Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging

J. McLean SLOUGHTER, Tilmann GNEITING, and Adrian E. RAFTERY

The current weather forecasting paradigm is deterministic, based on numerical models. Multiple estimates of the current state of the atmosphere are used to generate an ensemble of deterministic predictions. Ensemble forecasts, while providing information on forecast uncertainty, are often uncalibrated. Bayesian model averaging (BMA) is a statistical ensemble postprocessing method that creates calibrated predictive probability density functions (PDFs). Probabilistic wind forecasting offers two challenges: a skewed distribution, and observations that are coarsely discretized. We extend BMA to wind speed, taking account of these challenges. This method provides calibrated and sharp probabilistic forecasts. Comparisons are made between several formulations.

KEY WORDS: ECME algorithm; Gamma distribution; Numerical weather prediction; Skewed distribution; Truncated data; Wind energy.

1. INTRODUCTION

While deterministic point forecasts have long been the standard in weather forecasting, there are many situations in which probabilistic information can be of value. In this paper, we consider the case of wind speed. Often, ranges or thresholds can be of interest—recreational sailors are likely to be more interested in the probability of there being enough wind to go out sailing than in simply the best guess at the wind speed, and farmers may be interested in the chance of winds being low enough to safely spray pesticides. Possible extreme values are of particular interest, where it can be important to know the chance of winds high enough to pose dangers for boats or aircraft.

In situations calling for a cost/loss analysis, the probabilities of different outcomes need to be known. For wind speed, this issue often arises in the context of wind power, where underforecasting and overforecasting carry different financial penalties. The optimal point forecast in these situations is often a quantile of the predictive distribution (Roulston et al. 2003; Pinson, Chevallier, and Kariniotakis 2007; Gneiting 2008). Different situations can require different quantiles, and this needed flexibility can be provided by forecasts of a full predictive probability density function (PDF). Environmental concerns and climate change have made wind power look like an appealing source of clean and renewable energy, and as this field continues to grow, calibrated and sharp probabilistic forecasts can help to make wind power a more financially competitive alternative.

Purely statistical methods have been applied to short-range forecasts for wind speed only a few hours into the future (Brown, Katz, and Murphy 1984; Kretzschmar et al. 2004; Gneiting et al. 2006; Genton and Hering 2007). A detailed survey of the literature on short-range wind forecasting can be found in Giebel, Brownsword, and Kariniotakis (2003).

Medium-range forecasts looking several days ahead are generally based on numerical weather prediction models, which can then be statistically postprocessed. To estimate the predictive distribution of a weather quantity, an ensemble forecast is often used. An ensemble forecast consists of a set of multiple forecasts of the same quantity, based on different estimates of the initial atmospheric conditions and/or different physical models (Palmer 2002; Gneiting and Raftery 2005). An example of an ensemble forecast for wind speed can be seen in Figure 1. Ensemble forecasts can give an indication of uncertainty, and a statistical relationship between forecast errors and ensemble spread has been established for several ensemble systems. However, it has also been shown that ensemble forecasts typically are uncalibrated, with a tendency for observed values to fall outside of the range of the ensemble too often (Grimit and Mass 2002; Buizza et al. 2005; Gneiting and Raftery 2005).

In this light, a number of methods have been proposed for statistically postprocessing ensemble forecasts of wind speed or wind power. These approaches have largely focused on the use of quantile regression to generate forecast bounds and/or intervals (Bremnes 2004; Nielsen et al. 2004; Nielsen, Madsen, and Nielsen 2006; Pinson et al. 2007; Møller, Nielsen, and Madsen 2008). However, in order to obtain a full PDF from these methods, it requires running a large number of regression models to generate many quantiles, and then interpolating between them (and correcting for possible problems with quantiles crossing over one another), rather than explicitly modeling a full predictive PDF.

Bayesian model averaging (BMA) was introduced by Raftery et al. (2005) as a statistical postprocessing method for producing probabilistic forecasts from ensembles in the form of predictive PDFs. The BMA predictive PDF of any future weather quantity of interest is a weighted average of PDFs centered on the individual bias-corrected forecasts, where the weights can be interpreted as posterior probabilities of the models generat-

J. McLean Sloughter is Assistant Professor of Mathematics, Mathematics Department, Seattle University, Seattle, WA 98122 (E-mail: *sloughtj@seattleu.edu*). Tilmann Gneiting is Professor of Mathematical Statistics, Institut für Angewandte Mathematik, Universität Heidelberg, 69120 Heidelberg, Germany (E-mail: *t.gneiting@uni-heidelberg.de*). Adrian E. Raftery is Blumstein–Jordan Professor of Statistics and Sociology, Department of Statistics, University of Washington, Seattle, WA 98195-4322 (E-mail: *raftery@u.washington.edu*). We are grateful to Jeff Baars, Chris Fraley, Eric Grimit, Clifford F. Mass, Pierre Pinson, the referees, and the editor for helpful discussions and for providing code and data. This research was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under grant N00014-01-10745, by the National Science Foundation under awards ATM-0724721 and DMS-0706745, and by the Joint Ensemble Forecasting System (JEFS) under subcontract S06-47225 from the University Corporation for Atmospheric Research (UCAR).

Journal of the American Statistical Association
March 2010, Vol. 105, No. 489, Applications and Case Studies
DOI: 10.1198/jasa.2009.ap08615

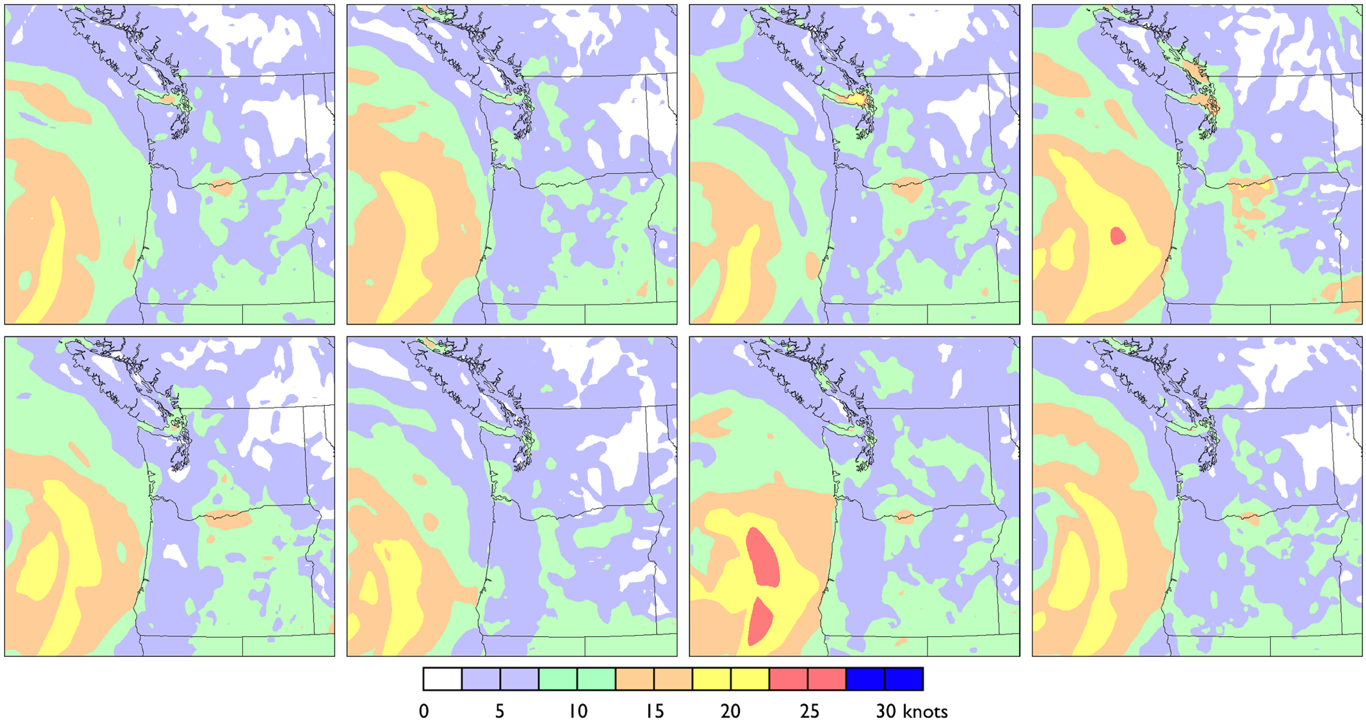

Figure 1. 48-hour-ahead ensemble forecast of maximum wind speed over the Pacific Northwest on August 7, 2003 using the eight-member University of Washington mesoscale ensemble (Grimit and Mass 2002; Eckel and Mass 2005).

ing the forecasts and reflect the forecasts' contributions to overall forecasting skill over a training period. The original development of BMA by Raftery et al. (2005) was for weather quantities whose predictive PDFs are approximately normal, such as temperature and sea-level pressure. This method was modified by Sloughter et al. (2007) to apply to quantitative precipitation forecasts. These forecasts used component distributions that had a positive probability of being equal to zero, and, when not zero, were skewed, and modeled using power-transformed gamma distributions.

