

Probabilistic Visibility Forecasting Using Bayesian Model Averaging

RICHARD M. CHMIELECKI

Department of Mathematics, United States Coast Guard Academy, New London, Connecticut

ADRIAN E. RAFTERY

Department of Statistics, University of Washington, Seattle, Washington

(Manuscript received 26 May 2010, in final form 7 September 2010)

ABSTRACT

Bayesian model averaging (BMA) is a statistical postprocessing technique that has been used in probabilistic weather forecasting to calibrate forecast ensembles and generate predictive probability density functions (PDFs) for weather quantities. The authors apply BMA to probabilistic visibility forecasting using a predictive PDF that is a mixture of discrete point mass and beta distribution components. Three approaches to developing predictive PDFs for visibility are developed, each using BMA to postprocess an ensemble of visibility forecasts. In the first approach, the ensemble is generated by a translation algorithm that converts predicted concentrations of hydrometeorological variables into visibility. The second approach augments the raw ensemble visibility forecasts with model forecasts of relative humidity and quantitative precipitation. In the third approach, the ensemble members are generated from relative humidity and precipitation alone. These methods are applied to 12-h ensemble forecasts from 2007 to 2008 and are tested against verifying observations recorded at Automated Surface Observing Stations in the Pacific Northwest. Each of the three methods produces predictive PDFs that are calibrated and sharp with respect to both climatology and the raw ensemble.

1. Introduction

The accurate prediction of visibility even over short-term (0–6 h) forecasting periods is challenging. Since most numerical weather prediction models do not explicitly model visibility, forecasts of visibility must first be derived from other meteorological parameters such as cloud water content, relative humidity, and precipitation. There are a handful of different specifications of visibility as a function of other weather parameters (Knapp 1999; Doran et al. 1999; Stoelinga and Warner 1999; Smirnova et al. 2000; Smith and Benjamin 2002; Benjamin et al. 2004; Gultepe and Isaac 2006; Gultepe and Milbrandt 2010). Typically, model output statistics (MOS; Glahn and Lowry 1972) or another statistical postprocessing technique is applied to deterministic forecasts to calibrate them and remove bias.

In the past few decades, researchers have developed several methods of forecasting visibility and postprocessing

forecasts, nearly all focusing on short-term-forecasting horizons. Vislocky and Fritsch (1997) compared the performance of observation-based, MOS-based, and persistence climatology models for short-term deterministic ceiling and visibility forecasts. Leyton and Fritsch (2003, 2004) extended the observation-based approach using high-density and, later, high-frequency networks of surface observations to produce probabilistic forecasts. Pasini et al. (2001), Bremnes and Michaelides (2007), and Marzban et al. (2007) applied neural networks to probabilistic visibility forecasting. Zhou et al. (2009) described the use of a short-range ensemble forecast system to generate probabilistic visibility forecasts. Roquelaure and Bergot (2008, 2009) and Roquelaure et al. (2009) were the first to use Bayesian model averaging (BMA) in visibility forecasting; their studies modeled binary low-visibility outcomes using a local ensemble prediction system at Charles de Gaulle Airport in Paris.

Operationally, aviation-related interests in the United States typically use two forms of ceiling and visibility forecasts provided by the National Weather Service's Global Forecast System (GFS) model: GFS MOS forecasts and GFS Localized Aviation Model Output Statistics Program (LAMP) forecasts. GFS MOS forecasts

Corresponding author address: Richard M. Chmielecki, Department of Mathematics, United States Coast Guard Academy, 15 Mohegan Ave., New London, CT 06320.
E-mail: richard.m.chmielecki@uscga.edu

are generated every 6 h and yield categorical probabilities of ceiling and visibility for forecast periods between 6 and 72 h. GFS LAMP forecasts use observational data to update GFS MOS forecasts between cycles and are available every hour. Both the GFS MOS and GFS LAMP produce probabilistic forecasts for seven ranges of visibility and return to the user a categorical forecast corresponding to the most likely range of values.

However, none of the existing methods provides a general framework for generating a full predictive probability density function (PDF) for visibility. Predictive PDFs are attractive for visibility forecasts because thresholds of ceiling and visibility correspond directly to the conditions governing flight rules. A predictive PDF allows the user to determine the probability of visibility falling below any threshold of interest rather than a single prespecified threshold. Raftery et al. (2005) introduced BMA as a way of producing calibrated predictive PDFs for any weather parameter of interest. The method has now been successfully used to generate probabilistic forecasts for temperature, sea level pressure, quantitative precipitation (Sloughter et al. 2007), wind speed (Sloughter et al. 2010), and wind direction (Bao et al. 2010). Here, BMA is applied to probabilistic visibility forecasting.

Deterministic visibility forecasts are obtained via a translation algorithm that converts numerical weather model information from the University of Washington Mesoscale Ensemble (UWME) into predicted visibility. The UWME translation algorithm incorporates two commonly used parameterizations of visibility: the extinction coefficient parameterization and the clear-air, relative humidity-based parameterization. BMA is then applied to these deterministic forecasts, generating predictive PDFs for visibility at any observing station.

While BMA was initially used for modeling parameters that can be reasonably well described by normal distributions (temperature and sea level pressure), a normal distribution is not appropriate for visibility because of the distinctly nonnormal character of its empirical distribution. At most observing stations, visibility is frequently recorded at the maximum value of 10 miles. If visibility is impaired, its value may range down to 0. However, the distribution of these subnominal values does not have a normal shape and is bounded below by 0 and above by 10. To represent this, we model visibility as a mixture of a point mass at 10 and a beta distribution corresponding to values less than 10. The BMA PDF is then a mixture of the member-specific mixtures of the respective point mass and beta distribution components.

The paper is organized as follows. In section 2, we describe how deterministic visibility forecasts are generated by the UWME visibility translation algorithm.

