Advanced Review

# A review on statistical postprocessing methods for hydrometeorological ensemble forecasting

Wentao Li,[1,2] Qingyun Duan,[1,2*] Chiyuan Miao,[1,2] Aizhong Ye,[1,2] Wei Gong[1,2] and Zhenhua Di[1,2]

Computer simulation models have been widely used to generate hydrometeorological forecasts. As the raw forecasts contain uncertainties arising from various sources, including model inputs and outputs, model initial and boundary conditions, model structure, and model parameters, it is necessary to apply statistical postprocessing methods to quantify and reduce those uncertainties. Different postprocessing methods have been developed for meteorological forecasts (e.g., precipitation) and for hydrological forecasts (e.g., streamflow) due to their different statistical properties. In this paper, we conduct a comprehensive review of the commonly used statistical postprocessing methods for both meteorological and hydrological forecasts. Moreover, methods to generate ensemble members that maintain the observed spatiotemporal and intervariable dependency are reviewed. Finally, some perspectives on the further development of statistical postprocessing methods for hydrometeorological ensemble forecasting are provided. 

**How to cite this article:**
*WIREs Water* 2017, e1246. doi: 10.1002/wat2.1246

## INTRODUCTION

Computer-based models have been an indispensable tool for hydrometeorological forecasting. However, all models are only abstract representations of the reality, and model-generated forecasts are plagued by uncertainties from various sources, including model inputs and outputs, initial and boundary conditions, and model structure and parameters.[1] Traditional hydrometeorological forecasts are generated in a deterministic manner, i.e., the forecasts of a certain hydrometeorological event are provided in the form of a single space–time series. This type of forecasts is inherently incapable of accounting for forecast uncertainty. To assess forecast uncertainty, the ensemble forecasting approach has gained popularity.[2,3] Ensemble forecasts are generated by running the model (or models) several times with perturbed factors such as model initial condition, model forcing, or model physics. This type of forecasts provides not only the most likely space–time scenario for a given event but also associated quantitative uncertainty information. Studies have shown that ensemble forecasts can improve the forecast accuracy and extend the forecast lead times over deterministic forecasts.[2,3]

Whether they are deterministic or ensemble forecasts, the raw forecasts generated by any model cannot be used directly by the end users because of various biases in them, which must be treated by statistical postprocessors, including 'meteorological postprocessor' and 'hydrological postprocessor'.[1] The

*Correspondence to: qyduan@bnu.edu.cn

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China

[2]Institute of Land Surface System and Sustainable Development, Faculty of Geographical Science, Beijing Normal University, Beijing, China

Conflict of interest: The authors have declared no conflicts of interest for this article.

 

'meteorological postprocessor' deals with the uncertainty in meteorological forecasts, which are inputs to the hydrological processor. The 'hydrological postprocessor' deals with the hydrological outputs from a hydrological processor. As shown in Figure 1, statistical postprocessors (both meteorological postprocessor and hydrological postprocessor) compose the hydrological ensemble forecasting system (HEFS) along with three other parts, namely, the data assimilator for uncertainty in initial conditions, parametric uncertainty processor for uncertainty in model parameters, and the hydrological processor.[4–7]

Statistical postprocessors are basically statistical models that relate observed variables of interest with the appropriate predictors derived from the direct model outputs (DMOs) of a meteorological or a hydrological model. Figure 2 illustrates how a statistical postprocessor works. First, the raw forecasts of a specific event and the corresponding observations collected in the training period are fed into a statistical model to derive the joint probability distribution between the raw forecasts and the observations. The training period can be a rolling period of the recent past, or a time window around a specific date in multiple years if reforecast datasets are available.[8–10] After the postprocessors for all relevant events are built, ensemble members are generated from the calibrated conditional probability distributions of all events. A well-constructed postprocessor achieves the following purposes: (1) it corrects the biases and dispersion errors in raw forecasts; (2) it preserves the predictive skill of the raw forecasts; (3) it downscales raw forecasts to the scale of applications (e.g., basin scale); and (4) it generates ensemble members of interested variables, which preserve the spatiotemporal and intervariable statistical dependency structure.[1,11–13]

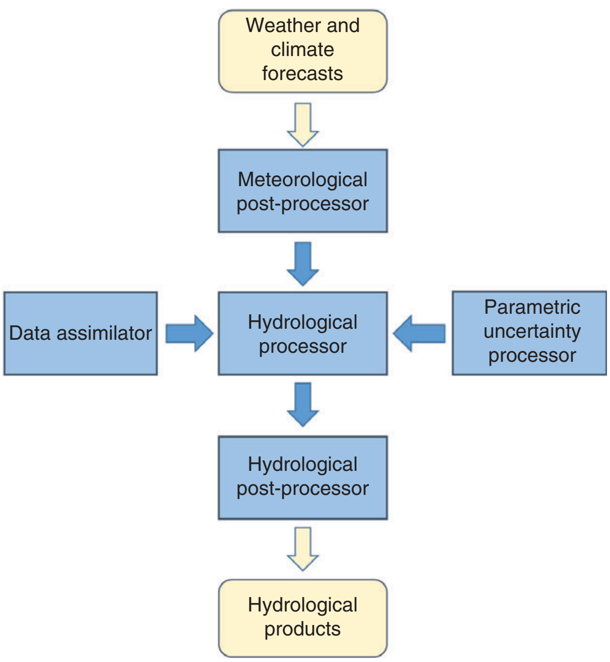


**FIGURE 1** | The components in a hydrological ensemble forecasting system (HEFS).

The importance of statistical postprocessing has long been recognized in meteorological forecasting.[14] Early works included models such as perfect prognosis,[15] model output statistics (MOS),[16] and the analog method (AM).[17–19] Over the recent years, many other postprocessing methods have been proposed, including rank histogram calibration,[20,21] quantile mapping (QM),[22,23] and ensemble preprocessor (EPP).[24,25] Several Bayes' theorem-based models were developed to combine prior climatology information with real-time forecasts, such as the Bayesian processor of output (BPO), Bayesian processor of forecast (BPF), and Bayesian processor of ensemble (BPE).[26–28] There exists a variety of regression-based models, including ensemble model output statistics (EMOS),[29–31] logistic regression (LR),[32–35] quantile regression (QR),[36,37] and member-by-member approach.[38] Ensemble dressing is another type of method for ensemble forecasts, which describes predictand with a mixture distribution.[39–41]

Postprocessing models for hydrological forecasts share similarities with those for meteorological forecasts, and regression or conditional distribution-based methods can also be applied for hydrological forecasts. One difference is that strong temporal autocorrelation exists in hydrological forecasts, so past recent observations or forecasts should be included as predictors in statistical postprocessing models. There are conditional distribution-based models for hydrological forecasts, such as hydrological uncertainty processor (HUP),[42–46] model conditional processor (MCP),[47–49] and Bayesian joint probability (BJP) model.[50–52] Regression-based models, such as general linear model postprocessor (GLMPP),[53–55] Kalman filter,[56,57] and autoregressive (AR) models[58–60] with wavelet transform methods,[61–63] are also proposed. Some nonparametric methods have been developed for hydrological forecasts to avoid parametric assumptions.[64,65]

Because all models are imperfect and no single model is expected to perform well under all circumstances, multimodel postprocessing methods have emerged as a way to achieve better forecast skills.[66] Methods such as the 'poor man's ensemble' method,[67–69] simple model average (SMA),[70,71] multimodel superensemble,[72–74] Bayesian model averaging (BMA),[75–77] and other heuristic methods[78,79]

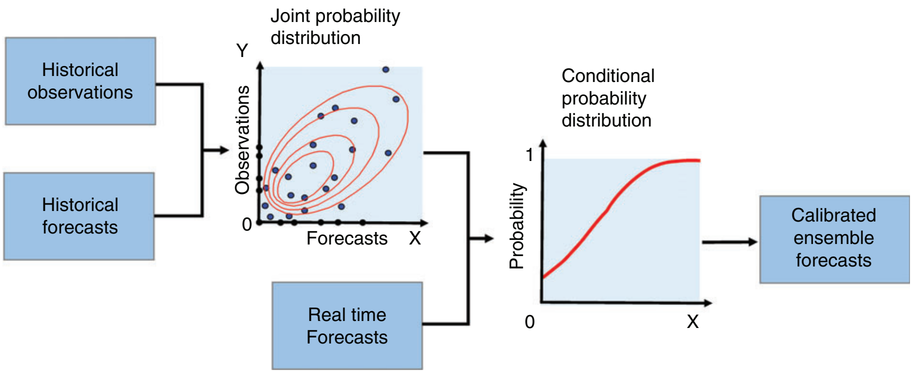


**FIGURE 2** | Statistical postprocessing flow for hydrometeorological ensemble forecasting.

are developed for both meteorological forecasts and hydrological forecasts.