As with precipitation, wind speed has a skewed distribution. Unlike for precipitation, there is no need to model the separate probability of wind speed being equal to zero, at least in the geographic region we consider. Here we develop a BMA method for wind speed by modeling the component distribution for a given ensemble member as a gamma distribution; the BMA PDF is then itself a mixture of such distributions.

In Section 2 we review the BMA technique and describe our extension of it to wind speed. Statistical approaches to wind forecasting offer a unique challenge in that observed values are reported discretized to the nearest whole knot, a much coarser discretization than is seen in other weather quantities. We compare a number of methods for estimating the parameters of the BMA PDF which account for the discretization in different ways. Then in Section 3 we give results for 48-hour-ahead forecasts of maximum wind speed over the North American Pacific Northwest in 2003 based on the eight-member University of Washington mesoscale ensemble (Grimit and Mass 2002; Eckel and Mass 2005). Throughout the paper we use illustrative examples drawn from these data, and we find that BMA is calibrated and sharp for the period we consider. Finally, in Section 4 we discuss alternative approaches and possible improvements to the method.

2. DATA AND METHODS

2.1 Forecast and Observation Data

This research considers 48-hour-ahead forecasts of maximum wind speed over the Pacific Northwest in the period from November 1, 2002 through December 31, 2003, using the eight-member University of Washington mesoscale ensemble (Eckel and Mass 2005) initialized at 00 hours UTC in international standard time, which is 5 p.m. local time in summer, when daylight saving time operates, and 4 p.m. local time otherwise. The dataset contains observations and forecasts at surface airway observation (SAO) stations, a network of automated weather stations located at airports throughout the United States and Canada. Maximum wind speed is defined as the maximum of the hourly "instantaneous" wind speeds over the previous 18 hours, where an hourly "instantaneous" wind speed is a 2-minute average from the period of two minutes before the hour to on the hour. Data were available for 340 days, and data for 86 days during this period were unavailable. In all, 35,230 station observations were used, an average of about 104 per day. The forecasts were produced for observation locations by bilinear interpolation from forecasts generated on a 12-kilometer grid, as is common practice in the meteorological community. The wind speed observations were subject to the quality control procedures described by Baars (2005).

The wind speed data we analyze are discretized when recorded—wind speed is rounded to the nearest whole knot. Additionally, any wind speeds below one knot are recorded as zero. One knot is equal to approximately 0.514 meters per second, or 1.151 miles per hour.

2.2 Bayesian Model Averaging

BMA (Leamer 1978; Kass and Raftery 1995; Hoeting et al. 1999) was originally developed as a way to combine inferences and predictions from multiple statistical models, and was applied to statistical linear regression and related models in the social and health sciences. Raftery et al. (2005) extended BMA to ensembles of deterministic prediction models and showed how it can be used as a statistical postprocessing method for forecast ensembles, yielding calibrated and sharp predictive PDFs of future weather quantities.

In BMA for forecast ensembles, each ensemble member forecast f_k is associated with a component PDF, $g_k(y|f_k)$. The BMA predictive PDF for the future weather quantity, y, is then a mixture of the component PDFs, namely

$$p(y|f_1,\ldots,f_K)=\sum_{k=1}^{K} w_k g_k(y|f_k), \tag{1}$$

where the BMA weight w_k is based on forecast k's relative performance in the training period. The w_k's are probabilities and so they are nonnegative and add up to 1, that is, $\sum_{k=1}^{K} w_k = 1$. Here K is the number of ensemble members.

The component PDF $g_k(y|f_k)$ can be thought of roughly as the conditional PDF of the weather quantity y given the kth forecast, f_k, conditional on f_k being the best forecast in the ensemble. This heuristic interpretation is in line with how operational weather forecasters often work, by selecting one or a small number of "best" forecasts from a potentially large number available, based on recent predictive performance (Joslyn and Jones 2008).

2.3 Gamma Model

For weather variables such as temperature and sea level pressure, the component PDFs can be fit reasonably well using a normal distribution centered at a bias-corrected forecast, as shown by Raftery et al. (2005). For precipitation, Sloughter et al. (2007) modeled the component PDFs using a mixture of a point mass at zero and a power-transformed gamma distribution.

Haslett and Raftery (1989) modeled the square root of wind speed using a normal distribution. Wind speed distributions have also often been modeled by Weibull densities (Justus, Hargraves, and Yalcin 1976; Hennessey 1977; Justus et al. 1978; Stevens and Smulders 1979). Tuller and Brett (1984) noted that the necessary assumptions for fitting a Weibull distribution are not always met. Here we generalize the Weibull approach by considering gamma distribution fits to power transformations of the observed wind speeds. We found that gamma distributions for the raw observed wind speeds themselves gave a good fit, and, perhaps surprisingly, fit better than using any power transformation. In determining the power transformation to use for the model, we fit gamma distributions to sets of the (possibly transformed) observed wind speeds, conditional on forecast values being within some bin (e.g., all observations for which the forecast was less than 5 knots). This is an approximation to the goal of our model, where we want to fit a distribution to observed wind speeds conditional on forecast values. Figure 2 shows examples of quantile–quantile plots for the untransformed observed wind speeds, demonstrating that the conditional gamma distributions provided a good fit.

In light of this, we model the component PDFs of wind speed as untransformed gamma distributions. The gamma distribution with shape parameter α and scale parameter β has the PDF

$$g(y)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}y^{\alpha-1}\exp(-y/\beta) \tag{2}$$

for $y \geq 0$, and $g(y) = 0$ for $y < 0$. The mean of this distribution is $\mu = \alpha\beta$, and its variance is $\sigma^2 = \alpha\beta^2$.

It remains to specify how the parameters of the gamma distribution depend on the numerical forecast. An exploratory data analysis showed that the observed wind speed is approximately linear as a function of the forecasted wind speed, with a standard deviation that is also approximately linear as a function of the forecast. This is illustrated in Figure 3, which shows the relation between the forecasted values (here represented by the midpoint of the forecast bin) and the means and standard deviations for gamma distribution fits to the observed wind speeds, conditional on the forecast being within the bin. These plots are based on the first ensemble member, though similar results were found for all ensemble members. Here the maximum bin considered was for forecasts between 30 and 35 knots, as there were very few instances of forecasts above 35 knots.

Putting these observations together, we get the following model for the component gamma PDF of wind speed:

$$g_k(y|f_k)=\frac{1}{\beta_k^{\alpha_k}\Gamma(\alpha_k)}y^{\alpha_k-1}\exp(-y/\beta_k). \tag{3}$$

The parameters of the gamma distribution depend on the ensemble member forecast, f_k, through the relationships

$$\mu_k = b_{0k} + b_{1k} f_k \tag{4}$$

and

$$\sigma_k = c_{0k} + c_{1k} f_k, \tag{5}$$

where $\mu_k = \alpha_k\beta_k$ is the mean of the distribution, and $\sigma_k = \sqrt{\alpha_k}\beta_k$ is its standard deviation. Here we restrict the standard deviation parameters to be constant across all ensemble members. This simplifies the model by reducing the number of parameters to be estimated, makes parameter estimation computationally easier, and reduces the risk of overfitting. We found that it led to no degradation in predictive performance. The c_{0k} and c_{1k} terms are replaced by c_0 and c_1.