TABLE 1. UWME members and their acronyms and sources: the National Centers for Environmental Prediction (NCEP), the Canadian Meteorological Centre (CMC), the Australian Bureau of Meteorology (ABM), the Japan Meteorological Agency (JMA), the Fleet Numerical Meteorological and Oceanographic Center (FNMO), the Taiwan Central Weather Bureau (TCWB), and the United Kingdom Meteorological Office (UKMO).

Acronym	Model	Source
GFS	Global Forecast System	NCEP
ETA	Limited-Area Mesoscale model	NCEP
CMCG	Global environmental multiscale	CMC
GASP	Global Analysis and Prediction Model	ABM
JMA	Global Spectral Model	JMA
NGPS	Navy Operational Global Atmospheric Prediction System	FNMO
TCWB	Global Forecast System	TCWB
UKMO	Unified Model	UKMO

We then review the BMA method and its application to visibility forecasting and give some examples of its use. We also discuss how the raw forecasts may be augmented with additional weather information. Finally, we propose a method by which predictive PDFs can be generated without using output from a standard translation algorithm. In section 3, we give results of each method for daily 12-h forecasts of visibility over the Pacific Northwest in 2007–08, and in section 4, we discuss the results.

2. Methods

a. Translation algorithm

The UWME is an eight-member ensemble that runs model physics from the nonhydrostatic fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) using 8 different sets of initial conditions (Eckel and Mass 2005). The eight ensemble components used in this analysis are given in Table 1. Within the UWME, visibility is parameterized according to the method of extinction coefficients described by Stoelinga and Warner (1999) and also by the clear-air Rapid Update Cycle (RUC) parameterization introduced by Smirnova et al. (2000); the forecast visibility is set equal to the lesser of the two resulting values.

The method of extinction coefficients for visibility parameterization uses four hydrometeorological extinction coefficients corresponding to cloud liquid water, cloud ice, rain, and snow, respectively. There is some disagreement in the literature about the values of the extinction coefficients. The extinction coefficients used by the UWME translation algorithm since its inception in 2002 are given in Table 2. The values used for rain and

TABLE 2. Extinction coefficients used by Stoelinga and Warner (1999) and the UWME translation algorithm, where C is mass concentration (g m^{-3}).

Hydrometeor	Stoelinga and Warner (1999)	UWME
Cloud liquid water	$144.7 C^{0.88}$	$144.7 C^{0.88}$
Rain	$1.1 C^{0.75}$	$2.24 C^{0.75}$
Cloud ice	$163.9 C^{1.00}$	$327.8 C^{1.00}$
Snow	$10.4 C^{0.78}$	$10.4 C^{0.78}$

cloud ice are those used by Bacon et al. (2002) but differ from the numbers suggested by Stoelinga and Warner (1999).

The clear-air parameterization models visibility as a decaying exponential function of relative humidity. Following Smirnova et al. (2000), the visibility forecast y at site s and time t associated with ensemble member k is given by

$$y_{kst} = 60 \exp(-2.5q_{rh}), \quad (1)$$

where

$$q_{rh} = \min(80, h_{kst}/100 - 0.15), \quad (2)$$

and h_{kst} is the corresponding member-specific model-based relative humidity forecast.

b. Bayesian model averaging

BMA models the predictive PDF of a weather quantity of interest y as a mixture of conditional PDFs $h_k(y|f_k)$, each PDF corresponding to one of the member-specific forecasts f_k . The BMA predictive PDF is given by

$$p(y|f_1, \dots, f_k) = \sum_{k=1}^K w_k h_k(y|f_k), \quad (3)$$

where w_k is the weight of the k th ensemble member and corresponds to its relative skill over a training period. The BMA weights are constrained to be nonnegative and sum to 1.

To determine what type of probability distribution is suitable for $h_k(y|f_k)$, we look at the collective distribution of station observations. A plot of the empirical distribution of daily 0000 UTC observations from the Pacific Northwest in 2007 and 2008 shows that a mixture of a discrete probability mass at 10 miles and a continuous probability distribution appears to be appropriate for $h_k(y|f_k)$. The majority (77%) of all observations is exactly at 10 miles—see the histogram in Fig. 1a.

The observations are also highly discretized. The Automated Surface Observing Station standard for

reporting visibility is in quarter-mile increments for observations of 0–2 miles, half-mile increments for observations of 2–3 miles, and 1-mile increments for observations greater than 3 miles. Figure 1b shows only observations below 10 miles and preserves the original bin widths according to the measurement rules for observations.

Figure 1c shows a histogram of the collective ensemble forecasts. There are a few significant differences between the distribution of forecasts and observations. First, and most obviously, the scales of the two distributions are quite different. Although actual visibility is often much higher than 10 miles, a feature correctly indicated by the forecasts, its recorded value is limited to 10 miles. Also, forecasts are continuous rather than discretized. We discuss potential issues caused by the discretization of observations in section 2d.

Given the above, we consider a two-component model for the conditional PDF $h_k(y|f_k)$. The first part consists of a point mass at 10 miles and corresponds to the probability that the recorded visibility is 10 miles conditionally on the k th forecast in the ensemble. We model this probability using logistic regression on the square root of the forecast as follows:

$$\text{logit } P(y = 10|f_k) \equiv \log \frac{P(y = 10|f_k)}{P(y < 10|f_k)} = a_{0k} + a_{1k} f_k^{1/2}. \quad (4)$$

Here, the square root of the forecast is used because we found it was a better predictor of $P(y = 10)$ than the raw forecast. Given the skewed nature of the distribution of the forecasts, a transformation of this form seems reasonable.

The second component of the model assigns a member-specific PDF to visibility given that it is less than 10. We use a beta distribution for this portion of the model because of its flexibility over a bounded range of values. Since the beta distribution is defined on $(0, 1)$, one must either rescale the observations, or equivalently, apply a scale transformation to the PDF resulting in a distribution defined on $(0, 10)$.

The beta distribution has two parameters, and we use the form introduced by Mielke (1975), in which one parameter can be interpreted as a location parameter and the other as a scale parameter. These correspond to the forecast and the uncertainty associated with it.