Most statistical postprocessors are developed for a given hydrometeorological event at a specific location and a specific lead time. However, hydrological applications generally require that the forecast products be in the form of continuous space–time series that preserve the spatiotemporal and intervariable statistical dependency structure of the observations.[80] Several methods have been developed to generate ensemble members that meet those requirements, including parametric methods such as spatial EMOS[81,82] and spatial BMA[83,84] or nonparametric methods such as Schaake shuffle[80] and ensemble copula coupling (ECC).[85]

There are already several review papers on hydrological ensemble forecasting[2,13] or probabilistic weather forecasting.[12] Wilks provided a comprehensive review of postprocessing methods for meteorological forecasts in 2007.[86] This paper provides a review of the commonly used statistical postprocessing methods for hydrometeorological forecasts, especially newly developed methods in recent years. The postprocessing methods for hydrometeorological forecasts from single models are reviewed in the second section. In the third section, multimodel postprocessing methods are discussed. In the fourth section, postprocessing methods that preserve spatial, temporal, and intervariable dependency are examined. Finally, some perspectives on the further development of statistical postprocessing methods are presented in the Conclusion section.

## STATISTICAL POSTPROCESSING METHODS FOR HYDROMETEOROLOGICAL FORECASTS GENERATED BY A SINGLE MODEL

In this section, the statistical postprocessing methods for meteorological forecasts generated from a single numerical weather prediction (NWP) model, or hydrological forecasts from a single hydrological model, are reviewed. For meteorological forecasts such as surface air temperature and atmospheric pressure, the forecast errors can be represented by Gaussian distributions and can be corrected relatively easily by conventional regression methods. Postprocessing models for precipitation forecasts are more complicated because (1) the distribution of forecast or observation of precipitation is a mixed discrete/continuous distribution; (2) the forecast error is heteroscedastic; and (3) the extreme events are hard to represent because of limited samples.[31] The postprocessing model for hydrological forecasts, such as streamflow forecasts, have to deal with similar difficulties as those for precipitation forecasts. One unique feature is that due to the strong temporal autocorrelation of hydrological forecasts, past recent observations or forecasts should be included as predictors in the postprocessing models. This section focuses mostly on methods for postprocessing precipitation forecasts and hydrological forecasts. Some of the methods, such as QM and AM, are regarded as empirical methods, while others are typical statistical methods, including condition distribution-based methods, regression-based methods, and ensemble dressing methods.

### Early Ad-hoc Methods

There are several simple ad-hoc methods, such as DMO and rank histogram recalibration methods. DMO is used to estimate probability forecasts from raw ensemble forecasts directly and is not often viewed as a postprocessing method.[86] Here, we mainly introduce the rank histogram recalibration method, which is designed to calibrate forecasts based on the rank histogram of historical ensemble forecasts.[87] Firstly, the constant biases in raw forecasts are removed. Then, the rank histogram is constructed based on the debiased forecasts and observations in the training datasets. In the rank histogram, the frequency of past observations with rank $i$ is estimated as $w_i$. Finally, a probability forecast for an interested event can be obtained given new ensemble forecasts. Specifically, if the interested threshold value is bounded by the range of ensemble, the probability of an event less than the threshold can be estimated using the frequency of observations in the rank histogram as follows:[87]

$$P(y \le q) = \sum_{j=1}^{i} w_j + w_{i+1} \frac{q - \widetilde{x}_{(i)}}{\widetilde{x}_{(i+1)} - \widetilde{x}_{(i)}}, \; \widetilde{x}_{(i)} \le q \le \widetilde{x}_{(i+1)}, \tag{1}$$

where $y$ is the predictand, $q$ is the threshold, and $\widetilde{x}_{(i)}$ is the debiased ensemble forecast member in ascending order. When the interested threshold is outside the range of all ensemble members, the probability estimation will be less accurate as extrapolation needs to be performed. Wilks conducted an experiment using synthetic data and found that the rank histogram recalibration performs better than other ad-hoc methods but worse than other advanced methods, such as EMOS or ensemble dressing.[86]

## Quantile Mapping

Quantile mapping (QM, also called quantile-to-quantile transform or cumulative distribution function (CDF) matching) is a simple postprocessing method that adjusts the CDF of the forecasts according to the CDF of the observations. The forecast distribution is adjusted using the following formula[22,23]:

$$\widetilde{x}_{\text{adjusted}} = F_{\text{obs}}^{-1}(F_{\text{sim}}(x_{\text{sim}})), \tag{2}$$

where $x_{\text{sim}}$ is the raw forecast, $\widetilde{x}_{\text{adjusted}}$ is the adjusted forecast, $F_{\text{sim}}$ is the CDF of raw model simulation, and $F_{\text{obs}}^{-1}$ is the inverse of the observation CDF. In performing QM, each forecast value is 'mapped' to the corresponding quantile in the observation CDF.

QM has been applied to the postprocessing of both precipitation forecasts[23] and streamflow forecasts.[22,88] Zhu et al. applied a method similar to QM, called the frequency matching method, to postprocess precipitation forecasts from the Global Ensemble Forecast System (GEFS), which uses the frequency value of forecasts and observations instead of their CDF value during the matching process.[89] Verkade et al. also applied QM to adjust some unconditional bias in ECMWF precipitation and temperature forecasts.[90]

However, as an unconditional method, QM does not preserve the connection between each pair of forecast and observation values. Thus, QM might sometimes adjust the raw forecasts to the wrong direction for some forecast values and cannot provide satisfying results as conditional methods.[91] Moreover, Zhao et al. did an in-depth study on whether QM is suitable for postprocessing GCM precipitation forecasts.[92] They found that although QM is able to correct the bias, it cannot ensure the reliability and coherence of forecasts ('coherence' here means forecasts are at least as skillful as climatology).[92] The reason is that QM does not consider the correlation between raw forecasts and observations. Therefore, Zhao et al. concluded that QM was not a satisfactory method for postprocessing forecasts that suffer from not only bias but also reliability and coherence problems.[92]

## Analog Method

AM mainly searches reforecast datasets for past forecasts that are similar to the current forecasts and forms calibrated ensemble forecasts from the observations corresponding to the analog forecasts. The analogs can be established using distance measurements, such as the root mean square, or correlation measures.[19] Then, the observations in these similar states are applied to establish a distribution, termed the calibrated probabilistic forecast, by calculating the frequency of observations in the similar states.[19]

Specifically, let $y^{t|r} = (y^{t|r}(1), \ldots, y^{t|r}(s))$ denote the $s$ past observations on the dates of the analogs. Here, '$t$' is shorthand for 'truth' and '$r$' for 'reforecast'. Then, the probabilistic quantitative forecasts can be obtained using the following formula[19]:

$$P(y^t > T) = \frac{1}{s} \sum_{k=1}^{s} I\left(y^{t|r}(k), T\right), \tag{3}$$

which means $I(y^{t|r}(k), T) = 1$ when $y^{t|r}(k) > T$; otherwise, $I(y^{t|r}(k), T) = 0$. Compared to other statistical processing methods, AM is simple, but a large archive of observations and reforecasts is needed.[19] Besides being suited for the postprocessing of ensemble forecasts, AM can also be applied to generate ensemble forecasts from deterministic forecasts.[93]

## Conditional Distribution-based Methods

As the aim of a postprocessor is to estimate the conditional distribution of the observation $y$, given the model simulation $x$, we review two such methods for deriving the conditional distribution for meteorological forecasts:

the BPO and the EPP. Three other conditional distribution-based methods originally designed for hydrological forecasts, namely hydrological uncertainty processor (HUP), MCP and BJP, are also reviewed.