Our BMA model for the predictive PDF of the weather quantity, y, here the maximum wind speed, is thus (1) with g_k as defined in (3).

2.4 Parameter Estimation

Parameter estimation is based on forecast-observation pairs from a training period, which we take here to be the N most recent available days preceding initialization. The training period is a sliding window, and the parameters are reestimated for

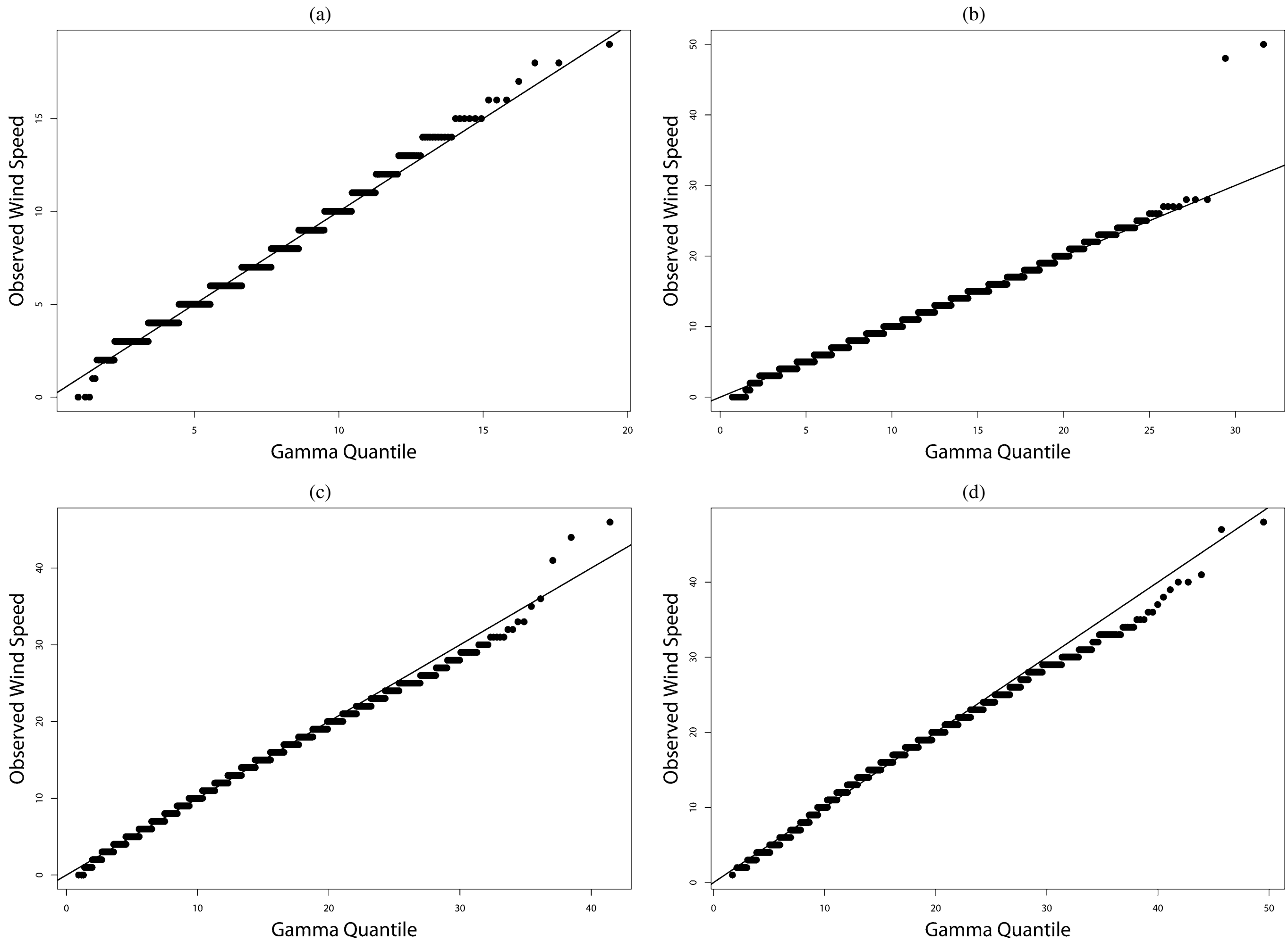

Figure 2. Gamma quantile–quantile plots for observed wind speeds conditional on the first ensemble member forecast being (a) less than 5 knots, (b) between 5 and 10 knots, (c) between 10 and 15 knots, and (d) between 15 and 20 knots. Data used is the full dataset as described above, from November 2002 through December 2003.

each new initialization period. We considered training periods ranging from the past 20 to 45 days. An examination of the sensitivity of our results to training period length showed very similar performance across potential training period lengths. Differences in average errors were only seen three decimal places out. Within this range, we consistently saw that a 25-day train-

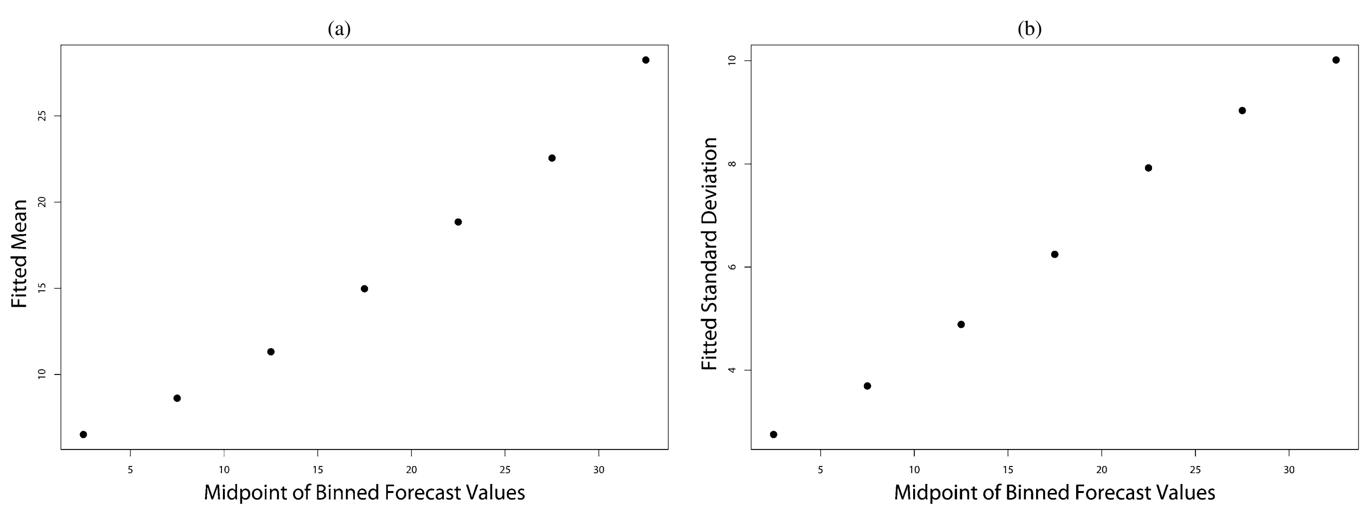

Figure 3. Fitted means and standard deviations for the forecast bins. Again using the full November 2002 through December 2003 data.

ing period gave very slightly better performance, and we will present results here based on this period.

2.4.1 Standard Method. We first consider a standard method of parameter estimation similar to the method used for quantitative precipitation in Sloughter et al. (2007). We estimate the mean parameters, b_{0k} and b_{1k}, by linear regression. These parameters are member specific, and are thus estimated separately for each ensemble member, using the observed wind speed as the dependent variable and the forecasted wind speed, f_k, as the independent variable.

We estimate the remaining parameters, $w_1, \ldots, w_K$, c_0, and c_1, by maximum likelihood from the training data. Assuming independence of forecast errors in space and time, the log-likelihood function for the BMA model is

$$\ell(w_1, \ldots, w_K; c_0; c_1) = \sum_{s,t} \log p(y_{st}|f_{1st}, \ldots, f_{Kst}), \quad (6)$$

where the sum extends over all station locations, s, and times, t, in the training data. As noted above, wind speed observations below one knot are recorded as zero knots. The log-likelihood requires calculating the logarithm of each observed wind speed, which is not possible with values of zero.