The PDF of the beta distribution is

$$p(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}, \quad (5)$$

for $0 \leq y \leq 1$, where the mean and variance of the distribution are $\alpha/(\alpha + \beta)$ and $\alpha\beta/(\alpha + \beta)^2(\alpha + \beta + 1)$,

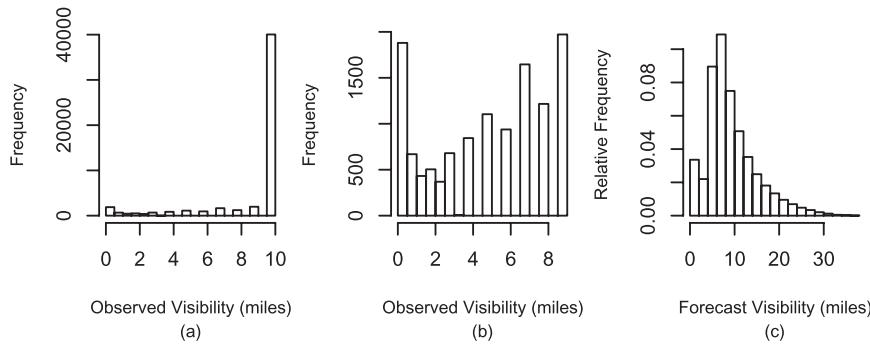


FIG. 1. Histograms of visibility observations and forecasts: (a) all observations, (b) observations <10 miles, and (c) collective ensemble forecasts.

respectively. Mielke suggested making the substitutions $p = \alpha/(\alpha + \beta)$ and $\gamma = \alpha + \beta$ resulting in a PDF of the form

$$p(y) = \frac{\Gamma(\gamma)}{\Gamma(p\gamma)\Gamma[\gamma(1-p)]} y^{p\gamma-1} (1-y)^{(1-p)\gamma-1}, \quad (6)$$

for $0 < p < 1$ and $\gamma > 0$, where the mean and variance are given by p and $p(1-p)/(\gamma + 1)$, respectively.

We found that the mean p and standard deviation $\sqrt{p(1-p)/(\gamma + 1)}$ of this beta distribution were both approximately linear as a function of the square-root-transformed forecast. Thus, we specify the mean and the standard deviation as

$$\mu_k = b_{0k} + b_{1k} f_k^{1/2}, \quad (7)$$

and

$$\sigma_k = c_{0k} + c_{1k} f_k^{1/2}. \quad (8)$$

Putting the two components of the model together we arrive at a final conditional PDF for visibility, given the forecast f_k ; namely,

$$h_k(y|f_k) = P(y < 10|f_k)g_k(y|f_k)I(y < 10) + P(y = 10|f_k)I(y = 10), \quad (9)$$

where $g_k(y|f_k)$ is a member-specific beta distribution.

As has been found in previous applications of BMA to weather prediction (Wilson et al. 2007; Sloughter et al. 2007), the variance parameters c_{0k} and c_{1k} did not vary much across ensemble members and were constrained to be constant. The imposition of this constraint has the two beneficial effects of avoiding potential model overfitting and improving computational efficiency by restricting the number of parameters to be estimated.

The final BMA model for the predictive PDF of visibility y is then

$$p(y|f_1, \dots, f_K) = \sum_{k=1}^K w_k [P(y < 10|f_k)g_k(y|f_k)I(y < 10) + P(y = 10|f_k)I(y = 10)], \quad (10)$$

where

$$g_k(y|f_k) = \frac{\Gamma(\gamma_k)}{\Gamma(p_k\gamma_k)\Gamma[\gamma_k(1-p_k)]} y^{p_k\gamma_k-1} (1-y)^{(1-p_k)\gamma_k-1}. \quad (11)$$

c. Parameter estimation

Parameters are estimated from a set of training data preceding the forecasting period of interest. We discuss choosing the length of the training period in section 2g below. The member-specific parameters a_{0k} and a_{1k} are estimated from a logistic regression model using the square root of the member forecasts in the training period as predictors and a vector of binary indicators of $y = 10$ as outcomes. The parameters b_{0k} and b_{1k} are estimated via linear regression, where observations of visibility less than 10 miles are regressed on the square-root-transformed forecasts.

The weight parameters w_1, \dots, w_K and standard deviation parameters c_0 and c_1 are estimated using the method of maximum likelihood (Fisher 1922). The log-likelihood function for the BMA model is

$$\ell(w_1, \dots, w_K; c_0, c_1) = \sum_{s,t} \log p(y_{st}|f_{1st}, \dots, f_{Kst}), \quad (12)$$

where s and t denote the sites and times, respectively, of observations in the training data.

Since the log-likelihood function cannot be maximized directly, we obtain maximum likelihood estimates using the expectation-maximization (EM) algorithm (Dempster et al. 1977; McLachlan and Krishnan 1997).

The EM algorithm is a general method of finding maximum likelihood estimates in a missing data or latent variable problem. In our current application, the missing data consist of a vector of binary indicators corresponding to whether forecast k is the most skillful forecast among ensemble members for site s at time t . We introduce unobserved quantities z_{kst} such that $z_{kst} = 1$ if forecast k is the best forecast for verification site s at time t and $z_{kst} = 0$ otherwise. For each value of s and t , only one of $z_{1st}, \dots, z_{kst} = 1$. Although the true values are binary, their estimates \hat{z}_{kst} are real numbers between 0 and 1 such that $\sum_{k=1}^K \hat{z}_{kst} = 1$ for each s and t .

In the E step of the algorithm, which computes the expected complete-data log likelihood given current parameter estimates, values of z_{kst} are estimated using the equation

$$\hat{z}_{kst}^{(j+1)} = \frac{w_k^{(j)} p^{(j)}(y_{st} | f_{kst})}{\sum_{l=1}^K w_l^{(j)} p^{(j)}(y_{st} | f_{kst})}. \quad (13)$$

Here, the superscript j refers to the j th iteration of the algorithm, $w_k^{(j)}$ refers to the estimate of w_k at the j th iteration, and $p^{(j)}(y_{st} | f_{kst})$ is evaluated using standard deviation estimates c_0 and c_1 from the previous iteration.