### *Bayesian Processor of Output*

The BPO model combines the climatic prior information and the forecast information from a NWP model based on Bayes' theorem. The posterior probability density function (PDF) of the predictand $y$, given the predictor vector $\mathbf{x}$, is as follows[26]:

$$h(y|\mathbf{x}) = \frac{f(\mathbf{x}|y)g(y)}{\kappa(\mathbf{x})}, \tag{4}$$

where $g(y)$ is the prior density distribution of the predictand, $f(\mathbf{x}|y)$ is the conditional density distribution of predictor vector $\mathbf{x}$ conditioned on the predictand $y$ (this part is also called likelihood function), and $\kappa(\mathbf{x})$ is the expected density function of the predictor vector $\mathbf{x}$. The prior density distribution of the predictand is estimated from historical observations, which represents the prior climatological information of the predictand. The likelihood function represents the relationship between observations and the corresponding forecasts. In this way, the conditional distribution of observations, given forecasts, can be obtained, which combines both the prior climatological information and the model forecasts information.

To model the distribution of the intermittent precipitation amount, the probabilistic quantitative precipitation forecast (PQPF) is constructed from two parts, namely, the probability of precipitation (PoP) occurrence and the distribution of amount (DoA), as follows[94,95]:

$$P(y|\mathbf{x}) = (1-\pi) + \pi \cdot H(y|\mathbf{x}, y>0), \tag{5}$$

where $\pi = P(y > 0|\mathbf{x})$ represents the conditional PoP occurrence given forecasts $\mathbf{x}$, and $H(y|\mathbf{x}, y > 0)$ is the posterior distribution of the precipitation amount $y$, conditional on forecasts $\mathbf{x}$ and occurrence of precipitation.

To deal with non-Gaussian predictand, such as precipitation, normal quantile transform (NQT) is applied to transform the predictors and predictand into standard Normal distribution before fitting the statistical model.[96–98] NQT is defined as follows[26]:

$$u = \Phi^{-1}(F_x(x)), \tag{6}$$

where $x$ is the original variable; $F_x$ is the CDF of $x$; $\Phi^{-1}$ is the inverse of standard Normal distribution; and $u$ is the transformed variable, which follows standard Normal distribution. NQT is always applied before fitting the statistical model. After obtaining the coefficients of the conditional distribution-based models, the inverse of NQT is then applied to transform data into original space.

Krzysztofowicz found that BPO outperformed the traditional MOS model because it provided the full distribution of the predictand with fewer parameters.[26] Later, Krzysztofowicz and Evans developed a similar model called the BPF for continuous quantities such as temperature.[27] Hamill et al. tested the BPF using surface temperature reforecasts and found that the NQT in BPF still needed improvements and that whether the verification situation was consistent with the climate situation affected its accuracy.[99] Krzysztofowicz also developed BPE for ensemble weather forecasts.[28] Reggiani and Weerts applied BPO for precipitation postprocessing as a component in their Bayesian forecasting system (BFS).[95]

### *Ensemble Preprocessor*

Besides Bayes' theorem-based methods, the conditional distribution formula can be directly applied to postprocess the raw forecasts. Schaake et al. developed the EPP to generate temperature and precipitation ensemble forecasts from single-value forecasts.[24] It is called 'preprocessor' because it deals with meteorological forecasts, which are inputs for hydrological models. The single-value forecasts can be the ensemble mean forecasts from an NWP model.

To model the distribution of precipitation with an intermittent property, a mixture of discrete and continuous distribution is applied. The marginal PDF of forecasts x is as follows[24]:

$$f_x(x) = (1-p_x)\delta(x) + p_x f_{xc}(x|x>0), \tag{7}$$

where $p_x$ is the PoP occurrence, $\delta(x)$ is the Dirac delta function, and $f_{xc}$ is the conditional distribution of precipitation amount given that precipitation occurs. The marginal PDF of observations can be modeled similarly.

The NQT procedure is also employed to convert the observations and raw forecasts into standard Normal space. Here, the transformed predictand is noted as $u$, and the transformed predictor is noted as $v$. The joint distribution of $u$ and $v$ follows a bivariate Normal distribution. According to the conditional distribution formula and the property of the Normal distribution, the conditional distribution of the predictand $u$, given the predictors $v$, namely,

$$f_{v|u}(v|u) = \frac{f_{uv}(u,v)}{f_u(u)} \tag{8}$$

is also a Normal distribution, with mean and variance as follows[24]:

$$\mu_{v|u} = \rho_{uv}u, \tag{9}$$

$$\sigma^2_{v|u} = 1 - \rho^2_{uv}, \tag{10}$$

where $\rho_{uv}$ is the correlation coefficient between $u$ and $v$. Thus, the conditional distribution of the predictand can be obtained. Finally, an inverse of NQT is applied to transform data to original space.

After the conditional distribution is estimated, the Schaake shuffle[80] is used to obtain ensemble members that can maintain spatiotemporal and intervariable dependency of the observations. A modified version of EPP, in which the parameters are optimized by minimizing the mean continuous ranked probability score (CRPS), has been applied successfully in the National Weather Service River Forecast System.[25] Its effectiveness was verified using GFS or GEFS precipitation reforecasts in China.[100,101]

### *Conditional Distribution-based Methods for Hydrological Forecasts*

In this subsection, three conditional distribution-based methods originally designed for hydrological forecasts, namely, HUP, MCP, and BJP, are reviewed. Like BPO, Bayes' theorem is also suited for the postprocessing of hydrological forecasts. Krzysztofowicz et al. implemented Bayes' theorem for hydrological forecasts and named it the HUP.[42,43] Due to the strong autocorrelation property of hydrological time series, recent past observations are added in prior distribution and the likelihood function in HUP. After NQT is applied, a first-order Markov process for prior distribution is assumed; thus, the observed river stage $y_t$ at time $t$ is determined only by the previous value $y_{t-1}$ at time $t$ – 1. Then, the prior distribution is modeled by linear regression as follows[42,43]:

$$y_t = c_t y_{t-1} + \varepsilon_t, \tag{11}$$

where $c_t$ is the coefficient. For the likelihood function, the observed river stage at time $t - 1$ and $t$ and the observed river stage at the forecast time $t_0$ are added to the linear regression as predictors[42,43]:

$$x_t = a_t y_t + e_t y_{t-1} + d_t y_0 + b_t + \varepsilon_t, \tag{12}$$

where $a_t$, $b_t$, $e_t$, and $d_t$ are coefficients.

Then, a posterior distribution of observations, given corresponding forecasts, can be obtained by Bayes' theorem, similar to Eq. (3) for BPO.

