knots (2.57 m/s), 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.

Before showing composite verification results, we give a specific example of 48-hour BMA and BMA$^+$ forecasts of wind direction at Castlegar Airport, British Columbia (station code CYCG), valid 0000 UTC on 26 August 2003. The member-specific circular-circular regression schemes for bias correction and the BMA parameters were estimated on a 28-day sliding training period, using data at Castlegar only. Table 2 shows the eight raw and bias-corrected UWME member forecasts and the respective BMA and BMA$^+$ weights. The bias correction technique results in member-specific counter-clockwise rotations, which range from two to twelve degrees. The UKMO member receives the highest BMA and BMA$^+$ weights, but the weights do not differ much between the ensemble members.

Figure 1 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 median error climatology (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 (that is, best) for the BMA$^+$ forecast distribution, at 17.5 degrees, 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 2.a 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 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, 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 two to three degrees, 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.

Table 2: Raw and bias-corrected UWME ensemble forecasts of wind direction at Castlegar Airport, British Columbia, valid 0000 UTC on 26 August 2003. The member-specific circular-circular regression schemes for the bias correction and the BMA and BMA$^+$ parameters were fit on a 28-day sliding 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	Unif
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

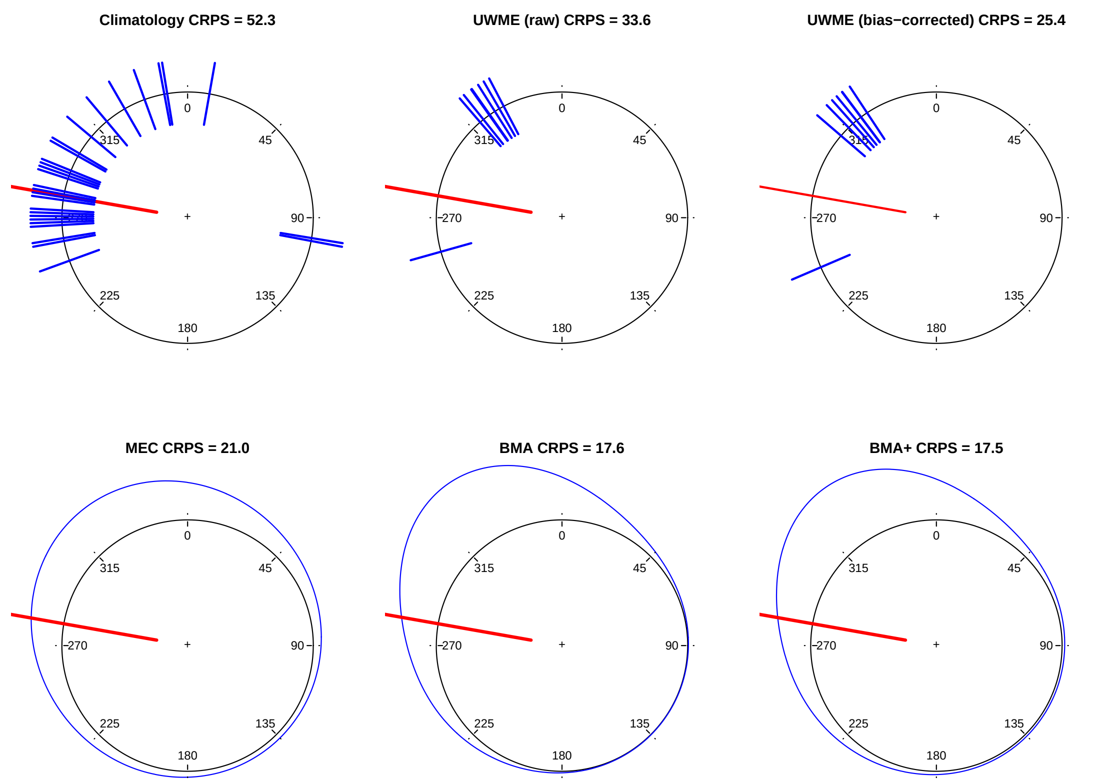

Figure 1: Circular diagrams of forecast distributions for wind direction at Castlegar Airport, British Columbia, valid 0000 UTC on 26 August 2003. Each panel in the figure shows the respective discrete forecast probability mass function (upper row: climatology, UWME raw, and UWME bias-corrected) or continuous forecast probability density function (lower row: MEC, BMA, and BMA$^+$). The blue lines and graphs represent the forecast distributions; the solid red line represents the verifying observation, at 280 degrees. The circular continuous ranked probability score is also shown, in units of degrees.