In the M step, parameters are updated by maximizing the expected complete-data log likelihood as a function of w_1, \dots, w_K, c_0 , and c_1 . Updated weight estimates are obtained using the equation

$$w_k^{(j+1)} = \frac{1}{n} \sum_{s,t} \hat{z}_{kst}^{(j+1)}, \quad (14)$$

where n is the number of cases in the training set. However, there is no analytic expression for updating c_0 and c_1 , and so we obtain estimates of these by numerically maximizing the expected complete-data log likelihood using current estimates of w_1, \dots, w_K .

In this application, we modify the EM algorithm by repeating the E step 50 times within each iteration, as was done by Sloughter et al. (2010). This is the so-called ECME algorithm (Liu and Rubin 1994). This modification reduces the number of numerical optimization steps needed for convergence and thus improves computational efficiency. Sloughter et al. (2007) found that supplying initial parameter values for day $t + 1$ equal to the converged solution for day t usually yielded good results; we also found that this method worked well. The EM algorithm is stopped when successive differences in either of the log likelihood or the weight or variance parameters are tolerably small.

d. Discretization of observations

We noted above that the observed values of visibility are discretized; this poses a potential problem since we model them as being generated in part by a continuous distribution. The legitimacy of modeling discrete observations with continuous distributions has been explored previously in terms of other weather parameters (Wilks 1990; Sloughter et al. 2010), and here we apply one of those methods to assess whether the discretized visibility observations are appropriately modeled using a continuous distribution.

In their application of BMA to wind speed forecasting, Sloughter et al. (2010) represented each component of the log likelihood not by the observation itself but by the integrated probability of the range of values that are rounded to the observed value. As an example using visibility, an observation of 8 miles is replaced by the integral of the BMA PDF between 7.5 and 8.5 miles because observed values falling within that range are rounded to the recorded value of 8. We refer to this as the “discretized” model.

e. Augmented model

We now assess whether adding additional predictors results in improved BMA forecasts. The available predictors were surface forecasts of sea level pressure, temperature, dewpoint, relative humidity, mixing ratio, ceiling, u and v component and resultant wind speed, maximum hourly and instantaneous wind speed, and quantitative precipitation for 3, 6, and 12 h. We found that only relative humidity and the quantitative precipitation forecasts were significantly associated with observed visibility, so we limit the inclusion of additional predictors to these only.

We considered each of several standard transformations for the precipitation and relative humidity variables for inclusion in each component of the model. We found that the square root of 3-h quantitative precipitation and exponentiated relative humidity provided the best logistic regression and linear regression fits, respectively. Precipitation has previously been modeled using a cubed or fourth-root transformation (Hamill et al. 2004; Sloughter et al. 2007), but here we found that a square root transformation performed best for visibility. Thus, the model for the binary outcome $y = 10$ is represented by the equation

$$\begin{aligned} \text{logit } P(y = 10 | f_k, r_k, p_k) \\ = a_{0k} + a_{1k} f_k^{1/2} + a_{2k} p_k^{1/2} + a_{3k} \exp(r_k), \end{aligned} \quad (15)$$

where f_k , p_k , and r_k are the member-specific visibility, quantitative precipitation, and relative humidity forecasts,

respectively. Given that $y < 10$, the mean and standard deviation of the associated member-specific beta distribution are specified by the relationships

$$\mu_k = b_{0k} + b_{1k}f_k^{1/2} + b_{2k}p_k^{1/2} + b_{3k}\exp(r_k) \quad (16)$$

and

$$\sigma_k = c_0 + c_1f_k^{1/2} + c_2\exp(r_k). \quad (17)$$

Precipitation is omitted in the standard deviation model since including it did not significantly contribute to estimating the standard deviation.

f. No-forecast model

Lastly, we consider a model in which we do not perform BMA on ensemble visibility forecasts but rather member-specific regression-based estimates of visibility using relative humidity and quantitative precipitation as the predictors. In this model, we predict $y = 10$ in the same manner as above except that visibility is removed from the model specification:

$$\text{logit } P(y = 10|r_k, p_k) = a_{0k} + a_{1k}p_k^{1/2} + a_{2k}\exp(r_k). \quad (18)$$

Similarly, we specify the component corresponding to $y < 10$ by the relationships

$$\mu_k = b_{0k} + b_{1k}p_k^{1/2} + b_{2k}\exp(r_k), \quad (19)$$

and

$$\sigma_k = c_0 + c_1\exp(r_k). \quad (20)$$

The performance of this model is of particular interest; since it consists of a different parameterization of visibility as a function of relative humidity and precipitation, it is a closely related competitor of the translation algorithm currently used.

g. Model scoring and training period length

The measures we considered for comparing model performance were the Brier score (BS) for the binary outcomes $y = 10$ and $y \leq 3$, the continuous ranked probability score (CRPS) of the BMA PDF, and the mean absolute error (MAE) of the BMA deterministic forecast, equal to the median of the predictive PDF.

Here, the BS (Brier 1950) is defined as the mean-squared error of the forecast probabilities for $y = 10$ and $y \leq 3$, respectively, given by

$$\text{BS} = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2, \quad (21)$$

where n is the number of observations, p_i is the forecasted probability of either $P(y = 10)$ or $P(y \leq 3)$, and o_i is the associated binary indicator of $y = 10$ or $y \leq 3$. The Brier score takes a value between 0 and 1—greater skill in forecasts is associated with a lower Brier score.