Krzysztofowicz implemented HUP combined with BPO in the BFS to account for input and output uncertainty together in a formal and consistent statistical framework.[102] However, Reggiani et al. found that HUP with a Markov chain assumption was not suitable for large basins.[44] The original HUP cannot discern rising and falling limbs, and its forecast skill decreases rapidly with lead time.[44] To adapt HUP for a large basin, such as the Rhine river basin, the Bayesian ensemble uncertainty processor (BEUP) was developed, which takes several observations at upstream stations as predictors in prior and likelihood functions to include more information.[45]

There are other conditional distribution-based models, including MCP and BJP.[50] Todini et al. developed a conditional distribution-based model, MCP, which is similar to EPP for univariate situations. Coccia and Todini demonstrated that univariate MCP is an alternative to HUP.[49] The advantage of MCP is that the derivation for univariate situations can be conveniently extended to multivariate situations, making it suitable for multimodel, multisite, and multilead time problems.[48] Moreover, MCP can deal with the heteroscedasticity of the residuals through the use of the multivariate truncated Normal distributions.[48] Recently, this model was also applied to the postprocessing of ensemble forecasts.[103]

The BJP[50] approach is also based on the conditional distribution formula. The advantage of BJP is that the Bayesian inference technique, Markov chain Monte Carlo (MCMC), is applied for the estimation of parameters and the corresponding uncertainty. The Bayesian inference technique allows the model to include datasets with nonconcurrent, missing data, or zero-value occurrences.[51] Similar to the other methods mentioned in this section, data transformations, such as Yeo–Johnson transformation[104] or log-sinh transformations,[105] should also be applied before parameter fitting. The BJP method has been applied for the postprocessing of seasonal streamflow,[50] precipitation,[106,107] and daily streamflow.[52]

## Regression-based Methods

Regression-based methods are convenient tools for modeling the statistical correlation between the predictand (i.e., the observation) and the predictors (i.e., the model forecasts). Here, we review several regression-based methods for hydrometeorological forecasts, including the MOS, EMOS, LR, QR, variance inflation, and member-by-member (MBM) regression method. Two other regression methods specifically designed for hydrological forecasts, namely, autoregressive models and GLMPP, are also reviewed.

### *Model Output Statistics*

MOS[16] is one of the earliest statistical postprocessing methods along with perfect prognosis[15] (Perfect Prog). Here, we focus on the MOS method for postprocessing precipitation forecasts. The MOS is a linear regression model between the conditional PoP exceeding a threshold $T_j$ and the predictors from NWP forecasts[16,108]:

$$P\left(y \geq T_j\right) = a_{0j} + \sum_{i=1}^{I} a_{ij} u_i, \tag{13}$$

where $u_i$ represents the NWP forecasts transformed by the grid-binary transform,[109] $a_{ij}$ are the coefficients, $y$ is the actual precipitation amount, and $P(y \geq T_j)$ is the conditional exceedance probability of $y$ exceeding the cutoff amount $T_j$. The predictors could be precipitation forecasts and other meteorological forecasts, such as vorticity or vertical velocity.[109] The grid-binary transform mainly maps each predictor value in each grid point into '1' or '0,' which indicates the exceedance or nonexceedance of a specified cutoff level and then interpolates the grid point value into the value at each station.[109]

For postprocessing of ensemble forecasts, Erickson proposed to apply MOS to each member simply, but this method leads to results that converge toward the climatology.[86,110] Coelho et al. developed a Bayesian version of MOS, also called 'forecast assimilation.'[111,112] Wilks showed that the forecast assimilation method performed worse than later methods, such as EMOS or ensemble dressing.[86] Kalman filter can also be considered a special type of MOS approach, which includes dynamic regression coefficients and is suitable for NWP models undergoing frequent upgrading and with a short training dataset.[113]

### *Ensemble Model Output Statistics*

Classical linear regression-based models, such as MOS, assume that the predictand follows Gaussian distribution with static variance, neglecting the relationship between the predictive error and the ensemble spread. In fact, the variance of postprocessed forecast error should increase with the increase of ensemble spread. To make use of the ensemble spread information, Gneiting et al. proposed the EMOS, also called nonhomogeneous Gaussian regression (NGR).[12,29] As shown in the following formula, the mean of the predictand $y$ is a linear function of the ensemble member $x_i$; moreover, the variance of the residual is a linear function of the ensemble spread $S$[26]:

$$y = a + \sum_{i=1}^{I} b_i x_i + \varepsilon, \tag{14}$$

$$\mathrm{Var}(\varepsilon) = c + dS^2. \tag{15}$$

The coefficients $a$, $b_i$, $c$, and $d$ in the regression formula are obtained by minimizing the CRPS, and $b_i$, $c$, and $d$ are constrained to be nonnegative. In this way, the variance of the forecasts maintains the information of ensemble spread.

To model variables that are truncated and non-Gaussian, several models were developed in the following years. EMOS models with predictand following left-censored generalized extreme value (GEV) distribution or censored and shifted gamma (CSG) distribution have been developed for precipitation postprocessing.[31,114,115] EMOS has also been applied to the postprocessing of hydrological forecasts after Box–Cox transform.[116]

### *Logistic Regression and Extended Logistic Regression (ExLR)*

LR is suitable for predicting the probability of binary events, such as the probability of precipitation accumulation to exceed or not exceed a given threshold. The LR formula is as follows[76]:

$$P\left(y < q_j | \mathbf{x}\right) = \frac{\exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right)}{1 + \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right)} = \Lambda\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right), \tag{16}$$

where $y$ is the predictand, $q_j$ is the $i$th threshold, $\mathbf{x}$ is the predictor vector, $\boldsymbol{\beta}$ is the coefficient vector, and $\Lambda$ is the logistic distribution. To obtain the probability of predictand falling below various thresholds, LRs can be established separately for each threshold.[117] However, this method faces several problems. For example, the LR lines estimated separately may cross with each other, leading to unreasonable results.[33] To deal with those problems, Wilks et al. developed ExLR, which adds a function of the exceeding thresholds in the regression formula as follows[33]:

$$P\left(y < q_j | \mathbf{x}\right) = \Lambda\left[\alpha g\left(q_j\right) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right], \tag{17}$$

where $g(q_j)$ is a monotonic function of the thresholds $q_j$, and $\alpha$ is a coefficient to be estimated. The regression coefficient vector $\boldsymbol{\beta}$ is the same for all thresholds, resulting in parallel regression lines to avoid the crossing problem of the original LR method.

Moreover, the traditional LR only predicts probability for exceeding several discrete thresholds, which limits its performance for extreme events.[118] In contrast, the ExLR provides predictions of the full distribution and performs well, even for extreme events.[118] Besides, traditional LR includes many more parameters than ExLR, so traditional LR is more sensitive to the training window size, while ExLR is not severely affected by sample size.[86]

To improve the extended LR for heteroscedastic predictand, such as precipitation, Messner et al. assumed that the dispersion of the regression is a linear function of the ensemble spread,[34,119] which is similar to Gneiting's EMOS. The extended LR has also been applied to hydrological postprocessing.[120]

### *Quantile Regression*

Classical linear regression predicts the mean value of $y$ conditional on $\mathbf{x}$; similarly, the QR model predicts the quantile values of $y$ conditional on $\mathbf{x}$.[121] The following is the formula of the QR for a specific quantile $p$:[121]

$$Q^{(p)}(y|\mathbf{x}) = a_0^{(p)} + \sum_{i=1}^{I} a_i^{(p)} x_i,\ 0 < p < 1. \tag{18}$$

The coefficients $\mathbf{a} = a_0, a_1\ldots, a_I$ are obtained by minimizing the least absolute deviation (LAD) function.[121]