Wilks (1990) suggested a method for maximum likelihood estimation of gamma distribution parameters with data containing zeroes due to rounding by, for each value of zero, replacing the corresponding component of the log-likelihood with the aggregated probability of the range of values that would be rounded to zero (in our case, between 0 and 1 knots). To incorporate this, for each observed y_{st} recorded as a zero, we replace $p(y_{st}|f_{1st}, \ldots, f_{Kst})$ above by

$$p(y_{st}|f_{1st}, \ldots, f_{Kst}) = P(1|f_{1st}, \ldots, f_{Kst}), \quad (7)$$

where

$$P(a|f_{1st}, \ldots, f_{Kst}) = \int_0^a p(y|f_{1st}, \ldots, f_{Kst})\, dy. \quad (8)$$

The log-likelihood function cannot be maximized analytically, and instead we maximize it numerically using the ECME algorithm (Liu and Rubin 1994), a variant of the EM algorithm (Dempster, Laird, and Rubin 1977; McLachlan and Krishnan 1997). Like the EM algorithm, the ECME algorithm is iterative, alternating between expectation (E) steps, and a series of conditional maximization (CM) steps, each maximizing either the expected complete data log-likelihood or the mixture log-likelihood conditional on the current estimates of the weights. It uses unobserved quantities z_{kst}, which are latent variables equal to 1 if observation y_{st} comes from the kth mixture component, and to 0 otherwise.

In the E step, the z_{kst} are estimated given the current estimate of the parameters. Specifically,

$$\hat{z}_{kst}^{(j+1)} = \frac{w_k^{(j)} p^{(j)}(y_{st}|f_{kst})}{\sum_{l=1}^K w_l^{(j)} p^{(j)}(y_{st}|f_{lst})}, \quad (9)$$

where the superscript j refers to the jth iteration of the ECME algorithm, and thus $w_k^{(j)}$ refers to the estimate of w_k at the jth iteration. The quantity $p^{(j)}(y_{st}|f_{kst})$ is defined using the estimates of c_0 and c_1 from the jth iteration, and is either $g_k(y_{st}|f_{kst})$ as defined in (3), if y_{st} is nonzero, or is $G_k(1|f_{kst})$, if y_{st} is zero, where

$$G_k(a|f_{kst}) = \int_0^a g_k(y|f_{kst})\, dy. \quad (10)$$

Note that, although the z_{kst} are equal to either 0 or 1, the $\hat{z}_{kst}$ are real numbers between 0 and 1. The $\hat{z}_{1st}, \ldots, \hat{z}_{Kst}$ are nonnegative and sum to 1 for each (s, t).

The CM-1 step then consists of maximizing the expected complete data log-likelihood as a function of $w_1, \ldots, w_K$, where the expectation is taken over the distribution of z_{kst} given the data and the previous parameter estimates. This is the same as maximizing the log-likelihood given the z_{kst}, c_0, and c_1, evaluated at $z_{kst} = \hat{z}_{kst}^{(j+1)}$. Thus

$$w_k^{(j+1)} = \frac{1}{n} \sum_{s,t} \hat{z}_{kst}^{(j+1)}, \quad (11)$$

where n is the number of cases in the training set, that is, the number of distinct values of (s, t).

The CM-2 step maximizes the mixture log-likelihood as a function of c_0, and c_1, conditional upon the estimated $w_1, \ldots, w_K$. There are no analytic solutions for the CM-2 step estimates, and so they must be found numerically.

The E and CM steps are then iterated to convergence, which we define as a change no greater than some small tolerance in the mixture log-likelihood in one iteration. The log-likelihood is guaranteed to increase at each ECME iteration (Liu and Rubin 1994), which implies that in general it converges to a local maximum of the likelihood. Convergence to a global maximum cannot be guaranteed, so the solution reached by the algorithm can be sensitive to the starting values. Choosing the starting values based on equal weights and the marginal variance usually led to a good solution in our experience.

Due to the numerical estimates in the CM-2 step, this algorithm can be computationally intensive. We found that computing time could be significantly reduced through a modification to the algorithm, in which we computed a CM-2 step only once for every 50 E and CM-1 steps. For our data, this modified algorithm gives nearly identical parameter estimates to the original algorithm.

2.4.2 Fully Discretized Method. While our standard method addresses the computational issues associated with recorded observations of zero, it does not more generally address the discretization of the wind speed values. We therefore consider a fully discretized method of parameter estimation that generalizes the method of Wilks (1990). Each component of the log-likelihood is replaced by the aggregated probability of the range of values that would be rounded to the recorded value. In our case, a recorded observation of 0 indicates a true value between 0 and 1, a recorded observation of 1 indicates a true value between 1 and $\frac{3}{2}$, and for any integer $i > 1$, a recorded observation of i indicates a true value between $i - \frac{1}{2}$ and $i + \frac{1}{2}$.

To extend the approach of our initial method, we first leave $p(y_{st}|f_{1st}, \ldots, f_{Kst})$ as defined in (7) for observed values of 0. For observed values of 1, we have

$$p(y_{st}|f_{1st}, \ldots, f_{Kst}) = P\left(\tfrac{3}{2}|f_{1st}, \ldots, f_{Kst}\right) - P(1|f_{1st}, \ldots, f_{Kst}), \quad (12)$$

and for observed values i where $i > 1$,

$$p(y_{st}|f_{1st},\ldots,f_{Kst}) = P\big(i+\tfrac{1}{2}|f_{1st},\ldots,f_{Kst}\big) - P\big(i-\tfrac{1}{2}|f_{1st},\ldots,f_{Kst}\big). \quad (13)$$

Analogously, in the E step of the ECME algorithm, for observed values of 0, we put

$$p^{(j)}(y_{st}|f_{kst}) = G_k(1|f_{kst}) \quad (14)$$

for observed values of 1,

$$p^{(j)}(y_{st}|f_{kst}) = G_k\big(\tfrac{3}{2}|f_{kst}\big) - G_k(1|f_{kst}) \quad (15)$$

and for observed values i where $i > 1$,

$$p^{(j)}(y_{st}|f_{kst}) = G_k\big(i+\tfrac{1}{2}|f_{kst}\big) - G_k\big(i-\tfrac{1}{2}|f_{kst}\big). \quad (16)$$

The rest of the ECME algorithm remains unchanged.

2.4.3 Doubly Discretized Method. In the fully discretized method, the discretization of observations is taken into account in the log-likelihood. This allows us to account for the discretization when estimating the BMA weights and the standard deviation parameters, which are estimated via maximum likelihood. However, this does not address the mean parameters, which are fit via linear regression. To take account of the discretization in estimating the mean parameters, we additionally discretize the forecasts in the same manner that the observations have been. The parameters are then estimated as in the fully discretized method, replacing the ensemble member forecasts with the discretized forecasts.

2.4.4 Pure Maximum Likelihood Method. We next investigate the possibility of estimating the mean parameters by maximum likelihood as well. To avoid computational problems with a parameter space of too high a dimension, we restrict the mean parameters to be constant across ensemble members, similar to the constraint already placed on the standard deviation parameters. We then estimate the BMA weights, mean parameters, and standard deviation parameters simultaneously via maximum likelihood, using the discretized log-likelihood function from the fully discretized method. The log-likelihood is optimized numerically.

2.4.5 Parsimonious Method. We finally consider one additional method, taking the partially discretized log-likelihood from the standard method but adding the constraint that the mean parameters must be constant across ensemble members. This represents the most parsimonious model, in that it has the smallest number of parameters.

3. RESULTS

We begin by looking at aggregate results over the entire Pacific Northwest domain, for the full 2003 calendar year, with the data available from late 2002 used only as training data, to allow us to create forecasts starting in January. The following section will then look at some more specific examples of results for individual locations and/or times.