The CRPS (Grimm et al. 2006; Wilks 2006; Gneiting and Raftery 2007) is a proper scoring rule defined as

$$\begin{aligned} \text{crps}(P, x) &= \int_{-\infty}^{\infty} [P(y) - I(y \geq x)]^2 dy \\ &= E_P|X - x| - \frac{1}{2}E_P|X - X'|, \end{aligned} \quad (22)$$

where P is the predictive cumulative distribution function, x is the observed visibility, and X and X' are independent random variables with distribution P . CRPS is a generalization of MAE that assesses the sharpness of the predictive PDF over its entire range, and it is thus a more general measure of model fit than the Brier score.

We chose the length of the training period by comparing the above scores for models using training period lengths between 15 and 50 days. We found that training periods of 25–30 days performed best. Figure 2 shows CRPS and MAE scores for the standard BMA model evaluated on data from the years 2007 and 2008 for various training period lengths.

h. Examples

We give three examples of BMA applied to stations in the Pacific Northwest. These examples are representative of the different shapes of predictive PDFs commonly generated by BMA and show that BMA effectively calibrates the ensemble, which is typically underdispersed and occasionally quite inaccurate. Each of the results presented is for the standard BMA model using a training period length of 25 days. Figure 3 shows the BMA PDFs and contributions from the weighted ensemble PDFs. We note that the probabilities corresponding to the beta distribution component of the PDF do not reflect the fact that the beta distribution has been scaled to accommodate observations from 0 to 10 miles. We favor this display, however, because it correctly indicates the relative magnitudes of the probabilities associated with the different components. Table 3 shows details of each ensemble forecast and the corresponding BMA output.

The first example is for Station KONP in Newport, Oregon, on 6 May 2008. The range of ensemble forecasts

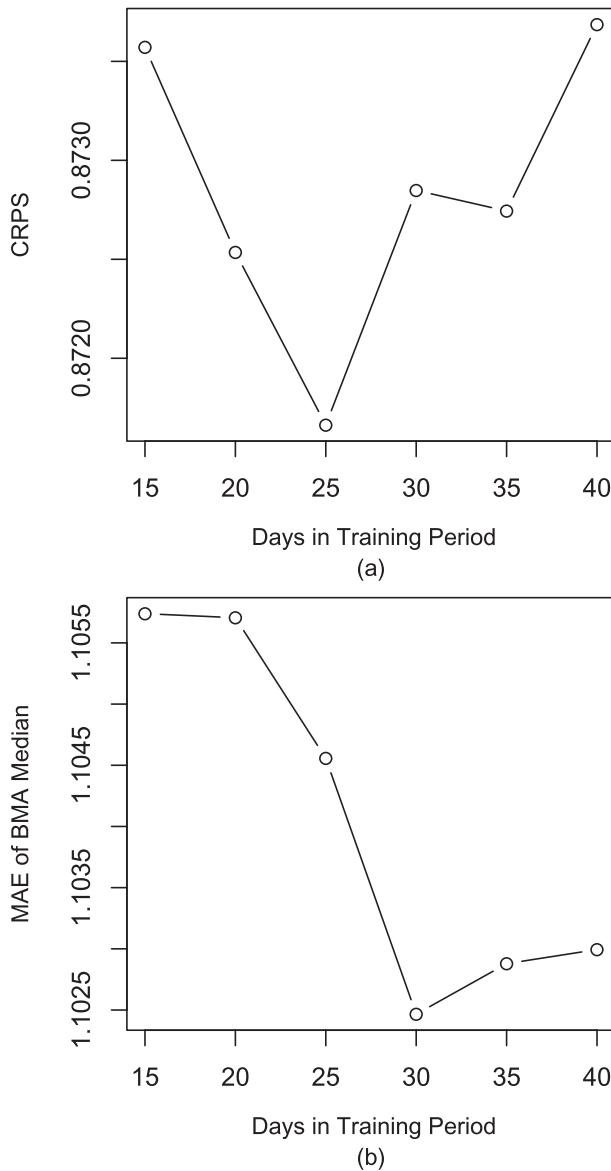


FIG. 2. Comparison of training period lengths: (a) CRPS of BMA forecasts and (b) MAE of BMA deterministic forecast (median).

was approximately 4.6–6.9 miles, and the verifying observation was 7 miles. Although the observation lay outside the range of ensemble forecasts, it was within the 80% central predictive interval given by BMA, the 10th and 90th quantiles of which were 5.57 and 10 miles, respectively. The concave-down distribution for the $y < 10$ component of the predictive PDF was one of the common shapes observed in this dataset.

The second example is for Station KNUW in Oak Harbor, Washington, on 14 November 2007. In this example, the range of the ensemble was roughly 13–19 miles and the verifying observation was 7 miles. Here,