For precipitation, Bremnes developed a two-step method to estimate the probability by probit regression and to predict the amount of precipitation by QR.[36] Friederichs and Hense applied a censored QR for precipitation downscaling.[37] There are also several applications of QR in hydrological forecasts.[63,122,123] The advantage of QR is that it directly estimates the quantiles of the predictand and is suitable for heteroscedastic quantities such as precipitation.[36] However, Coccia and Todini found that the QR method does not work well when the error variance increases nonlinearly with the magnitude of the predictand.[49] Besides, the QR method also suffers from the 'quantile crossing' problem, which means that the estimated lower quantile is higher than the estimated higher quantiles.[63,122] Another problem is the extrapolation of the model to extremes not included in the training sample.[63] To solve these problems, López et al. developed several alternative configurations of QR.[122] Bogner et al. applied the QR combined with a neural network (QRNN), which includes quantile rearranging and log-normal distribution fitting to eliminate the problems in traditional QR.[63]

### *Autoregression-based Methods*

To postprocess a streamflow or river stage with strong autocorrelation, autoregressive models that include time-lagged predictors are often applied. Here, we present a parsimonious method based on a autoregressive model described by Seo et al. as an example.[58] The NQT was first applied to transform the predictors and the predictand into a Gaussian distribution before fitting the statistical model. The main part of Seo's model is an autoregressive exogenous model, or ARX (1, 1), which is as follows[58]:

$$y_{t+1} = (1-b)y_t + bx_{t+1} + \varepsilon_{t+1}, \tag{19}$$

where $y_t$ and $y_{t+1}$ are the NQT-transformed streamflow observations at time step $t$ and $t+1$, respectively; $x_{t+1}$ is the NQT-transformed streamflow predicted by a hydrological model at time step $t+1$; $\varepsilon_{t+1}$ is the aggregate hydrological uncertainty at time step $t+1$; and parameter $b$ is the weight of the model forecast. As a result, the postprocessed streamflow is the weighted combination of the model forecasts at time step $t+1$ and the observation at time step $t$. The parameter is estimated by minimizing the forecast mean square error and the difference between the CDF of forecast and that of the observation at the same time.[58] This ARX model was adopted as the hydrological ensemble postprocessor (EnsPost) in the HEFS of the U.S. National Weather Service.[4]

To consider the dependency at different time scales, Bogner et al. applied wavelet transform to decompose the original time series into different time scales and then fitted a vector autoregressive model with exogenous input (VARX).[62] Then, HUP was combined with VARX to estimate the predictive conditional distribution.[62] There are other autoregressive models, such as autoregressive moving average (ARMA)[59] and AR for multimodel combinations.[60] Methods based on Kalman filter and ensemble Kalman filter have a close relationship with autoregressive models, and they have also been applied to the postprocessing of streamflow forecasts.[57,124]

### *General Linear Model Postprocessor*

Zhao et al. developed the GLMPP[53] specifically for hydrological postprocessing. Similar to autoregressive models, GLMPP is a linear regression model that predicts future observations with future forecasts and recent past observations and simulations. The NQT is first applied to transform the predictors and predictand into standard Normal distribution before fitting the regression model. The GLMPP model is as follows:[53]

$$\mathbf{Y} = \mathbf{A}\cdot\mathbf{X} + \mathbf{B}\cdot\mathbf{E}, \tag{20}$$

where $\mathbf{Y} = \widetilde{\mathbf{Q}}^{f}_{\text{obs}}$ is the NQT-transformed predictand vector, including the observations in the forecast period; $\mathbf{X} = \left[\widetilde{\mathbf{Q}}^{f}_{\text{sim}}, \widetilde{\mathbf{Q}}^{a}_{\text{obs}}, \widetilde{\mathbf{Q}}^{a}_{\text{sim}}\right]^{T}$ is the NQT-transformed predictor vector, including the simulations in the future forecast period (denoted by '$f$'), the observations, and corresponding simulations in the recent past analysis period (denoted by '$a$'); $\mathbf{E}$ is the error vector; and $\mathbf{A}$ and $\mathbf{B}$ are the coefficient matrix.

GLMPP is a convenient and effective tool to predict future observations at different lead times based on recent past forecasts and observations. Its performance has been demonstrated in several experiments using streamflow data from the Model Parameter Estimation Experiment (MOPEX).[54,55]

### *Variance Inflation and Member-by-Member Methods*

The variance inflation method was originally developed to calibrate ensemble seasonal streamflow forecasts.[88] The aim is to adjust the ensemble mean and spread so that the adjusted ensemble forecasts are statistically indistinguishable from the truth in a climatological sense.[125] The regression model is as follows[125]:

$$g^{i}_{t} = \alpha\overline{f}_{t} + \beta\varepsilon^{i}_{t}, \tag{21}$$

where $g^{i}_{t}$ is the $i$th adjusted ensemble member at time $t$; $\overline{f}_{t}$ is the unadjusted ensemble mean; $\varepsilon^{i}_{t}$ is the deviation of the unadjusted ensemble member to the ensemble mean, namely $\varepsilon^{i}_{t} = f^{i}_{t} - \overline{f}_{t}$, $\alpha$ and $\beta$ are the coefficients to adjust the mean and the spread of the new ensemble members, respectively.

To calibrate the ensemble mean and spread, the parameters are estimated based on the following two constraints[125]:

1. The average climatological variance of the adjusted ensemble members should equal the climatological variance of the real observation, namely,

$$\sigma_{g} = \sigma_{o}. \tag{22}$$

2. The correlation of the adjusted ensemble members with the unadjusted ensemble mean should equal the correlation of the real observation with the unadjusted ensemble mean, namely,

$$\text{corr}\left(g_{t}, \overline{f}_{t}\right) = \text{corr}\left(o_{t}, \overline{f}_{t}\right). \tag{23}$$

Van Schaeybroeck and Vannitsem extended the variance inflation method to the situation of multiple predictors, called the MBM approach.[38] The member-by-member approach achieved CRPS scores similar to those of the EMOS method and also avoided the undercorrection of the ensemble spreads problem of the MOS.[38] Moreover, the MBM method maintains the rank of the ensemble and thus preserves the correlation structure of the raw forecasts, which is especially important for further application in hydrology.[38]

## Ensemble Dressing Method

The ensemble dressing method first adds ('dress') the raw ensemble members with the error distribution and then forms a mixture probability distribution of the predictand using the sum of these dressed kernels. Here, we introduce the 'best member' method developed by Roulston and Smith.[40] In this method, the error distribution is estimated from the 'best member', which is defined as the member closest to the observation among all the ensembles. Let $y_t$ be the real observation at time $t$ and $\mathbf{x}_t = \{x_{t,k}, k = 1,2, \ldots, K\}$ the corresponding ensemble forecasts with $K$ members. The best member $x^{*}_{t}$ is the one that minimizes the $\|y_t - x_{t,k}\|$ for a given norm $\|\cdot\|$ as follows:[39]

$$x^{*}_{t} = \underset{x_{t,k}}{\text{argmin}} \left\|y_{t} - x_{t,k}\right\|. \tag{24}$$

Then, the distribution $p^{*}_{\varepsilon}$ of the best member's error $\varepsilon^{*} = y_{t} - x^{*}_{t}$ is estimated from the archive. Typically, a certain kernel function is assumed (e.g., Gaussian kernel) for the distribution of the error. The final mixture distribution is obtained as the average of the error distributions for each member as follows[39]:

$$p(y_{t}|\mathbf{x}_{t}) \approx \frac{1}{K}\sum_{k=1}^{K} p^{*}_{\varepsilon}\left(y_{t} - x_{t,k}\right). \tag{25}$$

After Roulston and Smith introduced the 'best member' method, several improvements have been made. Wang and Bishop developed the second-moment constraint dressing method to make the forecasts more reliable.[41] As the 'best member' method is only used for correcting the underdispersion problem, Fortin et al. developed a weighted ensemble dressing