3.1 Results for the Pacific Northwest

In assessing probabilistic forecasts of wind speed, we aim to maximize the sharpness of the predictive PDFs subject to calibration (Gneiting, Balabdaoui, and Raftery 2007). Calibration refers to the statistical consistency between the forecast PDFs and the observations. To assess calibration, we consider Figure 4, which shows the verification rank histogram for the raw ensemble forecast and probability integral transform (PIT) histograms for the BMA forecast distributions. In both cases, a more uniform histogram indicates better calibration. The verification rank histogram plots the rank of each observed wind speed relative to the combined set of the observation and the eight ensemble member forecasts. That is, a rank of 1 indicates that the observed value was the lowest number in the set and thus below all eight ensemble members, a rank of 2 indicates that the observed value was greater than only one of the ensemble members, and so on up to a rank of 9 indicating that the observed value was greater than all eight ensemble members. We then record the frequency with which each possible rank occurs. If the observation and the ensemble members come from the same distribution, then the observed and forecasted values are exchangeable so that all possible ranks are equally likely. The PIT is the value that the predictive cumulative distribution function attains at the observation and is a continuous analog of the verification rank.

For our data, the verification rank histogram illustrates the lack of calibration in the raw ensemble, which is underdispersed, similar to results reported by Eckel and Walters (1998), Hamill and Colucci (1998), and Mullen and Buizza (2001) for other ensembles. From the PIT histograms for the BMA forecast distributions, all five methods of parameter estimation gave similar results, in all cases substantially better calibrated than the raw ensemble.

If the eight-member raw ensemble were properly calibrated, there would be a $\frac{1}{9}$ probability of the wind speed observation falling below the ensemble range, and a $\frac{1}{9}$ probability of it falling above the ensemble range. As such, to allow direct comparisons to the raw ensemble, we will consider $\frac{7}{9}$ or 77.8% central prediction intervals from the BMA PDF. Table 1 shows the empirical coverage of 77.8% prediction intervals, and the results echo what we see in the verification rank and PIT histograms. The raw ensemble was highly uncalibrated. The BMA intervals were well calibrated. The table also shows the average width of the prediction intervals, which characterizes the sharpness of the forecast distributions. While the raw ensemble provides a narrower interval, this comes at the cost of much poorer calibration.

Scoring rules provide summary measures of predictive performance that address calibration and sharpness simultaneously. A particularly attractive scoring rule for probabilistic forecasts of a scalar variable is the continuous ranked probability score (CRPS), which generalizes the mean absolute error (MAE), and can be directly compared to the latter. It is a proper scoring rule and is defined as

$$\begin{aligned}\mathrm{crps}(P,x) &= \int_{-\infty}^{\infty} (P(y) - I\{y \geq x\})^2\,\mathrm{d}y \\ &= \mathrm{E}_P|X-x| - \frac{1}{2}\mathrm{E}_P|X-X'|, \end{aligned} \quad (17)$$

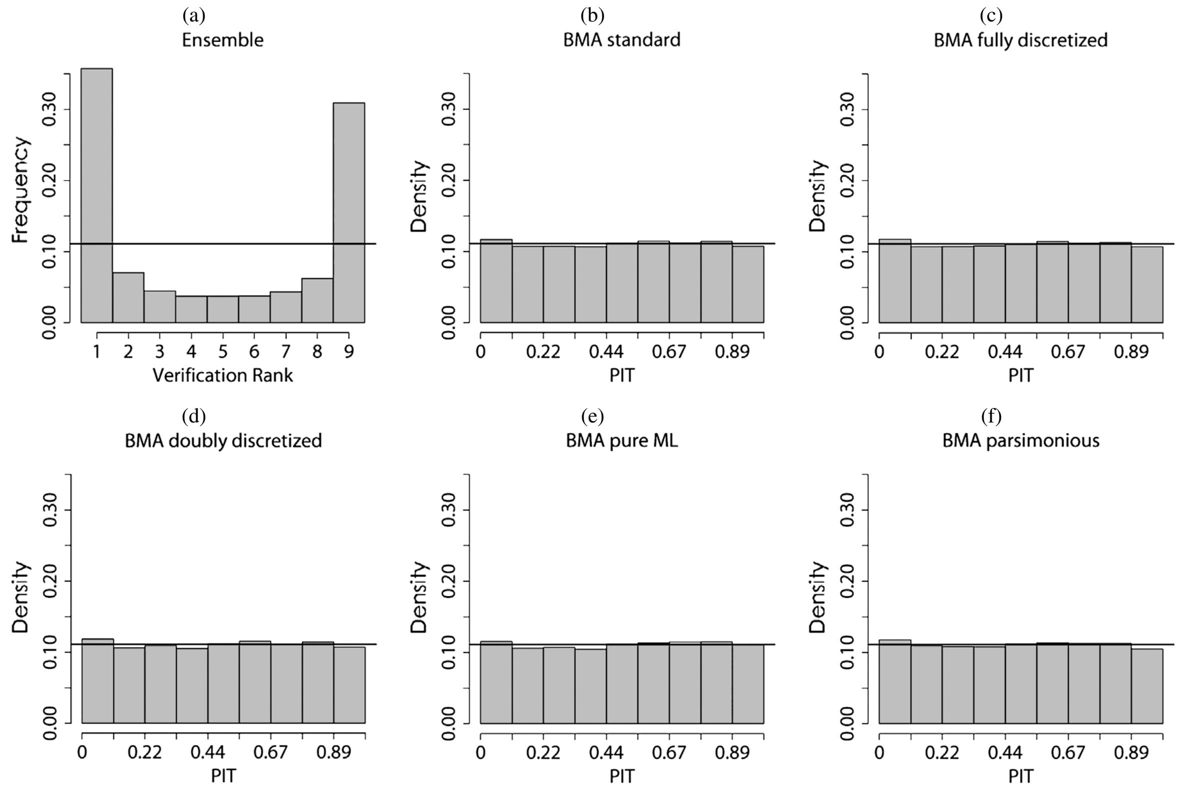

Figure 4. Calibration checks for probabilistic forecasts of wind speed over the Pacific Northwest in 2003. (a) Verification rank histogram for the raw ensemble, and PIT histograms for the BMA forecast distributions estimated using (b) the standard method, (c) the fully discretized method, (d) the doubly discretized method, (e) the pure maximum likelihood method, and (f) the parsimonious method.

where P is the predictive distribution, here taking the form of a cumulative distribution function, x is the observed wind speed, and X and X' are independent random variables with distribution P (Matheson and Winkler 1976; Grimit et al. 2006; Wilks 2006; Gneiting and Raftery 2007). Both CRPS and MAE are negatively oriented, that is, the smaller the better.

Table 1 shows CRPS and MAE values for climatology, that is, the marginal distribution of observed wind speed across space and time for the dataset, the raw ensemble forecast, and the BMA forecasts, all in units of knots. A point forecast can be created from the BMA forecast distribution by finding the median of the predictive distribution, and the MAE refers to this forecast. As the median does not have any closed form solution for our model, it is approximated numerically, in our case by the bisection method. Similarly, we show the MAE for the median of the eight-member forecast ensemble, with the results for BMA being by far the best. The results for the CRPS were similar, in that BMA outperformed the raw ensemble and climatology.

Table 1. Mean continuous ranked probability score (CRPS) and mean absolute error (MAE), and coverage and average width of 77.8% central prediction intervals for probabilistic forecasts of wind speed over the Pacific Northwest in 2003. Coverage in percent, all other values in knots. The MAE refers to the point forecast given by the median of the respective forecast distribution

Forecast	CRPS	MAE	Coverage	Width
Climatology	2.972	4.143	77.8	13.00
Ensemble	2.918	3.544	33.3	3.98
BMA standard	2.397	3.382	77.5	10.32
BMA fully discretized	2.397	3.382	77.3	10.27
BMA doubly discretized	2.397	3.383	77.3	10.28
BMA pure MLE	2.392	3.376	77.3	10.20
BMA parsimonious	2.398	3.384	77.7	10.36

These results show that all the parameter estimation methods considered performed similarly. While the pure maximum likelihood method gave very slightly better results, it was much slower computationally. We therefore recommend the use of the parsimonious method, which is the simplest in terms both of number of parameters to be estimated and of computational needs.