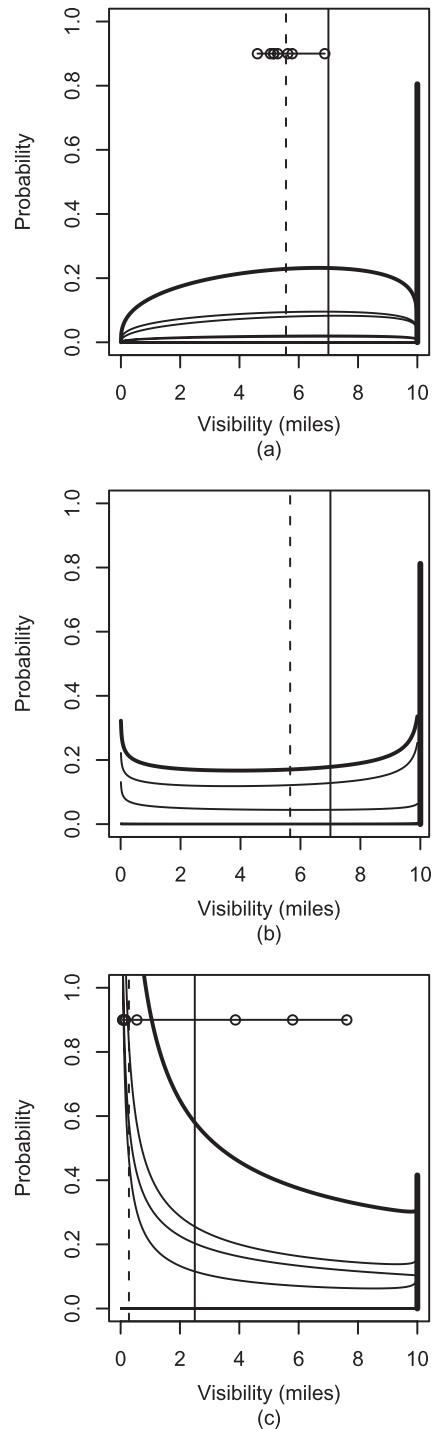


FIG. 3. BMA PDFs for (a) station KONP on 6 May 2008, (b) station KNUW on 14 Nov 2007, and (c) station KBLI on 23 Jul 2007. The vertical line at 10 represents the BMA probability of visibility of 10 miles, the vertical line is the verifying observation, and the vertical dashed line is the 10th percentile of the BMA predictive PDF. The thick curves are the BMA PDF of visibility given that it is < 10 miles, and the thin curves represent the individual ensemble contributions toward the weighted BMA. Dots represent the ensemble member forecasts.

TABLE 3. Three examples of the BMA method. The member forecast, BMA median, BMA lower bound, and observation fields are in units of miles. The BMA lower bound corresponds to the 10th quantile of the BMA predictive PDF.

	BMA	GFS	CMCG	ETA	GASP	JMA	NGPS	TCWB	UKMO
Station KONP on 6 May 2008									
Member forecast		5.15	5.16	4.61	5.63	5.29	5.78	6.88	5.05
BMA weight		0.10	0.00	0.40	0.41	0.00	0.00	0.00	0.08
Member $P(y = 10)$		0.83	0.79	0.80	0.84	0.83	0.83	0.86	0.81
BMA $P(y = 10)$	0.80								
BMA median	10								
BMA lower bound	5.57								
Observation	7								
Station KNUW on 14 Nov 2007									
Member forecast		17.39	18.32	19.35	13.10	15.95	18.23	15.91	16.40
BMA weight		0.00	0.76	0.00	0.00	0.00	0.01	0.00	0.23
Member $P(y = 10)$		0.77	0.82	0.81	0.73	0.73	0.82	0.73	0.78
BMA $P(y = 10)$	0.81								
BMA median	10								
BMA lower bound	5.66								
Observation	7								
Station KBLI on 23 Jul 2007									
Member forecast		0.16	3.87	0.14	7.62	5.80	0.08	0.55	0.07
BMA weight		0.33	0.00	0.00	0.00	0.00	0.41	0.00	0.26
Member $P(y = 10)$		0.39	0.72	0.50	0.88	0.83	0.35	0.51	0.47
BMA $P(y = 10)$	0.42								
BMA median	7.32								
BMA lower bound	0.28								
Observation	2.5								

although the observation fell rather far outside of the range of the ensemble, it was still within the 80% central predictive interval given by the BMA PDF. The U-shaped beta distribution in this example was another of the common shapes for $y < 10$ seen in our analysis.

The third example is for Station KBLI in Bellingham, Washington, on 23 July 2007, and it shows a scenario in which there is considerable disagreement in the forecasts. While some of the members forecast visibility very close to 0, there are forecasts as high as 7 miles. In this example, BMA correctly favors the members with lower forecasts, resulting in a BMA lower bound of 0.28 miles. Here, the verifying observation is within both the range of the ensemble members and the BMA predictive interval.

3. Results

We now compare the performance of the three models on 12-h forecasts from the years 2007 and 2008 verifying at 0000 UTC daily. For each of the model runs we used a training period length of 25 days, which was found to be roughly optimal for this data. A total of 52 303 station observations were available over 578 days, with the first 25 days of 2007 being reserved for training data. In the calculation of performance measures, all

ensemble forecasts greater than 10 miles were set to 10 to facilitate comparison between BMA and the ensemble.

While point forecasts from the BMA model are obtained by evaluating the median of the predictive PDF, it remains to be specified how comparable point forecasts are derived from both climatology and the raw ensemble. In the results that follow, we use the median of observations in the training set at each time step as the climatological point forecast and the empirical distribution of observations in the training set as the climatological predictive distribution. Similarly, we take the median of ensemble forecasts at each station date to be the point forecast associated with the raw ensemble while the set of forecasts represents the predictive distribution of the ensemble.

First, we consider the performance of the point mass component of the model, $P(y = 10)$. Figure 4a shows a reliability diagram (Wilks 2006, section 7.4.4) comparing the calibration of the raw ensemble to the standard BMA model. While the raw ensemble is severely underdispersed with respect to forecasts of $P(y = 10)$, BMA is well calibrated throughout the range of forecast probabilities. We also assess the calibration of the standard BMA model versus the ensemble for the operationally important threshold of 3 miles. Figure 4b

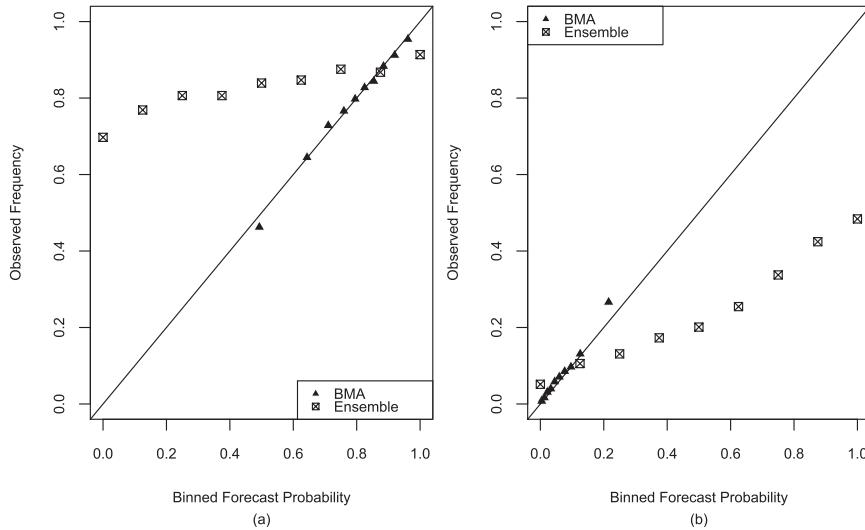


FIG. 4. Reliability diagram of binned forecast of (a) $P(y = 10)$ vs observed relative frequency of $y = 10$ for the raw ensemble and BMA, and (b) $P(y \leq 3)$ vs observed relative frequency of $y \leq 3$ for the raw ensemble and BMA.