**TABLE 1** | List of Representative Works on Single-Model Postprocessing Methods (Not Exhaustively)

| Type | Method | Abbreviation | References for Meteorological Forecasts | References for Hydrological Forecasts |
|---|---|---|---|---|
| Ad-hoc | Rank histogram recalibration | RHR | Hamill and Colucci (1998)[87] | |
| Distribution transform | Quantile mapping | QM | Piani et al. (2010)[23] | Hashino et al. (2007)[22] |
| Nonparametric | Analog method | AM | Hamill and Whitaker (2006)[19] | |
| | Conditional bias-penalized indicator cokriging | CBP-ICK | | Brown and Seo (2010)[64] and (2013)[65] |
| Conditional distribution | Bayesian processor of output/forecast | BPO/BPF | Krzysztofowicz (2006)[94] | |
| | Bayesian processor of ensemble | BPE | Krzysztofowicz (2008)[28] | |
| | Ensemble preprocessor | EPP | Schaake et al. (2007)[1,24] | |
| | Hydrological uncertainty processor | HUP | | Krzysztofowicz et al. (2000)[43] |
| | Model conditional processor | MCP | | Todini et al. (2008)[47] |
| | Bayesian joint probability | BJP | Robertson et al. (2013)[106] | Wang et al. (2009)[50] |
| Regression | Perfect prognosis | PP | Klein et al. (1959)[15] | |
| | Model output statistic | MOS | Glahn and Lowry (1972)[16] | |
| | Ensemble Kalman filter | EnKF | Crochet (2004)[113] | Vrugt et al. (2005)[124] |
| | Autoregressive model | AR | | Seo et al. (2006)[58], Bogner et al. (2011)[62] |
| | General linear model postprocessor | GLMPP | | Zhao et al. (2011)[53] |
| | Ensemble MOS | EMOS | Gneiting et al. (2005)[29], Scheuerer et al. (2014)[115] | Hemri et al. (2015)[116] |
| | Logistic regression | LR | Wilks (2009)[33] | |
| | Extended logistic regression | ExLR | Messner et al. (2015)[35] | Fundel and Zappa (2011)[120] |
| | Quantile regression | QR | Bremnes et al. (2004)[36] | Weerts et al. (2011)[123] |
| | Variance inflation | VI | | Wood and Schaake (2008)[88] |
| | Member-by-member regression | MBM | Van Schaeybroeck and Vannitsem (2015)[38] | |
| | Error-in-variable MOS | EVMOS | Vannitsem (2009)[127] | Roulin and Vannitsem (2014)[128] |
| Kernel density | Ensemble dressing | ED | Roulston and Smith (2003)[40] | Pagano et al. (2013)[129] |

method to add a different weight to each member, which is suitable for both under- and overdispersion problems.[39] Boucher compared these three kinds of dressing methods and found that Fortin's method performed the best, although more training data were needed for this method.[126] Furthermore, bias correction should be applied when using some of the ensemble dressing methods.[126]

## Further Discussion on Postprocessing Methods for Single-Model Forecasts

Table 1 shows the postprocessing methods reviewed in this section. The applications of these methods for meteorological forecasts are represented in the fourth column in Table 1. Among these methods, the early ad-hoc methods perform less well than other newly developed methods. As mentioned in Section *Quantile Mapping*, QM performs poorly when raw forecasts suffer from problems of reliability or coherence. AM is a nonparametric method that does not need the assumption of the distribution form of the predictand. However, its performance depends heavily on the sufficiency of training datasets. For example, AM may perform less well for extreme events that rarely appear in the historical archives.[31] Parametric methods are able to avoid this problem by extrapolating the statistical relations established by the datasets of normal events with extreme events as long as the statistical assumptions are reasonable.[31]

Several comparison experiments have been conducted for the parametric postprocessing methods, including LR, ExLR, EMOS, and ensemble dressing.[86,118,126,130] Wilks concluded that LR, EMOS, and ensemble dressing were the three most promising methods among postprocessing methods in their experiments.[86] Wilks and Hamill later compared these three methods based on precipitation forecasts from the Global Forecast System (GFS) and found that there is no single best method for all applications among these three methods, and the difference in the length of the training dataset leads to larger differences in forecast skill than different methods do.[86,130] As mentioned in Section *Logistic Regression and Extended Logistic Regression*, the traditional LR is more sensitive to training sample size than other methods and thus performs less well, especially for extreme events.[118] There are still several other methods that are not introduced in this section due to space limitation, such as the error-in-variable model (EVMOS).[127,131]

Several representative postprocessing methods for hydrological forecasts are listed in the fifth column of Table 1. Some methods originally designed for the postprocessing of meteorological variables can also be applied for hydrological forecasts, such as EMOS, QR, LR, and ensemble dressing.[116,120,123,129] Besides, there are other methods that have been developed specifically for hydrological forecasts, such as HUP, autoregressive model, and GLMPP. A unique feature of these models is that the temporal correlation between hydrological quantities is strong, and thus, past recent observations or forecasts should be added as predictors in the models to utilize the temporal dependence information.

Several comparisons among the postprocessing methods for hydrological forecasts have been conducted. Van Andel et al. conducted the intercomparison experiment based on observation and forecast datasets from the MOPEX,[132] where hydrological models were driven by real observations so that the uncertainty of meteorological input is removed.[7] Preliminary results showed that the skill of postprocessed forecasts is not very sensitive to the choice of postprocessors, but is sensitive to other factors such as the choice of predictors.[133] To further compare the regression-based methods with dressing methods for hydrological forecasts, Boucher et al. conducted several synthetic experiments and found that although regression-based and ensemble dressing methods have similar overall performance, they perform differently with regard to resolution and reliability: the former leads to better resolution, while the latter achieves better reliability.[126] Boucher et al. recommended ensemble dressing, rather than a simplified EMOS, for most of the cases in their experiments to obtain forecast reliability, especially for underdispersed, asymmetric forecasts.[126] Recently, Mendoza et al. conducted a comparative experiment among seven postprocessing methods, including linear blending, QM, extended LR, QR, AM, and GLMPP.[134] Preliminary results showed that no one postprocessor outperforms other methods for all situations, and the performance of postprocessors depends on factors such as basin types.[134] There are also some studies on whether to apply postprocessors for meteorological forecasts, or postprocessors for hydrological forecasts, or both.[90,135–137] Results showed that postprocessors for hydrological forecasts are needed even after postprocessing of the meteorological forcing.[135,136] In seasonal hydrological forecasting, Yuan and Wood also found that combining postprocessors for precipitation forecasts and postprocessors for streamflow forecasts achieved the best performance.[137]

There are other postprocessing methods for hydrological forecasts that are not introduced here due to space limitation, but most of them are similar to the statistical models reviewed here. For example, machine learning methods, such as uncertainty estimation based on local errors and clustering (UNEEC), perform similar to piecewise linear regression.[138] Besides, to avoid the drawbacks of parametric distribution assumptions, some nonparametric methods have been developed, such as the postprocessor based on indicator cokriging.[64,65]

## MULTIMODEL POSTPROCESSING METHODS

Because all models are imperfect, many researchers resort to multimodel postprocessing methods to combine outputs from multiple models and achieve better forecast skill. This section reviews several postprocessing methods for forecasted variables from multiple models. Several linear model-averaging methods are reviewed in the first subsection, including the 'poor man's ensemble' method,[67] the SMA,[70] and the 'superensemble' method.[72,73] In the second subsection, the popular BMA[75] is reviewed.

### Linear Model-Averaging Methods

This subsection reviews three linear model-averaging methods: poor man's ensemble method, SMA, and superensemble method. The 'poor man's ensemble' method simply combines forecasts from different forecasting centers to generate ensemble forecasts.