In particular, for the parsimonious method the parameters b_{0k} and b_{1k} in the regression equation (4) do not depend on the ensemble member forecast, k. Over the course of the entire 2003 year, the average estimate for the common intercept b_0 was 2.94. The average estimate for the slope b_1 was 0.72. Turning to the variance parameters in (5), the average estimate for c_0 was 1.41, while the average estimate for c_1 was 0.25.

3.2 Examples

To illustrate the BMA forecast distributions for wind speed, we show an example, on August 7, 2003 at Shelton, Washington. Figure 5 shows the ensemble values, the BMA component distributions, the BMA PDF, the BMA central 77.8% forecast interval, and the observation. The observed wind speed of 12 knots fell just above the ensemble range, while it was within the range of the BMA interval.

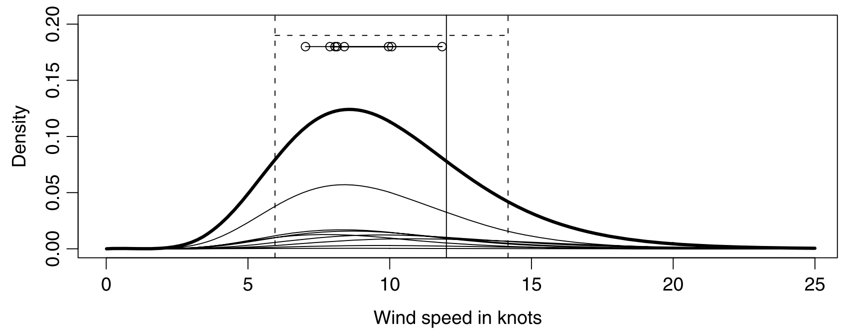

Figure 5. 48-hour-ahead BMA predictive PDF for maximum wind speed at Shelton, Washington on August 7, 2003. The upper solid curve is the BMA PDF. The lower curves are the components of the BMA PDF, namely, the weighted contributions from the ensemble members. The dashed vertical lines represent the 11th and 89th percentiles of the BMA PDF; the dashed horizontal line is the respective prediction interval; the circles represent the ensemble member forecasts; and the solid vertical line represents the verifying observation.

Figure 6 shows maps of the BMA median and BMA 90th percentile upper bound forecast for August 7, 2003. If we compare these to the ensemble forecast in Figure 1 we see that the general spatial structure is largely preserved in the BMA forecast. Figure 7 shows the verifying wind observations over the Pacific Northwest on August 7, 2003. It is evident that the raw ensemble was underforecasting in a number of areas where the BMA forecast was not.

Finally, we compare the BMA forecasts at Shelton, Washington over the 2003 calendar year to the raw ensemble forecast and the station climatology, that is, the marginal distribution of all observed values at Shelton over the time period considered. Table 2 shows CRPS and MAE scores along with prediction interval coverage and average width for station climatology, the raw ensemble, and BMA at this location. The BMA forecast showed substantially better CRPS and MAE than both station climatology and the raw ensemble. Furthermore, BMA gave sharper intervals on average than station climatology, and a better calibrated interval than the raw ensemble.

It should be noted that wind forecasting is notoriously difficult. While in this example BMA outperformed station climatology, that is not to be taken for granted, and may not be the case at stations at which there is substantial topography at subgrid scales (Gneiting et al. 2008).

4. DISCUSSION

We have shown how to apply BMA to wind speed forecasts. This provides a statistical postprocessing method for ensembles of numerical weather predictions that yields a full predictive distribution for maximum wind speed. In our experiments with the University of Washington mesoscale ensemble, the BMA forecast PDFs were better calibrated than the raw ensemble, which was underdispersed on average. The BMA median forecast had lower MAE than the ensemble median, and the BMA forecast PDFs had substantially lower CRPS than the raw ensemble or climatology.

Nielsen et al. (2004) presented a probabilistic forecasting method based on correcting quantiles of an ensemble, while

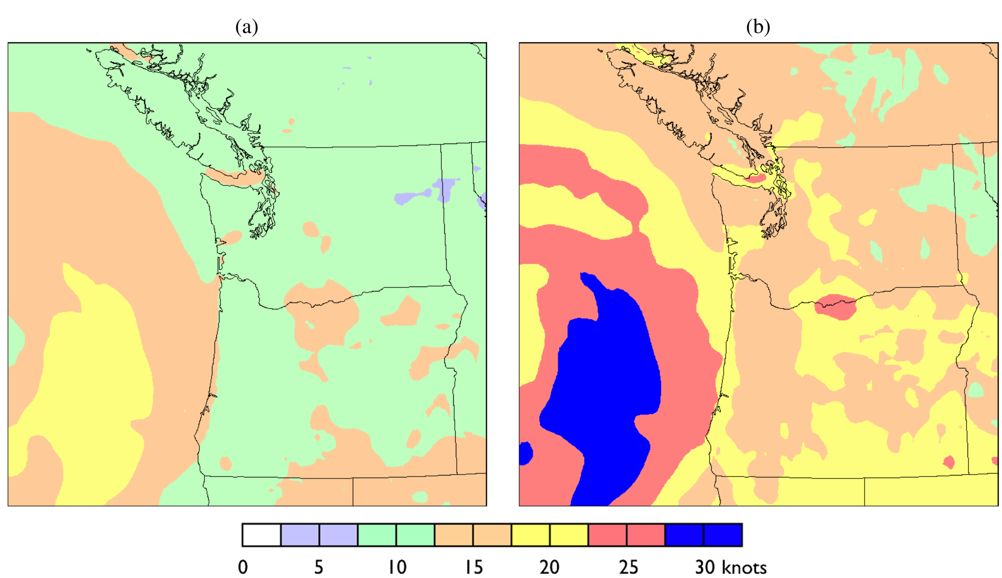

Figure 6. 48-hour-ahead (a) BMA median forecast, and (b) BMA 90th percentile forecast of maximum wind speed over the Pacific Northwest on August 7, 2003.

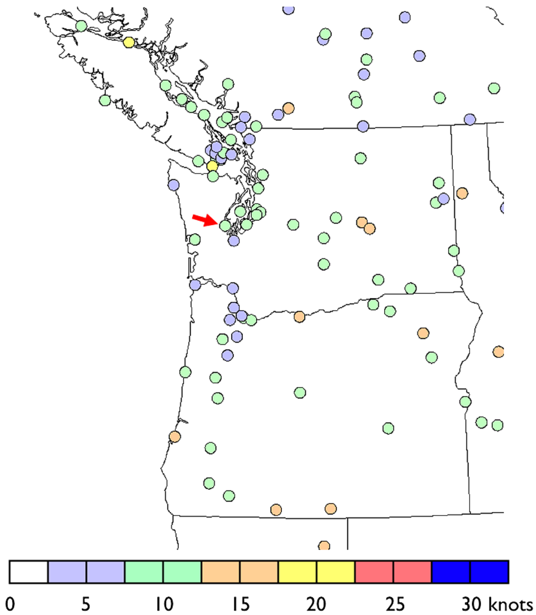

Figure 7. Observed maximum wind speeds at meteorological stations over the Pacific Northwest on August 7, 2003. The arrow indicates the station at Shelton, Washington.

other approaches have used climatology conditioned on ensemble forecasts (Roulston et al. 2003) or on numerical weather prediction model output (Bremnes 2006). The model based approach of BMA allows for the fitting of a predictive distribution using sparse training data. By modeling the relation between forecasts and observations, BMA allows for the creation of forecast distributions in situations where the raw forecasts may not have any close analogues in the training set.

The method described here differs in some respects from the more traditional sense of BMA, and can also be thought of as a mixture model, where the component weights are estimated from training data. Because the training data are changing every day, the component weights are also changing day by day, a paradigm that is in line with operational weather forecasting, where forecasters adapt the attention they give to different models each day based on recent past performance.

Raftery et al. (2007) and Pinson and Madsen (2009) proposed the use of recursive-adaptive techniques for updating forecast parameters each day. These approaches could potentially be adapted to our model as well, potentially reducing the computing burden of fitting the model for successive days.