indicates that BMA is calibrated at this threshold as well, whereas the ensemble is again underdispersed. The Brier score with respect to forecasts of $P(y = 10)$ for the models we considered did not vary much from 0.15, compared with scores of 0.16 and 0.45 for climatology and the raw ensemble, respectively. For this portion of the model BMA outperforms sample climatology, and improves substantially over the ensemble, which is uncalibrated and underdispersed.

We next evaluate the performance of the complete predictive PDF generated by the standard BMA model. Figure 5a shows a verification rank histogram for the raw ensemble. A verification rank histogram resembling the shape of the uniform distribution would indicate a

calibrated ensemble; however, we see that the ensemble is far from calibrated and quite underdispersed since the verifying observation often falls either below or above all the ensemble forecasts. Figure 5b shows the Probability Integral Transform (PIT) histogram for the BMA forecasts; its nearly uniform distribution indicates that BMA is well calibrated over the range of the predictive PDF and indicates a substantial improvement over the raw ensemble. To generate the PIT histogram, each BMA cumulative distribution function was evaluated at the corresponding observation; for observations of 10 miles the resulting probability was sampled randomly from a uniform distribution on the interval between the quantity $[1 - P(y = 10)]$ and 1.

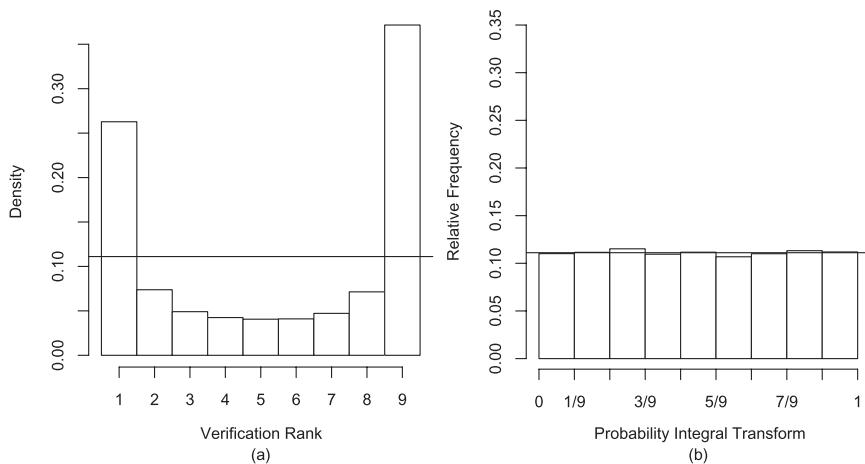


FIG. 5. (a) Verification rank histogram for raw ensemble forecasts and (b) PIT histogram for BMA forecast.

TABLE 4. Comparison of model performance using a 25-day training period and data from the 2007 and 2008 calendar years. BS is the Brier Score for the given threshold value, coverage and width refer to the 77.8% central predictive interval, and MAE is taken with respect to the BMA deterministic forecast (median).

Score	CRPS	BS	BS	MAE	Coverage	Width
Threshold		10	3			
Climatology	0.922	0.163	0.072	1.126	0.778	3.97
Ensemble	1.886	0.452	0.083	2.248	0.543	2.81
Standard BMA	0.872	0.151	0.069	1.105	0.790	3.77
Standard BMA (discretized)	0.872	0.151	0.068	1.105	0.791	3.77
Augmented model	0.869	0.150	0.069	1.106	0.788	3.66
No-forecast model	0.901	0.156	0.071	1.122	0.785	3.82

Table 4 shows model performance scores for the standard BMA model, the discretized model, and the augmented and no-forecast models. The sharpness of the predictive model is calculated with respect to the width of the central $\frac{7}{9}$ or 77.8% predictive interval. This is an appropriate interval for comparison between BMA and the ensemble because a perfectly calibrated ensemble would result in forecasts falling within the range of member forecasts $\frac{7}{9}$ of the time.

The standard BMA model performs better than both climatology and the raw ensemble across all scores. It produces wider intervals than the ensemble, but this is because it is calibrated and the ensemble is underdispersed. While each of the BMA models is well calibrated, the augmented model yields a slight improvement over the standard BMA approach with respect to CRPS and predictive interval width. Also of note is that the discretized model does not perform substantially better than the standard model. This suggests that using a continuous model for the observations is acceptable in spite of their discretized nature. Last, we note that the parameterization of visibility using only relative humidity and precipitation results in forecasts that are slightly less good than those from the translation algorithm.

While adding additional predictors in the augmented model did result in a slight improvement in model performance, it is perhaps not surprising that the improvement was relatively minor. The reason for this is that the predictors added into the model contain information quite similar to what is already used by the model, albeit in a slightly different form. Given that the augmented model is both more complex and computationally intensive, the standard BMA approach appears to be adequate.

4. Discussion

We have proposed a method for using BMA to generate predictive PDFs for visibility. The resulting predictive PDFs are calibrated and sharp with respect to the raw ensemble and climatology. This methodology

resolves a number of the shortcomings of the ensemble forecasts including their underdispersion and the discrepancy in scale between forecasts and observations.

