# Probabilistic Precipitation Forecasting Based on Ensemble Output Using Generalized Additive Models and Bayesian Model Averaging

YANG Chi[1,2*](杨　赤), YAN Zhongwei[2](严中伟), and SHAO Yuehong[2](邵月红)

1 *College of Global Change and Earth System Science, Beijing Normal University, Beijing* 100875

2 *Key Laboratory of Regional Climate-Environment Research for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

(Received November 23, 2010; in final form December 9, 2011)

**ABSTRACT**

A probabilistic precipitation forecasting model using generalized additive models (GAMs) and Bayesian model averaging (BMA) was proposed in this paper. GAMs were used to fit the spatial-temporal precipitation models to individual ensemble member forecasts. The distributions of the precipitation occurrence and the cumulative precipitation amount were represented simultaneously by a single Tweedie distribution. BMA was then used as a post-processing method to combine the individual models to form a more skillful probabilistic forecasting model. The mixing weights were estimated using the expectation-maximization algorithm. The residual diagnostics was used to examine if the fitted BMA forecasting model had fully captured the spatial and temporal variations of precipitation. The proposed method was applied to daily observations at the Yishusi River basin for July 2007 using the National Centers for Environmental Prediction ensemble forecasts. By applying scoring rules, the BMA forecasts were verified and showed better performances compared with the empirical probabilistic ensemble forecasts, particularly for extreme precipitation. Finally, possible improvements and application of this method to the downscaling of climate change scenarios were discussed.

**Key words:** Bayesian model averaging, generalized additive model, probabilistic precipitation forecasting, TIGGE, Tweedie distribution

**Citation:** Yang Chi, Yan Zhongwei, and Shao Yuehong, 2012: Probabilistic precipitation forecasting based on ensemble output using generalized additive models and Bayesian model averaging. *Acta Meteor. Sinica*, **26**(1), 1–12, doi: 10.1007/s13351-012-0101-8.

## 1. Introduction

Numerical weather models have achieved great success in weather forecasting and been widely used as a fundamental tool for investigating and simulating synoptic processes. However, it is now more and more difficult to increase the effective forecast lead time by simply improving the physical mechanisms or raising resolutions in their numerical implementations. There are several sources of uncertainty in the model output that are almost impossible to eliminate. The most important two sources are the inherent chaotic behavior of the atmospheric system and the initial conditions subject to observational errors and/or data assimilation. Numerical weather forecasts without uncertainties specified are hard to be incorporated into operations and decision making of other downstream applications such as early warning of floods and rainfall-induced geological hazards. The ensemble forecasting technique provides an effective way to quantify those uncertainties. While ensemble forecasts from a single deterministic model running several times with different initial conditions only address some of the uncertainties inherent in the model, ensembles from multiple models, or a grand ensemble, can help address uncertainties arising from model physics, numerical implementation, and data assimilation. With additional information about uncertainties provided, the useful-

Supported by the National Basic Research and Development (973) Program of China (2010CB428402) and China Meteorological Administration Special Public Welfare Research Fund (GYHY200706001).

*Corresponding author: chi@bnu.edu.cn.

ness of forecasts particularly those at a lead time beyond 3 days can be greatly improved. For an overview of ensemble forecasting, see Zhu (2005).

The uncertainty of ensemble forecasting can be expressed in terms of forecast probability density function (PDF), based upon which probabilistic forecast is generated. To make full use of all the information available in an ensemble forecast, Bayesian model averaging (BMA) was introduced by Raftery et al. (2005) as a statistical post-processing method for producing probabilistic forecast from an ensemble. By BMA, the overall forecast PDF of any variable of interest is a weighted average of forecast PDFs based on each of the individual forecasts, where the weights are estimated by posterior probabilities of the models generating the forecasts and reflect the relative forecasting skills of the individual models in the training period. The BMA forecast PDF is not only calibrated but also sharp, meaning that the distribution intervals are narrower on average than those obtained from climatology. The BMA forecast variance consists of two parts: between-model and within-model variances, while the ensemble spread reflects only the first part. This feature can have a substantial impact on the forecast PDF, especially in its tail where extreme values are captured. Tail behavior of a forecast PDF might be vital for some applications like early warning of floods for which probabilities of extreme rainfall events must be sufficiently estimated.

Precipitation is one of the most important weather variables in all senses. Prior to applying BMA to an ensemble of precipitation forecasts, the distribution of precipitation must be modeled properly. However, for precipitation at daily or sub-daily time scale, this problem itself is still under active research. Difficulties arise from the natural features of precipitation: 1) non-negativity, 2) a point mass of probability at zero value, and 3) highly skewed distribution for positive values. Currently, among the various approaches to this issue, a two-stage approach proposed by Coe and Stern (1982) and Stern and Coe (1984) is widely adopted: precipitation occurrences and precipitation amounts conditional on occurrences are handled separately within the framework of generalized linear models (GLMs) (McCullagh and Nelder, 1989). Logistic regression (binomial distribution) and gamma distribution are fitted to precipitation occurrences and precipitation amounts, respectively. The mean of each distribution is linked to a linear combination of covariates with a function subject to some conditions. GLMs provide a flexible way to allow for space and time effects and external forcings (Chandler and Wheater, 2002; Yang et al., 2005). Following the two-stage strategy, Sloughter et al. (2007) developed a simple precipitation model (for forecast calibration and within-model variability) for BMA probabilistic forecasting. Ensemble forecast was the only predictor considered in the model. The full distribution of precipitation was a mixture of a point mass at zero and a gamma distribution.

For many applications such as hydrological forecasting using distributed hydrological models, gridded precipitation forecast over a basin is required. Therefore, spatial variation of precipitation as a function of spatial covariates should be explicitly formulated in the precipitation model. In addition, for the sake of operation, the model fitting and forecasting procedures should be able to be carried through without the need of any manual manipulation. Most important of all, extreme precipitation must be correctly estimated using the upper tail of the distribution given by the fitted precipitation model.