Here we assume that, for any given training period, the weights for each model are fixed. Given the performance of the BMA method presented here, this assumption seems reasonable. A more complicated model could be considered if there were reason to suspect that the relative value of each ensemble member differed based on differing forecasts. The approach of Greybush, Haupt, and Young (2008) addresses this issue, by restricting training data to atmospheric regimes. It could also be possible to adapt the mixture of experts model from Jacobs et al. (1991) to combine predictive models that are tailored to atmospheric regimes.

Table 2. Mean continuous ranked probability score (CRPS) and mean absolute error (MAE), and coverage and average width of 77.8% central prediction intervals for probabilistic forecasts of wind speed at Shelton, Washington in 2003. Coverage in percent, all other values in knots. The MAE refers to the point forecast given by the median of the respective forecast distribution

Forecast	CRPS	MAE	Coverage	Width
Station climatology	2.64	3.82	77.8	12.0
Ensemble	2.80	3.49	35.1	4.2
BMA	2.14	3.01	78.9	10.0

Our results showed no appreciable difference in performance between parameter estimation methods that took into account the discretization of the data and those that did not. A detailed simulation study looking at the effects of discretization for meteorological data was carried out by Cooley, Nychka, and Naveau (2007). Perrin, Rootzén, and Taesler (2006) found that discretized wind speed values can result in erroneously low standard errors for parameter estimates. However, we do not look at standard errors of our parameter estimates here, as we are interested primarily in prediction, and not in parameter estimation for its own sake. Vrugt, Diks, and Clark (2008) compared using the EM algorithm to using a fully Bayesian, Markov chain Monte Carlo (MCMC) based method to fit the BMA parameters for temperature forecasts, and found that the fully Bayesian approach gave comparable performance to the EM method. Their method allows for the examination of posterior distributions of the BMA parameters, which showed those parameters to have relatively little uncertainty. This is to be expected, as we have large amounts of data at hand in estimating these parameters.

Our results also showed that constraining the mean parameters to be equal across ensemble members did not impact the predictive performance. This suggests that there are no appreciable differences in the ideal mean parameters across members. This result, however, may not hold in other situations. If different ensemble members had distinct patterns of bias it may be necessary to preserve the distinct individual mean parameters. In our situation, by assuming equal mean parameters across members, we are able to greatly reduce the parameter space being considered.

Our approach has not incorporated temporal autocorrelation. While wind speeds in successive time periods are certainly autocorrelated, our method implicitly addresses forecast errors rather than raw wind speeds, because it models the predictive distribution conditional on the numerical forecasts. Exploratory work has shown that there is little or no autocorrelation in forecast errors of wind speeds, so the numerical forecasts effectively take care of the autocorrelation in the wind speeds themselves. To incorporate temporal autocorrelation, the method would have to be made considerably more complicated, and it seems unlikely that this would improve its predictive performance appreciably.

While our implementation has been for a situation where the ensemble members come from clearly distinguishable sources, it is easily modifiable to deal with situations in which ensemble members come from the same model, differing only in some random perturbations, such as the global ensembles that are currently used by the National Centers for Environmental Prediction and the European Centre for Medium-Range Weather Forecasts (Buizza et al. 2005). In these cases, members coming from the same source should be treated as exchangeable, and thus should have equal weight and equal BMA parameter values across members. As our recommended parsimonious method already constrains the mean and standard deviation parameters to be equal across all members, the only change that would need to be made would be to add the constraint that the BMA weights be equal, as described by Raftery (2005, p. 1170) and Wilson et al. (2007).

There are a number of potential wind quantities of interest. Our method was developed for forecasting maximum wind speed over a particular time interval. Instantaneous wind speed, or maximum wind speed over some other time interval, could well have different distributional properties. These differences could possibly be accounted for by fitting a gamma distribution to some power transformation of the wind speed, rather than to the raw wind speed. The power transformation appropriate for any given situation would need to be determined empirically.

Our method produces wind speed forecasts at individual locations, which is the focus of many, and possibly most, applications. As a result we have not had to model spatial correlation between wind speeds, although these definitely are present, as can be seen in Figure 7. However, in applications that involve forecasting wind speeds at more than one location simultaneously, it would be vital to take account of spatial correlation in forecast errors. Such applications include forecasting the maximum wind speed over an area or trajectory, for example for shipping or boating, and forecasting the total energy from several wind farms in a region. Methods for probabilistic weather forecasting at multiple locations simultaneously have been developed for temperature (Gel, Raftery, and Gneiting 2004; Berrocal, Raftery, and Gneiting 2007), for precipitation (Berrocal, Raftery, and Gneiting 2008), and for temperature and precipitation simultaneously (Berrocal et al. 2007). These methods could possibly be extended to wind speeds.

Our method estimates a single set of parameters across the entire domain. Nott et al. (2001) noted that localized statistical postprocessing can address issues of locally varying biases in numerical weather forecasts. A localized version of BMA for temperature, based on taking sets of forecasts and observations within a carefully selected neighborhood, has shown substantial improvement over the global version (Mass et al. 2009), and it is likely that similar improvements would be seen for a localized version of BMA for wind speed.

Either the global or localized parameter estimation can be used to create probabilistic forecasts of wind speed on a spatial grid. These methods have already been implemented for temperature and precipitation and provide real-time probabilistic weather forecasts over the Pacific Northwest (Mass et al. 2009), which are available to the general public at *http://probcast.washington.edu*. We intend to make similar, gridded probabilistic forecasts for wind speed available also. Additionally, for specialized uses where only a single location is of interest, such as a wind farm or a windsurfing or sailing location, BMA parameters could be fit using the same methodology as described in this paper, but restricting it to use only data from that particular location.

[Received November 2008. Revised July 2009.]

REFERENCES

Baars, J. (2005), "Observations QC Summary Page," available at *http://www.atmos.washington.edu/mm5rt/qc_obs/qc_obs_stats.html*. [26]

Berrocal, V. J., Raftery, A. E., and Gneiting, T. (2007), "Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts," *Monthly Weather Review*, 135, 1386–1402. [34]

——— (2008), "Probabilistic Quantitative Precipitation Field Forecasting Using a Two-Stage Spatial Model," *The Annals of Applied Statistics*, 2, 1170–1193. [34]

Berrocal, V. J., Raftery, A. E., Gneiting, T., and Steed, R. C. (2007), "Probabilistic Weather Forecasting for Winter Road Maintenance," Technical Report 511, University of Washington, Dept. of Statistics, Seattle, WA. Available at *http://www.stat.washington.edu/research/reports/*. [34]

Bremnes, J. B. (2004), "Probabilistic Wind Power Forecasts Using Local Quantile Regression," *Wind Energy*, 7, 47–54. [25]

——— (2006), "A Comparison of a Few Statistical Models for Making Quantile Wind Power Forecasts," *Wind Energy*, 9, 3–11. [33]

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984), "Time Series Models to Simulate and Forecast Wind Speed and Wind Power," *Journal of Climate and Applied Meteorology*, 23, 1184–1195. [25]

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y. (2005), "A Comparison of the ECMWF, MSC and NCEP Global Ensemble Prediction Systems," *Monthly Weather Review*, 133, 1076–1097. [25,33]

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840. [33]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1–39. [29]

Eckel, F. A., and Mass, C. F. (2005), "Effective Mesoscale, Short-Range Ensemble Forecasting," *Weather and Forecasting*, 20, 328–350. [26]

Eckel, F., and Walters, M. (1998), "Calibrated Probabilistic Quantitative Precipitation Forecasts Based on the MRF Ensemble," *Weather and Forecasting*, 13, 1132–1147. [30]

Gel, Y., Raftery, A. E., and Gneiting, T. (2004), "Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation (GOP) Method" (with discussion), *Journal of the American Statistical Association*, 99, 575–590. [34]

Genton, M., and Hering, A. (2007), "Blowing in the Wind," *Significance*, 4, 11–14. [25]

Giebel, G., Brownsword, R., and Kariniotakis, G. (2003), "The State-of-the-Art in Short-Term Prediction of Wind Power," Deliverable Report D1.1, Project Anemos. Available at *http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf*. [25]

Gneiting, T. (2008), "Quantiles as Optimal Point Predictors," Technical Report 538, University of Washington, Dept. of Statistics. Available at *http://www.stat.washington.edu/research/reports/*. [25]