Table 3: Mean circular absolute error for raw and bias-corrected 48-hour forecasts of wind direction 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

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 2.a, namely **climatology** and **median-error climatology (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 slightly over three degrees 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.

Turning now to results at individual stations, Figure 2 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 6 stations the BMA forecasts showed the lowest score. MEC and climatology performed best at one station each.

Table 4: Mean circular absolute error and mean circular continuous ranked probability score for 48-hour forecasts of wind direction over the Pacific Northwest, in units of 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.

	$\mathrm{AE}_{\mathrm{circ}}$	$\mathrm{CRPS}_{\mathrm{circ}}$
Climatology	56.9	35.9
UWME (raw)	42.8	35.0
UWME (bias-corrected)	39.3	31.2
MEC	39.3	28.8
BMA	39.4	27.8
BMA^+	39.3	27.6

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). 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 multi-year records of wind observations.

These 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 (that is, 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 equation (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 multi-model systems with groups of exchangeable and/or missing members, in ways analogous to those described by Fraley et al. (2009) for linear variables. For example, the 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. 2009), and thus will share common bias correction parameters as well as equal BMA or BMA^+ weights.

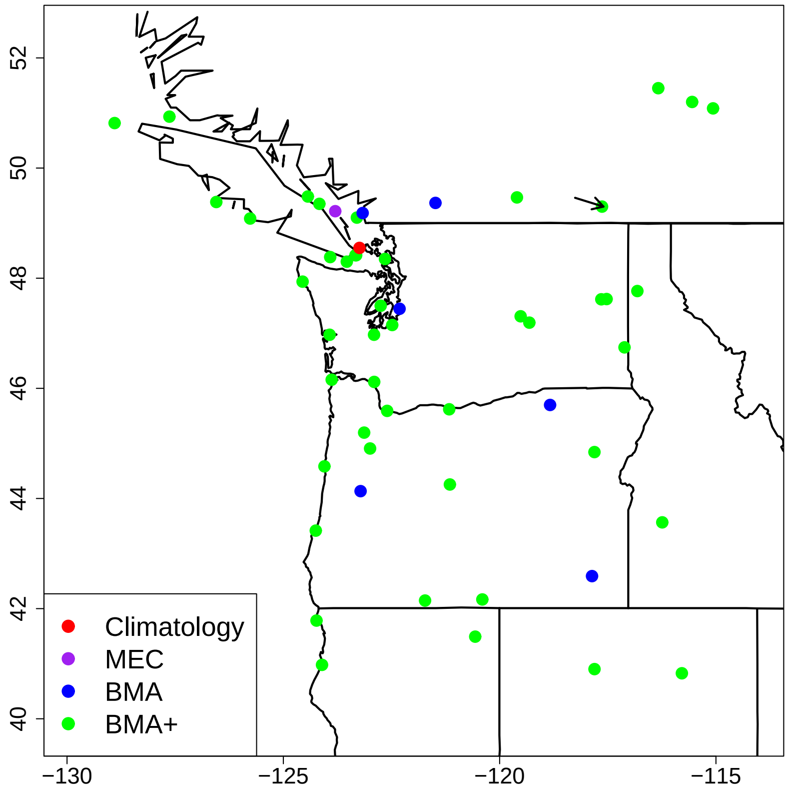

Figure 2: Pacific Northwest domain for the UWME system, with the locations of the SAO stations considered in this study. The color at each station location indicates which of the forecast methods in Table 4 performed best in terms of mean $\text{CRPS}_{\text{circ}}$ over calendar year 2003: green stands for BMA^+, blue for BMA, purple for MEC, and red for climatology. The station at Castlegar Airport, British Columbia is marked by an arrow.

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.