In this paper, a precipitation model based on the Tweedie distribution (Tweedie, 1984) within the framework of generalized additive models (GAMs) (Hastie and Tibshirani, 1990) is considered for BMA probabilistic forecasting. The Tweedie distribution is a compound Poisson-gamma distribution that can model occurrence and amount of precipitation simultaneously (Dunn, 2004). By assuming the Tweedie distribution, the precipitation model can be built straightforwardly with a single distribution. The GAMs are extensions of GLMs that allow for nonparametric regressions. Here, the thin plate regression splines (Wood, 2003) are used to represent the spatial variation of precipitation. By taking these measures, not only the BMA probabilistic forecasting can be made at any grid point, the performance of the

BMA probabilistic forecasting of extreme precipitation is improved, and the model building and forecasting procedures can be simplified to meet the requirements of operation as well.

In Section 2, we briefly review the techniques used in the BMA probabilistic precipitation forecasting. In Section 3, we give as example results of daily 24-h forecasts of 24-h accumulated precipitation over the Yishusi River basin in North China for July 2007 based on the ensemble forecast produced by the National Centers for Environmental Prediction (NCEP) for The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) (Park et al., 2008). Finally, in Section 4, we conclude and discuss possible improvements to the method described here and its potential application in the climate change impact study.

## 2. Methods

### 2.1 Tweedie distribution

Suppose that there are $N$ precipitation events that occur during a day and $N$ has a Poisson distribution with mean $\lambda$. If the $i$th precipitation event results in a precipitation amount of $R_i$, and each $R_i$ has a gamma distribution $G(-\alpha_i,\ \gamma_i)$ with mean $-\alpha_i\gamma_i$ and variance $-\alpha_i\gamma_i^2$, then the total daily precipitation $Y$ can be obtained by

$$Y = R_1 + R_2 + \cdots + R_N,$$

and it consequently has a Poisson-gamma distribution belonging to the Tweedie family of distributions. The resulting probability density function can be written as

$$\log f_p(y;\mu,\phi) = \begin{cases} -\lambda & \text{for } y=0, \\ -y/\gamma - \lambda - \log y & \\ \quad +\log W(y,\phi,p) & \text{for } y>0, \end{cases}$$

where $\gamma = \phi(p-1)\mu^{p-1}$ and $\lambda = \mu^{2-p}/[\phi(2-p)]$. $W$ can be expressed as the infinite summation

$$W(y,\phi,p) = \sum_{j=1}^{\infty} \frac{y^{-j\alpha}(p-1)^{\alpha j}}{\phi^{j(1-\alpha)}(2-p)^j j!\Gamma(-j\alpha)}, \qquad (1)$$

where $\alpha = (2-p)/(1-p)$. The mean of the distribution is $\mu$ and the variance is $\mathrm{Var}[Y] = \phi\mu^p$. For $1 < p < 2$, the distributions are supported on the nonnegative real numbers, and $p$ is the index that determines which Poisson-gamma distribution is used. Particularly, the probability of a dry day is given by

$$P(Y=0) = \exp(-\lambda) = \exp\Big[-\frac{\mu^{2-p}}{\phi(2-p)}\Big].$$

The Tweedie density function has no closed form except for some special cases. Numerical methods are needed to evaluate the infinite series in Eq. (1). Figure 1 shows the density functions of some Tweedie distributions for $1 < p < 2$. In addition to the continuous distribution for $Y > 0$, there is also a point mass of probability at $Y = 0$ for each Tweedie distribution.

### 2.2 Generalized additive model

GAMs are semi-parametric extensions of GLMs. Like a GLM, a GAM uses a link function to establish a relationship between the mean of the response variable $Y$ having a distribution from the exponential family and additive smooth functions of explanatory variables $x_1$, $x_2$,..., $x_p$, as follows:

$$g(\mu) = \beta_0 + s(x_1) + s(x_2, x_3) + \cdots + s(x_p), \qquad (2)$$

where $\mu = E[Y]$, $g(\cdot)$ is a smooth monotonic link function, $\beta_0$ is a constant, and various forms of $s(\cdot)$ are smooth functions of explanatory variables. It is possi-

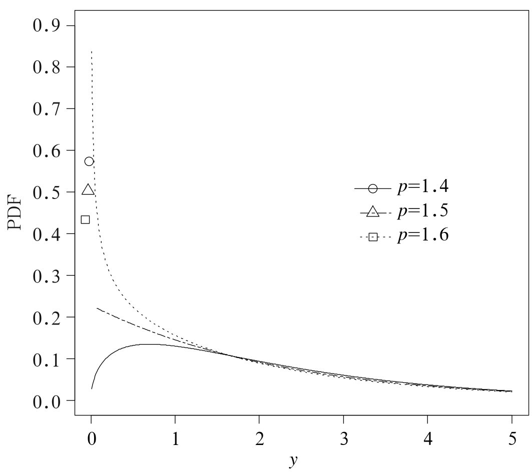


**Fig. 1.** Some Tweedie density functions for $1< p <2$. Geometric symbols indicate the point masses of probability at $Y= 0$. In each case, the mean and variance are fixed at 1 and 3, respectively.

ble to build this kind of model structures within the GLM framework using polynomials or orthogonal bases to represent nonlinearities, e.g., using Legendre polynomials to represent regional effects and Fourier series to represent seasonal cycles (Chandler and Wheater, 2002; Yang et al., 2005). However, such an approach has two main problems: model selection is rather cumbersome and the basis selected can have a strong influence on the fitted model. In the GAM framework, one efficient approach to implement the model is using penalized regression splines to represent the smooth terms (Wood and Augustin, 2002). The smooth functions are rewritten using a set of basis functions, each of which has an associated penalty that controls its effective degrees of freedom through a single smoothing parameter that can be determined by minimizing some objective functions. For a comprehensive introduction to this approach, refer to Wood (2006).

Usually, the spatial variation of precipitation over an area can be represented as a function of a two-dimensional vector variable $\boldsymbol{x}$ with its two components being longitude and latitude or projected $x$-$y$ coordinates. In GAM, the smooth function $s(\boldsymbol{x})$ can be implemented using thin plate regression splines (Wood, 2003). Seasonal cycles and longer-term time effects can be represented by one-dimensional smooth functions. Time consistencies of precipitation can be represented by introducing some auto-regression terms. In addition, large-scale external forcings on local precipitation can also be introduced linearly or nonlinearly. The Tweedie distribution itself is a member of the exponential family. Therefore, it is possible to build comprehensive spatial-temporal precipitation models by fitting a single distribution to a set of multi-site observations for a certain period in the GAM framework for various purposes. Here, GAMs are used to forecast precipitation conditional on an ensemble member individually.