Gneiting, T., and Raftery, A. E. (2005), "Weather Forecasting With Ensemble Methods," *Science*, 310, 248–249. [25]

——— (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378. [31]

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007), "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society, Ser. B*, 69, 243–268. [30]

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006), "Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space-Time Method," *Journal of the American Statistical Association*, 101, 968–979. [25]

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. (2008), "Assessing Probabilistic Forecasts of Multivariate Quantities, With Applications to Ensemble Predictions of Surface Winds" (with discussion and rejoinder), *Test*, 17, 211–264. [32]

Greybush, S. J., Haupt, S. E., and Young, G. S. (2008), "The Regime Dependence of Optimally Weighted Ensemble Model Consensus Forecasts of Surface Temperature," *Weather and Forecasting*, 23, 1146–1161. [33]

Grimit, E. P., and Mass, C. F. (2002), "Initial Results of a Mesoscale Short-Range Ensemble Forecasting System Over the Pacific Northwest," *Weather and Forecasting*, 17, 192–205. [25,26]

Grimit, E. P., Gneiting, T., Berrocal, V. J., and Johnson, N. A. (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, 3209–3220. [31]

Hamill, T. M., and Colucci, S. J. (1998), "Evaluation of Eta-RSM Ensemble Probabilistic Precipitation Forecasts," *Monthly Weather Review*, 126, 711–724. [30]

Haslett, J., and Raftery, A. E. (1989), "Space–Time Modelling With Long-Memory Dependence: Assessing Ireland's Wind Power Resource" (with discussion), *Journal of the Royal Statistical Society, Ser. C*, 38, 1–50. [27]

Hennessey, J. P. J. (1977), "Some Aspects of Wind Power Statistics," *Journal of Applied Meteorology*, 16, 119–128. [27]

Hoeting, J. A., Madigan, D. M., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial" (with discussion), *Statistical Science*, 14, 382–401. A corrected version is available at *http://www.stat.washington.edu/www/research/online/hoeting1999.pdf*. [27]

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991), "Adaptive Mixtures of Local Experts," *Neural Computation*, 3, 79–87. [33]

Joslyn, S., and Jones, D. W. (2008), "Strategies in Naturalistic Decision-Making: A Cognitive Task Analysis of Naval Weather Forecasting," in *Naturalistic Decision Making and Macrocognition*, eds. J. M. Schraagen, L. Militello, T. Ormerod, and R. Lipshitz, Aldershot, U.K.: Ashgate Publishing, pp. 183–202. [27]

Justus, C. G., Hargraves, W. R., Mikhail, A., and Graber, D. (1978), "Methods for Estimating Wind Speed Frequency Distributions," *Journal of Applied Meteorology*, 17, 350–353. [27]

Justus, C. G., Hargraves, W. R., and Yalcin, A. (1976), "Nationwide Assessment of Potential Output From Wind-Powered Generators," *Journal of Applied Meteorology*, 15, 673–678. [27]

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [27]

Kretzschmar, R., Eckert, P., Cattani, D., and Eggimann, F. (2004), "Neural Network Classifiers for Local Wind Prediction," *Journal of Applied Meteorology*, 43, 727–738. [25]

Leamer, E. E. (1978), *Specification Searches*, New York: Wiley. [27]

Liu, C., and Rubin, D. B. (1994), "The ECME Algorithm: A Simple Extension of EM and ECM With Faster Monotone Convergence," *Biometrika*, 81, 633–648. [29]

Mass, C., Joslyn, S., Pyle, J., Tewson, P., Gneiting, T., Raftery, A., Baars, J., Sloughter, J. M., Jones, D., and Fraley, C. (2009), "PROBCAST: A Web-Based Portal to Mesoscale Probabilistic Forecasts," *Bulletin of the American Meteorological Society*, 90, 1009–1014. [34]

Matheson, J. E., and Winkler, R. L. (1976), "Scoring Rules for Continuous Probability Distributions," *Management Science*, 22, 1087–1096. [31]

McLachlan, G. J., and Krishnan, T. (1997), *The EM Algorithm and Extensions*, New York: Wiley. [29]

Møller, J. K., Nielsen, H. A., and Madsen, H. (2008), "Time-Adaptive Quantile Regression," *Computational Statistics & Data Analysis*, 52, 1292–1303. [25]

Mullen, S. L., and Buizza, R. (2001), "Quantitative Precipitation Forecasts Over the United States by the ECMWF Ensemble Prediction System," *Monthly Weather Review*, 129, 638–663. [30]

Nielsen, H. A., Madsen, H., and Nielsen, T. S. (2006), "Using Quantile Regression to Extend an Existing Wind Power Forecasting System With Probabilistic Forecasts," *Wind Energy*, 9, 95–108. [25]

Nielsen, H. A., Madsen, H., Nielsen, T. S., Badger, J., Giebel, G., Landberg, L., Sattler, K., and Feddersen, H. (2004), "Wind Power Ensemble Forecasting," in *Proceedings of the 2004 Global Windpower Conference and Exhibition*, Chicago, IL. [25,32]

Nott, D. J., Dunsmuir, W. T. M., Kohn, R., and Woodcock, F. (2001), "Statistical Correction of a Deterministic Numerical Weather Prediction Model," *Journal of the American Statistical Association*, 96, 794–804. [34]

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. [25]

Perrin, O., Rootzén, H., and Taesler, R. (2006), "A Discussion of Statistical Methods Used to Estimate Extreme Wind Speeds," *Theoretical and Applied Climatology*, 85, 203–215. [33]

Pinson, P., and Madsen, H. (2009), "Ensemble-Based Probabilistic Forecasting at Horns Rev," *Wind Energy*, 12, 137–155. [33]

Pinson, P., Chevallier, C., and Kariniotakis, G. N. (2007), "Trading Wind Generation With Short-Term Probabilistic Forecasts of Wind Power," *IEEE Transactions on Power Systems*, 22, 1148–1156. [25]

Pinson, P., Nielsen, H. A., Møller, J. K., Madsen, H., and Kariniotakis, G. N. (2007), "Non-Parametric Probabilistic Forecasts of Wind Power: Required Properties and Evaluation," *Wind Energy*, 10, 497–516. [25]

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005), "Using Bayesian Model Averaging to Calibrate Forecast Ensembles," *Monthly Weather Review*, 133, 1155–1174. [25-27,34]

Raftery, A. E., Karný, M., Andrýsek, J., and Ettler, P. (2007), "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," Technical Report 525, University of Washington, Dept. of Statistics, Seattle, WA. Available at *http://www.stat.washington.edu/research/reports/*. [33]

Roulston, M. S., Kaplan, D. T., Hardenberg, J., and Smith, L. A. (2003), "Using Medium-Range Weather Forecasts to Improve the Value of Wind Energy Production," *Renewable Energy*, 28, 585–602. [25,33]

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C. (2007), "Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging," *Monthly Weather Review*, 135, 3209–3220. [26,27,29]

Stevens, M. J. M., and Smulders, P. T. (1979), "The Estimation of the Parameters of the Weibull Wind Speed Distribution for Wind Energy Utilization Purposes," *Wind Engineering*, 3, 132–145. [27]

Tuller, S. E., and Brett, A. C. (1984), "The Characteristics of Wind Velocity That Favor the Fitting of a Weibull Distribution in Wind Speed Analysis," *Journal of Applied Meteorology*, 23, 124–134. [27]

Vrugt, J. A., Diks, C. G. H., and Clark, M. P. (2008), "Ensemble Bayesian Model Averaging Using Markov Chain Monte Carlo Sampling," *Environmental Fluid Mechanics*, 8, 579–595. [33]

Wilks, D. S. (1990), "Maximum Likelihood Estimation for the Gamma Distribution Using Data Containing Zeros," *Journal of Climate*, 3, 1495–1501. [29]

——— (2006), *Statistical Methods in the Atmospheric Sciences* (2nd ed.), Burlington: Elsevier Academic Press. [31]

Wilson, L. J., Beauregard, S., Raftery, A. E., and Verret, R. (2007), "Reply," *Monthly Weather Review*, 135, 4231–4236. [34]