Acknowledgements

We are indebted to Cliff Mass and Jeff Baars for sharing their insights and providing data. 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 awards 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, $v_1, \ldots, v_n$, from von Mises distributions with known mean directions, $\mu_1, \ldots, \mu_k$, and unknown common concentration parameter, κ. The respective log-likelihood function is

$$-n \log I_0(\kappa) + \kappa \sum_{k=1}^{n} \cos(v_k - \mu_k), \tag{9}$$

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

$$\frac{I_0'(\kappa)}{I_0(\kappa)} = \bar{C} \qquad \text{where} \qquad \bar{C} = \frac{1}{n} \sum_{k=1}^{n} \cos(v_k - \mu_k). \tag{10}$$

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

$$\hat{\kappa} \doteq \frac{n+2}{2n(1-\bar{C})} \tag{11}$$

if $\bar{C}$ is small, and

$$\frac{1}{\hat{\kappa}} \doteq 2(1-\bar{C}) + \frac{1}{\bar{C}}\,(1-\bar{C})^2\,(.48794 - .82905\bar{C} - 1.3915\bar{C}^2) \tag{12}$$

if $\bar{C}$ is large, by equations (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 2.c, the M step requires an updated estimate, $\kappa^{(l+1)}$, of the common concentration parameter. The update is obtained by optimizing the complete data log-likelihood given the latent variables, that is, by maximizing

$$\sum_{k=1}^{n} \sum_{j=1}^{m} z_{jk}^{(l+1)} \log g(v_k | f_{jk}, \kappa) \tag{13}$$

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 (13) takes essentially the same form as the log-likelihood function (9), 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 (11) if $\bar{C} \leq 0.1$, and from the approximation (12) if $\bar{C} \geq 0.9$. Otherwise we solve eq. (10) numerically.

References

Atger, F., 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. *Monthly Weather Review*, **131,** 1509–1523.

Bougeault, P., and coauthors, 2009: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bulletin of the American Meteorological Society*, in press.

Bröcker, J., and L. A. Smith, 2008: From ensemble forecasts to predictive distribution functions. *Tellus Series A*, **60,** 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. *Monthly Weather Review*, **133,** 1076–1097.

Carter, G. M., 1975: Automated prediction of surface wind from numerical model output. *Monthly Weather Review*, **103,** 866–873.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series 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. *Weather and Forecasting*, **13,** 1132–1147.

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

Engel, C., and E. Ebert, 2007: Performance of hourly operational consensus forecasts (OCFs) in the Australian region. *Weather and Forecasting*, **22,** 1345–1359.

Fisher, N. I., 1993: *Statistical Analysis of Circular Data*. Cambridge University Press, 277 pp.

Fraley, C., A. E. Raftery, and T. Gneiting, 2009: Calibrating multi-model forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review*, in press.

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

Glahn, H. R., and D. A. Unger, 1986: A local AFOS MOS program (LAMP) and its application to wind prediction. *Monthly Weather Review*, **114,** 1313–1329.

Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310,** 248–249.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102,** 359–378.

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 Technical Note NCAR/TN-398+STR, available at `http://www.mmm.ucar.edu/mm5/documents/mm5-desc-doc.html`.

Grimit, E. P., and C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting*, **17,** 192–205.

Grimit, E. P., 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. *Quarterly Journal of the Royal Meteorological Society*, **132,** 2925–2942.

Guttorp, P., and R. A. Lockhart, 1988: Finding the location of a signal: A Bayesian analysis. *Journal of the American Statistical Association*, **83,** 322–330.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, **125,** 1312–1327.

Kato, S., K. Shimizu, and G. S. Shieh, 2008: A circular-circular regression model. *Statistica Sinica*, **18,** 633–645.

Lenth, R. V., 1981: On finding the source of a signal. *Technometrics*, **23,** 149–154.

Mardia, K. V., 1972: *Statistics of Directional Data*. New York: Academic Press.

Mass, C. F., 2003: IFPS and the future of the National Weather Service. *Weather and Forecasting*, **18,** 75–79.

Mass, C. F., and coauthors, 2003: Regional environmental prediction over the Pacific Northwest. *Bulletin of the American Meteorological Society*, **84,** 1353–1366.

McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, New York.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, **128,** 747–

774.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, **134,** 2029–2050.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, **133,** 1155–1174.

Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Monthly Weather Review*, **130,** 1792–1811.

Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, **135,** 3209–3220.

Sloughter, J. M., T. Gneiting, and A. E. Raftery, 2009: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, accepted for publication.

Torn, R. D., and G. J. Hakim, 2008: Performance characteristics of a pseudo-operational ensemble Kalman filter. *Monthly Weather Review*, **136,** 3947–3963.

Wilks, D. S., 2006a: *Statistical Methods in the Atmospheric Sciences*, 2nd edition. Elsevier Academic Press, 627 pp.

Wilks, D. S., 2006b: Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorological Applications*, **13,** 243–256.

Wilks, D. S., and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Review*, **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. *Monthly Weather Review*, **135,** 1364–1385.