### *2.3 BMA*

For ensemble probabilistic forecasting, BMA is a statistical technique that combines inferences and predictions based on individual ensemble members, so as to yield a more skillful and reliable probabilistic prediction. Assume that the forecast PDF of a weather variable $y$ conditional on the ensemble member forecast $f_k$ is $p_k(y|f_k)$, the BMA forecast PDF is then

$$p(y|f_1, f_2, \ldots, f_K) = \sum_{k=1}^{K} w_k p_k(y|f_k), \tag{3}$$

where $w_k$ is the posterior probability of forecast $k$, being the best one given the observations in the training period and subject to $\sum\limits_{k=1}^{K} w_k = 1$. The BMA ensemble forecast is essentially an average of forecasts based on individual members weighted by the likelihood that an individual forecasting model is correct given the observations. Denote $\mu_k = E[y|f_k]$ and $\sigma_k^2 = \mathrm{Var}[y|f_k]$, then the posterior mean and variance of BMA forecast can be expressed as

$$\begin{aligned}
\mu &= E[y|f_1, f_2, \ldots, f_K] = \sum_{k=1}^{K} w_k \mu_k, \\
\sigma^2 &= \mathrm{Var}[y|f_1, f_2, \ldots, f_K] \\
&= \sum_{k=1}^{K} w_k(\mu_k - \mu)^2 + \sum_{k=1}^{K} w_k \sigma_k^2.
\end{aligned}$$

That is, the BMA variance is a measure of BMA forecast uncertainty contributed by between-model uncertainty and within-model uncertainty.

Equation (3) is a finite mixture density (McLachlan and Peel, 2000) with a general form of

$$f(\boldsymbol{x}; \boldsymbol{w}, \boldsymbol{\theta}) = \sum_{k=1}^{K} w_k g_k(\boldsymbol{x}, \boldsymbol{\theta}_k),$$

where $\boldsymbol{x}$ is a random vector variable, $\boldsymbol{w} = (w_1, w_2, \ldots, w_K)$ is mixing proportions, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K)$ represents parameters of component densities $g_k$. Parameters $\boldsymbol{w}$ and $\boldsymbol{\theta}$ can be estimated by maximizing the log-likelihood function of the mixture density $L$, which is

$$L(\boldsymbol{w}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log f(\boldsymbol{x}_i; \boldsymbol{w}, \boldsymbol{\theta}),$$

where $N$ is the number of observations. Usually this cannot be maximized analytically, and instead can be maximized numerically using the expectation-maximization (EM) algorithm (McLachlan and Krishnan, 2008). The EM algorithm is iterative. For the

BMA model specified by Eq. (3), in brief, first introduce unobserved variables $z_{kst}$, where $z_{kst} = 1$ if the model $k$ produces the best forecast for precipitation observed at site $s$ and time $t$, and $z_{kst} = 0$ otherwise, and then proceed through the expectation (E) step

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

and the maximization (M) step

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

iteratively, where $j$ refers to the $j$th iteration and $f_{kst}$ is the forecast for site $s$ and time $t$ by the model $k$. Parameters $\boldsymbol{\theta}$ are also estimated at each iteration step using the current estimates of $\boldsymbol{w}$. The E step and M step are iterated till convergence, and the log-likelihood is guaranteed to increase at each step. However, for non-Gaussian component densities, convergence to a global maximization of the likelihood cannot be guaranteed. Starting values for the iteration should be carefully chosen in order to obtain a good result.

### *2.4 Model diagnosis*

Residual diagnostics is an effective way to assess fitted models. Here Pearson residuals are used to check the systematic structure captured by a model. The Pearson residual for the $i$th observation $y_i$ is defined as

$$r_i = \frac{y_i - \mu_i}{\sigma_i},$$

where $\mu_i$ and $\sigma_i$ are the mean and standard deviation of $y_i$ under the fitted model. If the model is correct, all values of $r_i$ have a distribution with mean zero and variance 1. By appropriate selection of subsets, the residuals can be used to check unexplained structures. For a large subset, the mean residual is defined as

$$\overline{r} = \frac{1}{N} \sum_{s=1}^{S} \sum_{t=1}^{T} \chi_{st} r_{st},$$

where $T$ is the number of times for observation in the subset and $S$ the number of sites; $\chi_{st}$ is an indicator taking the value 1 if a residual is available for site $s$ on time $t$ and 0 otherwise; $r_{st}$ is the residual at site $s$ on time $t$; and $N$ is the total number of observations in the subset, with $N = \sum\limits_{s=1}^{S} \sum\limits_{t=1}^{T} \chi_{st}$. Then 100(1–$\alpha$)% limits for this mean are approximately at $\pm Z_{\alpha/2}$s.e.$(\overline{r})$, where $Z_{\alpha/2}$ is the standard normal upper quantile at $\alpha/2$ and s.e.$(\overline{r})$ is the standard error of the mean residual. For a spatial-temporal precipitation model, by introducing appropriate covariates representing auto-regressions, residuals from the fitted model can be thought of as temporally independent. However, without the spatial dependence considered, they may still be dependent spatially unless all the sites are far enough from each other. The variance for the mean of residuals independent temporally but dependent spatially can be estimated by

$$\mathrm{Var}(\overline{r}) = \frac{1}{N^2} \sum_{t=1}^{T} \sum_{s_1=1}^{S} \sum_{s_2=1}^{S} \chi_{s_1 t} \chi_{s_2 t} \mathrm{cov}(r_{s_1 t}, r_{s_2 t}),$$

where $\mathrm{cov}(r_{s_1 t}, r_{s_2 t})$ is the covariance between residuals at sites $s_1$ and $s_2$ at the same time. The square root of the variance is the required standard error of this mean residual.

The mean residuals can be computed over subsets by site or by time. If the fitted model is correct, the mean residuals should not show significant trends and about 100(1–$\alpha$)% of them should be within the corresponding limits. Otherwise, there might be some unexplained structures that should be taken into account by the model.

### *2.5 Forecast verification*

Three widely used scoring rules, the mean absolute error (MAE), the continuous ranked probability score (CRPS), and the Brier score (BS) (Jolliffe and Stephenson, 2003), are considered here to verify the BMA probabilistic forecasts. The MAE is defined as

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|,$$

where $y_i$ is the verification, $\hat{y}_i$ is the deterministic forecast given by the median of the forecast PDF, and $N$ is the number of test data.