This method samples uncertainty in the ensemble forecast system via the different forecasting techniques used by operational centers.[139] It was applied in several early meteorological forecasting studies, and the multimodel averaging forecasts achieved better forecast skill than forecasts from a single model because of the improved forecast skill of the ensemble mean.[67–69,139]

SMA simply combines each individual model output with equal weights. SMA can be expressed by the following formula[70,71]:

$$S = \overline{O} + \sum_{i=1}^{N} \frac{F_i - \overline{F}_i}{N}, \tag{26}$$

where $S$ is the SMA prediction, $\bar{O}$ is the time mean of historical observations, $N$ is the total number of models, $\overline{F}_i$ is the time mean of the historical forecasts by model $i$, and $F_i$ is the prediction by the $i$th model. Hagedorn et al. applied SMA for DEMETER multimodel seasonal ensemble forecast systems and demonstrated the superiority of multimodel forecasts relative to single-model forecasts.[70] Georgakakos et al. applied SMA for hydrological forecasts and also obtained more skillful and reliable simulation results than a single model.[71]

Methods that are more complex include model-dependent weights, which reflect each model's performance in the training period. The superensemble method is an example where the weights are estimated by linear regression[73]:

$$S = \overline{O} + \sum_{i=1}^{N} a_i \left(F_i - \overline{F}_i\right), \tag{27}$$

The weights $a_i$ are estimated at each grid point by minimizing the sum of squared residuals. Krishnamurti et al. applied the superensemble method for a combination of multiple meteorological forecasts, including seasonal/multiseasonal precipitation.[73] Recently, Krishnamurti et al. provided a theoretical framework of multimodel superensemble methods for weather and climate application.[74] The superensemble method can also be applied for hydrological forecasts.[66]

## Bayesian Model Averaging

The BMA[75] is one of the most popular multimodel postprocessing methods developed in recent years. Similar to the 'best member' dressing method, the output of BMA is also a mixture probability distribution. However, instead of finding a 'best member,' as in dressing methods, in BMA, the weight for each member is estimated by the posterior probability of each member to be the best forecast. The final output of the BMA is a weighted average of forecasts as follows[75]:

$$p(y|F_1, F_2, \ldots, F_N) = \sum_{i=1}^{N} w_i \cdot g_i(y|F_i), \tag{28}$$

where $g_i(y|F_i)$ is the distribution of observation $y$, given the raw model forecast $F_i$. The weight $w_i$ is the posterior probability of forecast $F_i$ to be the best forecast, which reflects the forecast performance for the training data. The weights are assumed to be nonnegative and add up to 1. The weights and the parameters in the forecast distribution of each member can be estimated by the expectation–maximization (EM) algorithm.[140] As the BMA weight represents each model's performance for the training dataset, BMA is suitable for the postprocessing of distinguishable members from different models. On the contrary, the standard ensemble dressing methods mentioned in the last section are mainly for the postprocessing of undistinguishable members, such as members obtained by perturbing initial conditions of one model.[141]

As there are more parameters in BMA than in other postprocessing methods such as EMOS, the training sample size for BMA should be large enough to avoid overfitting.[142] Hamill et al. suggested that reforecast datasets should be applied to fit BMA if available.[142] However, as reforecast datasets are often not available for all models in the multimodel postprocessing of meteorological forecasts, alternative methods have been proposed, such as enlarging sample size by using samples from supplemental locations and applying simpler merging methods such as QM and ensemble dressing.[143]

In addition to Raftery's BMA in 2005, several variants of BMA have been developed. For postprocessing of precipitation forecasts, Sloughter et al. applied LR to determine the PoP and used Gamma distribution for the precipitation amount.[77] Duan et al. applied the BMA model with Box–Cox-transformed predictand.[76] Fraley et al. extended BMA to situations with missing members or exchangeable members.[144] More advanced methods, such as particle filter, Monte Carlo, and copula functions, have been applied to improve BMA for non-Gaussian variables in hydrology.[145–148] BMA has been combined with other methods for further applications. For example, Kleiber et al. developed the geostatistical model averaging (GMA) method, which combines BMA with geostatistical methods to interpolate forecasts between gauge stations.[149,150] Marty et al. developed the

Bayesian processor of ensemble member (BPEM) model , which combines BMA and Krzysztofowicz's BPO by hierarchical model to calibrate and downscale weather forecasts.[151]

For postprocessing and downscaling of seasonal climate forecasts, Luo and Wood developed a Bayes' theorem-based Bayesian merging method,[152,153] which has been applied to generate climate inputs for seasonal hydrological forecasting.[154,155] Wang et al. developed a modified version of BMA to merge forecasts from different GCM models.[156] In this model, the Dirichlet prior distribution function is applied to stabilize the weights when significant sampling variability exists.[156] Moreover, a cross-validation likelihood function is used to replace the classical likelihood function to better represent the predictive abilities of models.[156] Schepen et al. suggested that to obtain better forecasts from GCM models, the 'Calibration, Bridging, and Merging' (CBaM) framework should be applied, which combines three techniques together: (1) to calibrate raw forecasts from each model, (2) to bridge large-scale climatic indices (e.g., ENSO) with local weather forecasts, and (3) to merge all the calibrated and bridged forecasts from multiple models.[157] The CBaM method has been successfully applied for seasonal precipitation, temperature, and streamflow forecasts.[157–160]

### Further Discussion on Multimodel Postprocessing Methods

Representative works of multimodel postprocessing methods mentioned in this section are listed in Table 2. The aim of multimodel postprocessing is to achieve better predictions by combining the outputs of multiple models. The model combination methods mentioned in this section can be categorized into methods with equal weight (e.g., SMA) and methods with model-dependent weight (e.g., super-ensemble and BMA). The former type of methods is easier to implement than the latter type, and they are already able to obtain forecasts with better skill than single-model forecasts both for meteorological and hydrological forecasting.[67,70,71] The second type of methods that apply model-dependent weight according to each model's historical performance are theoretically appealing, but they are more complicated and still need further research. The benefits of applying model-dependent weight will be few if most raw ensembles are similar.[161]

Among the methods with model-dependent weight, the superensemble method is simpler than BMA, but the coefficients may be negative and not suitable for the averaging of probability.[161] The BMA weights, however, are constrained to be positive and represent the forecast performance of each member. However, the BMA method still suffers from several problems, such as overfitting,[142] problematic treatment of extreme events,[162] and overweighting of the climatology.[163] Aside from the methods mentioned in this section, there are many other methods for multimodel postprocessing,[164] such as the heuristic methods,[68,78,79] which could not be reviewed in detail here due to space limitations.

## POSTPROCESSING METHODS TO ACCOUNT FOR SPATIOTEMPORAL AND INTERVARIABLE DEPENDENCY

The aforementioned postprocessing methods are established for a single hydrometeorological quantity at a fixed location and lead time. However, for further application in hydrology, spatiotemporal weather trajectories should be generated from the calibrated probability distributions, which preserve the spatial, temporal, and intervariable dependency structure. In this section, methods to generate ensemble members that maintain those dependency structures are reviewed. These methods can be categorized into two types: parametric methods and nonparametric methods.