We also explored a method of using additional predictors to increase precision in the BMA forecasts. Though we used relative humidity and quantitative precipitation here, because they were the most informative predictors available in our data, this framework can be easily modified to include any other set of predictors. In this application, we found that the relative humidity and precipitation forecasts added a small amount of information to the visibility forecasts obtained from the translation algorithm.

Last, we described a method by which deterministic parameterizations of visibility not derived from a typical translation algorithm might be converted into predictive PDFs using BMA. Although this model did not perform as well as the standard BMA model, it proposed a framework for calibrating regression-based visibility forecasts generated from any set of predictors.

Acknowledgments. The authors thank Richard Steed, Clifford Mass, and Jeff Baars for helpful discussions and access to data from the UWME. 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.

REFERENCES

- Bacon, D. P., Z. Boybeyi, and R. A. Sarma, 2002: Aviation forecasting using adaptive unstructured grids. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., JP1.27. [Available online at <http://vortex.atgteam.com/papers/2002/10-AvWX-JP127.pdf>.]
- Bao, L., T. Gneiting, E. P. Gritter, P. Guttorp, and A. E. Raftery, 2010: Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Mon. Wea. Rev.*, **138**, 1811–1821.
- Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518.

- Bremnes, J. B., and S. C. Michaelides, 2007: Probabilistic visibility forecasting using neural networks. *Pure Appl. Geophys.*, **164**, 1365–1382.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, **39B**, 1–39.
- Doran, J. A., P. J. Roohr, D. J. Beberwyk, G. R. Brooks, G. A. Gayno, R. T. Williams, J. M. Lewis, and R. J. Lefevre, 1999: The MM5 at the AF Weather Agency—New products to support military operations. Preprints, *Eighth Conf. on Aviation, Range, and Aerospace Meteorology*, Dallas, TX, Amer. Meteor. Soc., 115–119.
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.
- Fisher, R. A., 1922: On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London*, **A222**, 309–368.
- Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211.
- Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378.
- Grimt, 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. *Quart. J. Roy. Meteor. Soc.*, **132**, 3209–3220.
- Gultepe, I., and G. A. Isaac, 2006: Visibility versus precipitation rate and relative humidity. Preprints, *12th Cloud Physics Conf.*, Madison, WI, Amer. Meteor. Soc., P2.55. [Available online at http://ams.confex.com/ams/Madison2006/techprogram/paper_113177.htm.]
- , and J. A. Milbrandt, 2010: Probabilistic parameterizations of visibility using observations of rain precipitation rate, relative humidity, and visibility. *J. Appl. Meteor. Climatol.*, **49**, 36–46.
- Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447.
- Knapp, D., 1999: An advanced algorithm to diagnose atmospheric turbulence using numerical model output. Preprints, *16th Conf. on Weather Analysis and Forecasting*, Phoenix, AZ, Amer. Meteor. Soc., 79–81.
- Leyton, S. M., and J. M. Fritsch, 2003: Short-term probabilistic forecasts of ceiling and visibility utilizing high-density surface weather observations. *Wea. Forecasting*, **18**, 891–902.
- , and —, 2004: The impact of high-frequency surface weather observations on short-term probabilistic forecasts of ceiling and visibility. *J. Appl. Meteor.*, **43**, 145–156.
- Liu, C., and D. B. Rubin, 1994: The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**, 633–648.
- Marzban, C., S. M. Leyton, and B. Colman, 2007: Ceiling and visibility forecasts via neural networks. *Wea. Forecasting*, **22**, 466–479.
- McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, 274 pp.
- Mielke, P. W., 1975: Convenient beta distribution likelihood techniques for describing and comparing meteorological data. *J. Appl. Meteor.*, **14**, 985–990.
- Pasini, A., V. Pelino, and S. Potesta, 2001: A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *J. Geophys. Res.*, **106**, 14 951–14 959.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.
- Roquelaure, S., and T. Bergot, 2008: A local ensemble prediction system for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.*, **47**, 3072–3088.
- , and —, 2009: Contributions from a Local Ensemble Prediction System (LEPS) for improving low cloud forecasts at airports. *Wea. Forecasting*, **24**, 39–52.
- , R. Tardif, S. Remy, and T. Bergot, 2009: Skill of a ceiling and visibility Local Ensemble Prediction System (LEPS) according to fog-type prediction at Paris-Charles de Gaulle Airport. *Wea. Forecasting*, **24**, 1511–1523.
- Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220.
- , T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35.
- Smirnova, T. G., S. G. Benjamin, and J. M. Brown, 2000: Case study verification of RUC/MAPS fog and visibility forecasts. Preprints, *Ninth Conf. on Aviation, Range, and Aerospace Meteorology*, Orlando, FL, Amer. Meteor. Soc., 2.3.
- Smith, T. L., and S. G. Benjamin, 2002: Visibility forecasts from the RUC20. Preprints, *10th Conf. on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc., JP1.27.
- Stoelinga, M. T., and T. T. Warner, 1999: Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event. *J. Appl. Meteor.*, **38**, 385–404.
- Vislocky, R. L., and J. M. Fritsch, 1997: An automated, observations-based system for short-term prediction of ceiling and visibility. *Wea. Forecasting*, **12**, 116–122.
- Wilks, D. S., 1990: Maximum likelihood estimation for the gamma distribution using data containing zeros. *J. Climate*, **3**, 1495–1501.
- , 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. Elsevier Academic Press, 627 pp.
- Wilson, L. J., S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 1364–1385.
- Zhou, B., J. Du, J. McQueen, and G. Dimego, 2009: Ensemble forecast of ceiling, visibility, and fog with NCEP Short-Range Ensemble Forecast system (SREF). Preprints, *Aviation, Range, and Aerospace Meteorology Special Symp. on Weather–Air Traffic Management Integration*, Phoenix, AZ, Amer. Meteor. Soc., 4.5. [Available online at http://www.emc.ncep.noaa.gov/mmb/SREF/AMS_2009CV_fog.pdf.]

Copyright of Monthly Weather Review is the property of American Meteorological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.