The CRPS is defined as

$$\mathrm{CRPS} = \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{\infty} [F(t) - H(t - y_i)]^2 \mathrm{d}t,$$

where $F(t)$ is the forecast cumulative distribution function (CDF), and $H(t-y)$ denotes the Heaviside function and takes 0 when $t<y$ and 1 otherwise. The CRPS can be viewed as a generalization of the MAE and can be directly compared to the latter (Gneiting and Raftery, 2007).

The BS for probability forecasts of $K$ complete and mutually exclusive categories is given by

$$\mathrm{BS}=\frac{1}{N}\sum_{k=1}^{K}n_k(\hat{p}_k-o_k)^2,$$

where $\hat{p}_k$ is the forecasted probability of the $k$th category subject to $\sum\limits_{k=1}^{K}\hat{p}_k=1$; $o_k=1$ if the event occurs and 0 otherwise such that there are $K-1$ zeros and a single 1 contained in $o_k$, $k=1, \ldots, K$; $n_k$ is the number of observations for the $k$th category and $n_k\times K=N$. The BS is essentially the mean squared error of the probability forecasts.

All these three scoring rules are negatively oriented, i.e., the smaller the values, the better the forecasts. The BS takes on values only in the range [0, 1].

### *2.6 Computational implementation*

Although the methods described above involve complicated computations, in the R environment (R Development Core Team, 2011), these can be easily implemented without heavy coding. The GAMs are fitted using the package "mgcv" (Wood, 2006). The Tweedie family of distributions is computed using the package "tweedie" (Dunn, 2009). Functions for numerical integration and optimization are also available in R.

In the fitting of each component of the BMA forecast PDF in Eq. (3) using the Tweedie family GAMs, the power $p$ is the parameter that determines which Tweedie distribution is used and should be specified prior to the fitting. The mean $\mu$ is expressed by smooth functions of covariates and the dispersion $\phi$ is assumed to be constant. These two parameters are member-specific and can be estimated by the individual fitting. The power $p$ can be assumed to be a constant for all the component PDFs in Eq. (3) so that except for weights $\boldsymbol{w}$ there is only one additional parameter $p$ in the mixture density function. Parameters $p$ and $\boldsymbol{w}$ can be estimated by maximizing the log-likelihood of the mixture density for the training data using the EM algorithm described in Subsection 2.3. The parameter $p$ is estimated numerically by optimizing the log-likelihood function using the current estimations of $\boldsymbol{w}$ at each M step. However, when practiced in the example given in the next section, this procedure showed so slow a convergence rate that it is actually impractical for operations. An alternative to this is searching $p$ for the maximum log-likelihood of each component PDF in Eq. (3) individually so that all the three parameters of the Tweedie family of distributions are member-specific and only $\boldsymbol{w}$ are estimated by the EM iteration. This procedure was much faster than the former one while the same maximum log-likelihood was reached. The estimated $w_k$ differed only slightly from the former results. The example BMA model described in the next section was finally fitted using this procedure.

## 3. Example results

### *3.1 Data description and model fitting*

The BMA probabilistic precipitation forecasting was applied to daily 24-h forecasts of 24-h accumulated precipitation over the Yishusi River basin in North China throughout July 2007. The Yishusi River consists of three main branches, namely, Yi River, Shu River, and Si River. It starts from Yimeng Mountain in Shandong Province and runs into the Yellow Sea crossing Henan, Anhui, and Jiangsu provinces. The total area of the basin is about $7.96\times10^4$ km$^2$. Example stations in the basin are shown in Fig. 2, 40 of which are training stations (triangles) and the other 4 test stations (dots). Precipitation data from these stations were provided by the China Meteorological Administration. There were several rainstorms that occurred over the basin in that month; the highest recorded 24-h accumulated precipitation was 259.6 mm at Shouxian station on the 8th. Contours in Fig. 2 show the averaged 24-h precipitation over the month. The 21-member $1°\times1°$ NCEP ensemble precipitation forecasts retrieved from TIGGE were interpolated at these stations as ensemble forecasts corresponding to

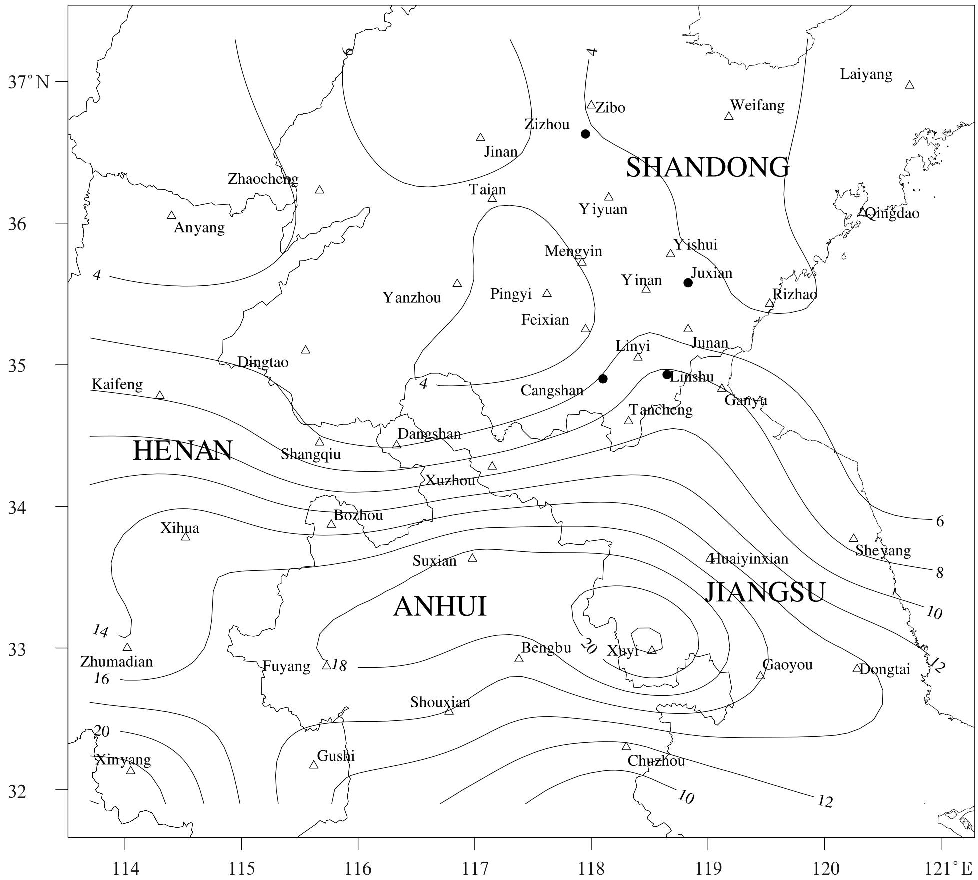


**Fig. 2.** Example stations in the Yishusi River basin. Triangles represent training stations and dots represent test stations. Contours show the averaged 24-h precipitation (mm) over July 2007.

the observations.

Sloughter et al. (2007) found out the optimized length of training period by minimizing the MAE or CRPS. They concluded through their examples that 30 days was enough and that longer period would not significantly reduce the scores. For the study area we chose, these scores are sensitive to the length of the training period, because over the East Asian monsoon region where the river basin is located, the seasonal variability of precipitation is very significant. Nevertheless, 30 days is still a proper length of time period during which the precipitation is approximately homogeneous for the study area. If the training period is long and the precipitation is inhomogeneous temporally, the model should include explicitly some covariates representing seasonality or even longer-time effects. Since here we intend to give an illustrative example rather than a complicated "perfect" solution, we take 30 days as the length of the training period during which we assume that the precipitation is homogeneous in time.

Exploratory analysis shows that the observational series at each station has high correlations with corresponding forecast series at lags from 0 to 2 days, with the lagged correlations even higher than its autocorrelations. By taking all these considerations together, the precipitation forecast model using GAMs based on the $k$th member of the ensemble forecast is constructed as

$$\log\mu_i^{(k)}(t) = s(\boldsymbol{x}_i) + s(f_i^{(k)}(t)) + s(f_i^{(k)}(t-1)) + s(f_i^{(k)}(t-2)), \qquad (4)$$

where $\mu_i^{(k)}(t)$ is the mean of the Tweedie distribution of precipitation at the $i$th station on day $t$ conditional on the $k$th member of the ensemble forecast, $\boldsymbol{x}_i$ is the

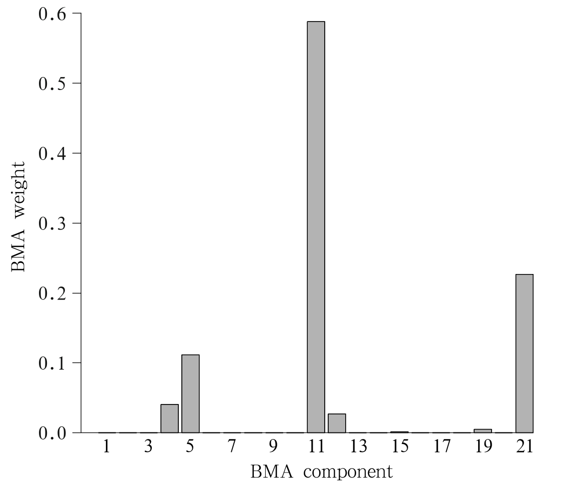


**Fig. 3.** Estimated weights for the 21 BMA components fitted to the period of 1–30 July 2007. The first 20 are for components conditional on perturbed runs of the ensemble forecast and the last one on the control run.

two-dimensional coordinates of the $i$th station, $f_i^{(k)}(t)$ is the $k$th member of the ensemble forecast for the $i$th station on day $t$, and $s$ represents one- or two-dimensional smoothing regression splines. To achieve a better fit, the cubic roots of the ensemble forecasts rather than the original values are taken in the model fitting, just as treated in Sloughter et al. (2007) and many other works.

After all the individual models are fitted, the resulted forecast PDFs are then properly weighted and summed up to produce the final BMA forecast PDF. Figure 3 shows the weights estimated using the EM algorithm for the 21 components fitted to the period of 1–30 July 2007. It can be seen that except for 5 components, all the other weights are almost cut down to 0. Figure 4 shows the mean Pearson residuals (circles) of the fitted BMA forecast model by site (Fig. 4a) and by day (Fig. 4b). Lines show the 95% confidence limits. There is an apparent outlier from Dingtao station on the 11th. The recorded precipitation for that day was 54.6 mm; however, not only all the ensemble members forecasted an amount of 0 for that day, but both the ensemble forecast and recordings for the previous 3 days were all 0s as well. This case exceptionally violated the relationship assumed in the model specified by Eq. (4) and resulted in the Pearson residual for that day as high as 22.3, and this is reflected on the mean residual plots. Except for this case, and for a little bit overestimation for a few stations, the systematic spatial and temporal patterns of precipitation were well captured by the BMA forecast model.

### *3.2 BMA forecasts*

The dynamic BMA forecasting for the 4 test stations proceeded through July 2007 using a sliding training period of 30 days. Empirical PDFs of the raw ensemble forecasts and the sample climatology were also used respectively to generate forecasts for comparison. Table 1 shows the MAE, CRPS, and

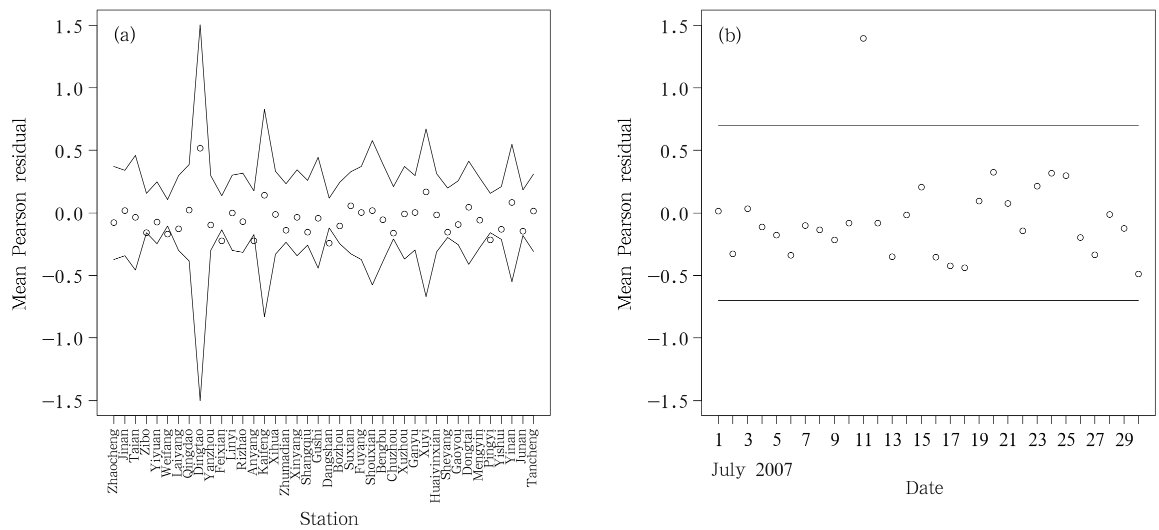


**Fig. 4.** Mean Pearson residuals (circles) of the fitted BMA forecast model by site (a) and by day (b). Lines show the 95% confidence limits, which are adjusted for spatial dependence between sites.

**Table 1.** MAE, CRPS, and BS for probabilistic precipitation forecasts for the four test stations and their averages

| Station | | MAE (mm) | CRPS | BS | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 0 mm | 10 mm | 25 mm | 50 mm | 100 mm | 250 mm | Average |
| Zichuan | B | 3.07 | 2.31 | 0.145 | 0.117 | 0.0696 | 0.00514 | 0.0216 | 0.000313 | 0.0599 |
| | E | 3.74 | 3.36 | 0.217 | 0.224 | 0.132 | 0.00358 | 0.0323 | 0 | 0.102 |
| | C | 4.19 | 3.94 | 0.219 | 0.142 | 0.0857 | 0.00597 | 0.0312 | 0.000162 | 0.0807 |
| Juxian | B | 3.53 | 2.50 | 0.175 | 0.203 | 0.0143 | 0.0872 | 0.00228 | 0.000115 | 0.0803 |
| | E | 3.15 | 2.81 | 0.245 | 0.255 | 0.0756 | 0.120 | 0 | 0 | 0.116 |
| | C | 4.84 | 4.00 | 0.250 | 0.248 | 0.00638 | 0.113 | 0.00139 | 0.000162 | 0.103 |
| Cangshan | B | 5.28 | 4.35 | 0.199 | 0.238 | 0.0689 | 0.0463 | 0.0280 | 0.00211 | 0.0969 |
| | E | 5.45 | 4.63 | 0.326 | 0.381 | 0.137 | 0.0676 | 0.0323 | 0 | 0.157 |
| | C | 6.13 | 5.29 | 0.268 | 0.273 | 0.0618 | 0.0631 | 0.0311 | 0.000162 | 0.116 |
| Linshu | B | 7.09 | 5.74 | 0.258 | 0.179 | 0.0360 | 0.0687 | 0.0598 | 0.00191 | 0.101 |
| | E | 8.18 | 7.40 | 0.487 | 0.501 | 0.110 | 0.0936 | 0.0645 | 0 | 0.209 |
| | C | 8.54 | 7.23 | 0.237 | 0.167 | 0.0357 | 0.0881 | 0.0613 | 0.000162 | 0.0983 |
| Average | B | 4.74 | 3.72 | 0.194 | 0.184 | 0.0472 | 0.0518 | 0.0279 | 0.00111 | 0.0844 |
| | E | 5.13 | 4.55 | 0.319 | 0.340 | 0.114 | 0.0713 | 0.0323 | 0 | 0.146 |
| | C | 5.93 | 5.12 | 0.244 | 0.207 | 0.0474 | 0.0675 | 0.0312 | 0.000162 | 0.0996 |

The letter B stands for BMA forecasts, E for empirical probabilistic ensemble forecasts, and C for sample climatology.

BS of BMA forecasts, empirical probabilistic ensemble forecasts, and sample climatology for each station and their averages over the four stations. It can be seen that the BMA forecasts outperformed the other two in almost every category under comparison.

To early warnings of floods, what is more important is the forecast for extreme precipitation that can result in floods. Among the 120 records from the 4 test stations, the top 10% of them can be regarded as extreme events for the period of interest, and 4 of which are greater than 50 mm. Figure 5 shows the BMA 50th (solid lines) and 90th (dashed lines) percentile forecasts compared with observations (circles) for the 4 test stations through July 2007. Figure 6 is the counterpart of Fig. 5 for empirical probabilistic ensemble forecasts. If we take the medians as the deterministic forecasts, although the BMA deterministic forecasts for heavy precipitation are smaller than the observations in general, almost all the BMA 90th percentile upper bounds exceed the observations (there is only one apparent exception). In contrast, almost all the corresponding upper bounds forecasted by the empirical ensemble PDFs are below the heavy rainfall observations. Have in mind that except for spatial coordinates and ensemble forecasts, no other information on the test stations was considered in the BMA forecasting, the BMA results are really impressive. This comparison also indicates that the BMA forecasts are much better calibrated than the empirical probabilistic ensemble forecasts.

## 4. Discussion

In this paper, we proposed a BMA probabilistic precipitation forecasting model. GAMs were used to fit the spatial-temporal precipitation models to individual ensemble member forecasts. The distributions of the precipitation occurrence and the cumulative precipitation amount were represented simultaneously by a single Tweedie distribution. BMA was then used to mix up the GAMs and to yield the final mixture model. The mixing weights were estimated using the EM algorithm maximizing the log-likelihood of the mixture model. The residual diagnostics was used to examine if the fitted BMA forecasting model had fully captured the spatial and temporal patterns of precipitation for the study area and the period of interest. The whole procedure was applied to an example dataset from the Yishusi River basin for July 2007 using the NCEP ensemble forecasts to generate the BMA probabilistic forecasts. The results were verified by applying three scoring rules and it is concluded that the BMA

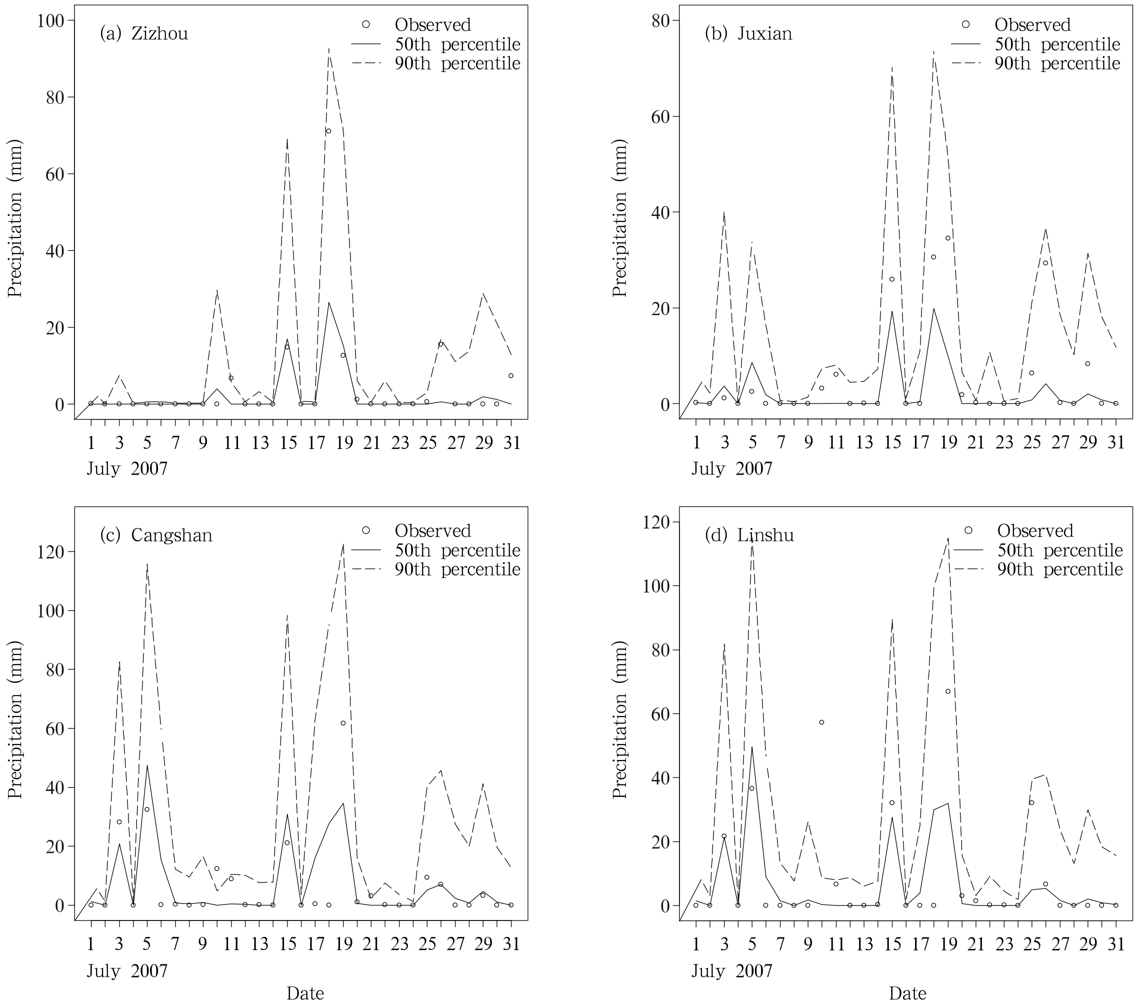


**Fig. 5.** The BMA 50th (solid lines) and 90th (dashed lines) percentile forecasts of 24-h precipitation and the observations (circles) for the 4 test stations through July 2007.

forecasts outperformed the empirical probabilistic ensemble forecasts particularly for extreme precipitation events.

The reason why the BMA probabilistic forecasts captured the extreme precipitation so well as shown in the example is that, within the GAM framework, the nonlinear relationship between the mean of the precipitation distribution and the covariates can be represented "as is" by the model described by Eq. (2). For highly skewed data like precipitation records, their variance is usually proportional to their mean, just like samples from the gamma distribution. Once the means for extreme precipitation events are underestimated, their variances are also underestimated. This is usually the case in GLMs in which the right-hand side of Eq. (2) is instead a linear combination of covariates that can be biased easily by the large amount of small-value events. Consequently, the forecasted upper bounds for extreme events are also underestimated. In GAMs, the additive smooth functions in Eq. (2) can represent unbiased relationships between the mean and the covariates for both the small-value and extreme precipitation events, which guarantees sufficient variances for extreme precipitation events and large enough forecasted upper bounds as well.

In the fitting of spatial-temporal precipitation model using GAMs, spatial independence between sites conditional on the fitted surface was assumed.

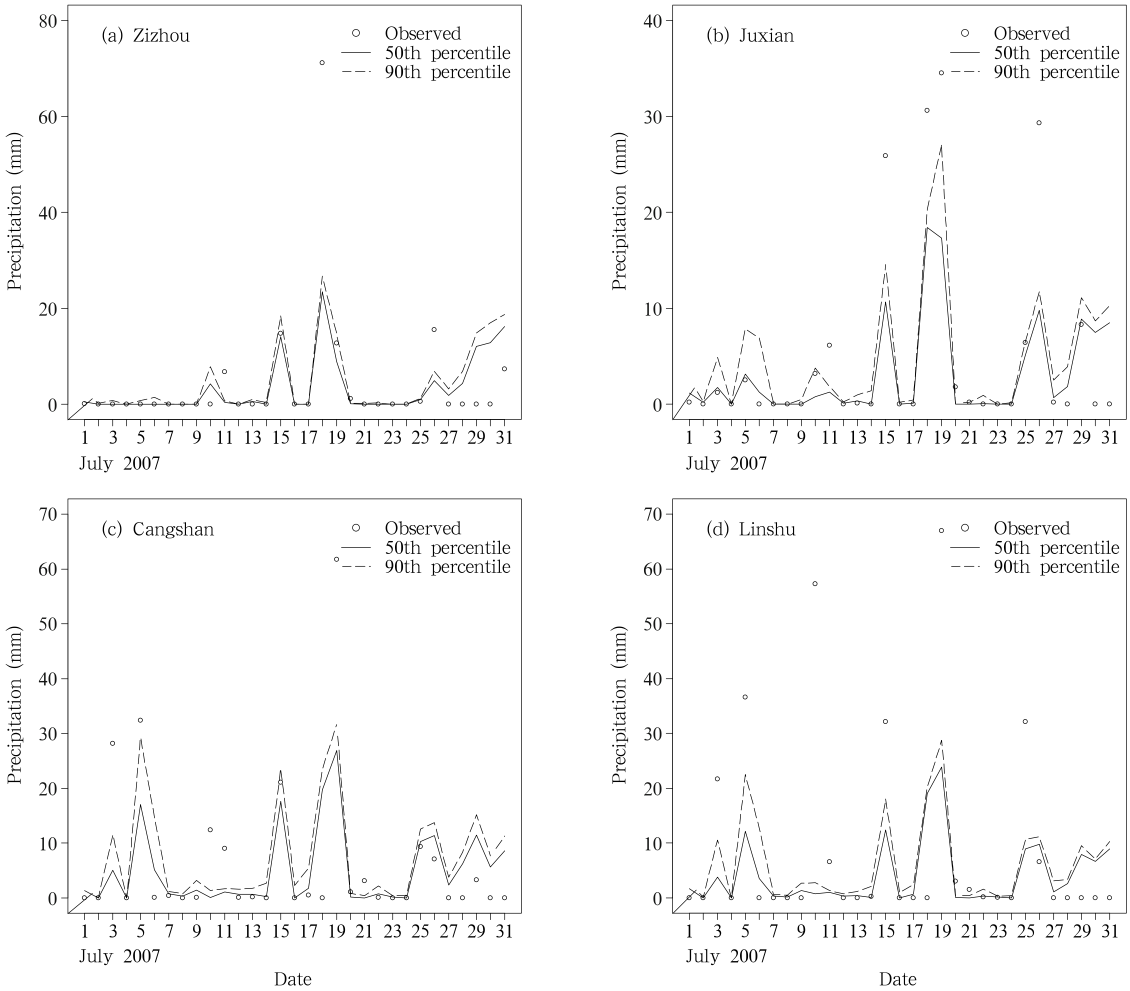


**Fig. 6.** The counterpart of Fig. 5 for the empirical probabilistic ensemble forecasts.

Only in the model diagnosis, adjustment was made for spatial dependence for the estimation of confidence limits of the mean Pearson residuals. This is not a problem in using the model for forecasting, since the systematic spatial variation has already been incorporated into the model and what is needed is only the marginal forecast distribution for each station. However, when the model is used for multi-site simulation of precipitation, it is necessary to allow for spatial dependence in the model to generate realistic scenarios. One approach to implement this is the generalized additive mixed-effect models (GAMMs) (e.g., Ruppert et al., 2003), in which spatial correlation structures can be specified. However, the incorporation of spatial correlation structures makes the computation of GAMMs intensive and much slower than that of GAMs. The whole BMA post-processing procedure to incorporate spatial dependence suitable for operations needs to be worked out ad hoc.

At the stage of the numerical estimation of BMA weights, weights for extreme precipitation events are still apt to be underestimated. A possible improvement to this is to use weighted likelihood estimator in the EM algorithm (Markatou, 2000). The weight function must be carefully designed such that the log-likelihoods of models in good agreement with observations of heavy precipitation can receive higher weights. This will be considered in our future work.

A direct application of the method proposed here is the downscaling of multimodel GCM simulations for climate change scenarios with uncertainty estimation. Since for this purpose the training period could be decades of years, time effects at different scales such as seasonal, interannual, and decadal variations should be explicitly represented in the model. Covariates other than spatial and temporal variables could be GCM outputs representing large-scale circulation patterns that result in the local precipitation. If spatial dependence is incorporated, the model could be used to simulate the local precipitation under climate change scenarios. A series of fine-scale precipitation fields could be generated by such simulation models as input to hydrological models for the assessment of climate change impact.

## REFERENCES


Chandler, R. E., and H. S. Wheater, 2002: Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resource Res.*, **38**(10), 1192–1202.

Coe, R., and R. D. Stern, 1982: Fitting models to daily rainfall data. *J. Appl. Meteor.*, **21**, 1024–1031.

Dunn, P. K., 2004: Occurrence and quantity of precipitation can be modeled simultaneously. *Int. J. Climatol.*, **24**, 1231–1239.

——, 2009: *Tweedie: Tweedie Exponential Family Models.* R package version 2.0.2, http://CRAN.R-project.org/package=tweedie.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, **102**, 359–378.

Hastie, T. J., and R. J. Tibshirani, 1990: *Generalized Additive Models.* Chapman and Hall, London, 335 pp.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Wiley and Sons, 240 pp.

Markatou, M., 2000: Mixture models, robustness, and the weighted likelihood methodology. *Biometrics*, **56**(2), 483–486.

McLachlan, G. J., and D. Peel, 2000: *Finite Mixture Models.* Wiley, 419 pp.

——, and T. Krishnan, 2008: *The EM Algorithm and Extensions.* 2nd Edition. Wiley, 359 pp.

McCullagh, P., and J. A. Nelder, 1989: *Generalized Linear Models.* 2nd Edition, Chapman and Hall, London, 532 pp.

Park, Y.-Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.*, **134**, 2029–2050.

R Development Core Team, 2011: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raftery, A. E., T. Gneiting, F. Balabdaoui, et al., 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.

Ruppert, D., M. P. Wand, and R. J. Carroll, 2003: *Semiparametric Regression.* Cambridge University Press, 386 pp.

Sloughter, J. M., A. E. Raftery, T. Gneiting, et al., 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220.

Stern, R. D., and R. Coe, 1984: A model fitting analysis of daily rainfall data. *J. Roy. Stat. Soc.*, **147A**, 1–34.

Tweedie, M. C. K., 1984: An index which distinguishes between some important exponential families in statistics: Applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, J. K. Ghosh and J. Roy, Eds., Indian Statistical Institute, Calcutta, 579–604.

Wood, S. N., 2003: Thin plate regression splines. *J. R. Statist. Soc. B*, **65**(1), 95–114.

——, 2006: *Generalized Additive Models: An Introduction with R.* CRC/Chapman & Hall, Boca Raton, Florida, 392 pp.

——, and N. H. Augustin, 2002: GAMs with integrated model selection using penalized regression splines and applications to environmental modeling. *Eco. Modelling*, **157**, 157–177.

Yang, C., R. E. Chandler, V. S. Isham, et al., 2005: Spatial-temporal rainfall simulation using generalized linear models. *Water Resource Res.*, **41**(11), W11415.

Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22**(6), 781–788.