### Parametric Methods

Parametric methods are generally extensions of univariate EMOS, BMA, or classical regression for multivariate situations considering the spatial, temporal,

**TABLE 2** | List of Representative Works on Postprocessing of Multimodel Forecasts (Not Exhaustively)

| Types | Method | References for Meteorological Forecasts | References for Hydrological Forecasts |
|---|---|---|---|
| Equal Weight | Poor man's ensemble | Mylne et al. (2002)[67] | |
| | Simple model average (SMA) | Hagedorn et al. (2005)[70] | Georgakakos et al. (2004)[71] |
| Model-dependent Weight | Superensemble | Krishnamurti et al. (1999)[72] | Ajami et al. (2006)[66] |
| | Bayesian model averaging (BMA) | Raftery et al. (2005)[75], Sloughter et al. (2007)[77] | Duan et al. (2007)[76] |

or intervariable dependencies. The dependency structures are usually modeled by parametric copulas.[165] Here, we consider a postprocessing model to preserve temporal dependency by Gaussian copula as an example.[166]

Gaussian copula is an $L$-variate function $C$ as follows[165]:

$$C(u_1,\ldots,u_L) = \Phi_\Sigma\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_L)\right) \\ = \Phi_\Sigma\left(\Phi^{-1}\left(F_{y_1}(y_1)\right),\ldots,\Phi^{-1}\left(F_{y_L}(y_L)\right)\right), \quad (29)$$

where $y_j$ is the predictand, $F_{y_j}$ is the CDF of $y_j$, $\Phi^{-1}$ is the inverse of the CDF of a standard univariate Normal distribution, and $\Phi_\Sigma$ is the CDF of a standard $L$-variate Normal distribution with covariance $\Sigma$.

We noted that $z_j = \Phi^{-1}(u_j)$ and then $(z_1, \ldots, z_L) \sim N(0, \Sigma)$. Then, the covariance between $z_j$ at different lead times can be modeled by mathematical functions. Here, an exponential function is applied as follows[166]:

$$\mathrm{cov}\left(z_{t,l_1}, z_{t,l_2}\right) = \exp\left(-\frac{|l_1 - l_2|}{v}\right),\ 0 < l_1, l_2 \le L, \quad (30)$$

where $z_{t,l_1}$ and $z_{t,l_2}$ are the transformed predictand at forecast date $t$ with lead time $l_1$ and $l_2$, respectively; $L$ is the maximum of the lead time; and $v$ is a parameter to be estimated.

To preserve spatial dependency for meteorological forecasts (e.g., precipitation), the geostatistical output perturbation (GOP) model was combined with traditional univariate postprocessing methods. Gel et al. added the GOP-based Gaussian error fields to the deterministic weather fields to incorporate the spatial dependency of weather forecasts at different locations.[167] Berrocal et al. proposed a spatial BMA model that combined GOP with the univariate ensemble BMA.[83,84] Later, the spatial EMOS model was also developed, which is a combination of EMOS and the GOP method.[81,82] Experiments show that the spatial EMOS model performed similar to the spatial BMA approach, but the former is more efficient.[82] As for intervariable dependency, methods have been developed to combine univariate EMOS or BMA with Gaussian joint distribution or Gaussian copula.[168–170]

To preserve the temporal correlation between predictand at different lead times for hydrological forecasts (e.g., streamflow), Engeland and Steinsland developed a Gaussian copula-based regression model.[171] Hemri et al. applied temporal BMA and temporal EMOS for streamflow postprocessing over different lead times and obtained a better forecast skill than univariate BMA or EMOS.[116,172]

## Nonparametric Methods

Nonparametric methods are generally reordering methods like Schaake shuffle[80] and ECC[85], both of which could be seen as empirical copula methods.[85] The aim of these methods is to reorder the ensemble forecast members to preserve the spatial, temporal, or intervariable dependency structure. The steps of Schaake shuffle and ECC for each margin $l$ are shown in Boxes 1 and 2. The two methods differ mainly in the choice of dependency template. For Schaake shuffle, the template is chosen from past observations in the historical archive.[80] The underlying assumption is that the dependency structure does not change over time.[80] For ECC, the template is the raw ensemble members, which assumes that the raw ensembles are able to represent the actual dependency structure.[85] Recently, Schefzik developed a variant of Schaake shuffle, namely SimSchaake method, which chooses the reordering template from past observations under similar situations.[173] The SimSchaake was reported to perform better than the ECC method with broader applicability.[173] Scheuerer et al. provided an alternative similarity criterion to select similar historical dates in the Schaake shuffle to preserve the spatial–temporal dependence of meteorological forecasts for hydrological application.[174] The above methods were originally developed for

**BOX 1**

**SCHAAKE SHUFFLE**

1. Draw $M$ samples $\mathbf{X}_l = \left(x_1^l,\ldots,x_M^l\right)$ from the calibrated cumulative distribution function $F_l$. Sort the forecast ensembles such that $\widetilde{\mathbf{X}}_l = \left(\widetilde{x}_1^l,\ldots,\widetilde{x}^{Ml}\right) = \left(x_{(1)}^l,\ldots,x_{(M)}^l\right)$ with $x_{(1)}^l \le x_{(2)}^l \le,\ldots,\le x_{(M)}^l$.
2. Choose $M$ samples of observations from the historical archive. Sort the observations $\mathbf{Y}_l = \left(y_1^l,\ldots,y_M^l\right)$ such that $\widetilde{\mathbf{Y}}_l = \left(\widetilde{y}_1^l,\ldots,\widetilde{y}_M^l\right) = \left(y_{(1)}^l,\ldots,y_{(M)}^l\right)$ with $y_{(1)}^l \le y_{(2)}^l \le,\ldots,\le y_{(M)}^l$. Denote the corresponding ranks of these observation samples as $rk_m^l$.
3. Arrange the forecast ensembles according to the rank dependency structure of the historical observations and obtain the reordered forecasts as $\mathbf{X}_l^{ss} = \left(\widetilde{x}_{rk_1^l}^l,\ldots,\widetilde{x}_{rk_m^l}^l\right)$.

**BOX 2**

**ENSEMBLE COPULA COUPLING**

1. Draw $M$ samples $\mathbf{X}_l = (x_1^l,\ldots,x_M^l)$ from the calibrated cumulative distribution function $F_l$. The sampling method varies among ECC-Q, ECC-R, and ECC-t.[85] Sort the forecast ensembles such that $\tilde{\mathbf{X}}_l = \left(\tilde{x}_1^l,\ldots,\tilde{x}_M^l\right) = \left(x_{(1)}^l,\ldots,x_{(M)}^l\right)$ with $x_{(1)}^l \le x_{(2)}^l \le,\ldots,\le x_{(M)}^l$.
2. Sort the raw forecast ensembles $\mathbf{R}_l = (r_1^l,\ldots,r_M^l)$ such that $r_{(1)}^l \le r_{(2)}^l \le,\ldots,\le r_{(M)}^l$. Denote the corresponding ranks of raw forecasts as $rk_m^l$.
3. Arrange the postprocessed samples according to the rank dependency structure of the raw forecast ensemble and obtain the reordered forecasts as $\mathbf{X}_l^{ECC} = \left(\tilde{x}_{rk_1^l}^l,\ldots,\tilde{x}_{rk_m^l}^l\right)$.

meteorological forecasts, but they could also applied for hydrological forecasts such as streamflow.[116]

## Further Discussion on Methods to Account for Spatiotemporal and Intervariable Dependency

Several representative references of both parametric and nonparametric methods to generate ensemble members that preserve the spatiotemporal and intervariable dependency structure are presented in Table 3. Between the two types of methods, parametric methods (e.g., the EMOS, BMA or classical regression-based methods) are based on parametric statistical assumptions and avoid the spurious and inaccurate predictions from the ensemble template in nonparametric methods.[82] One drawback is that these methods suffer from large computational burden; thus, parametric methods mainly address low-dimensional or highly structured correlations to avoid a large number of parameters to be fitted.[175] Moreover, temporal and spatial stationarity of the dependency structure is often assumed in parametric methods, which can be problematic in applications.[82,83] Contrarily, nonparametric methods (e.g., ECC and Schaake shuffle) are based on simple reordering processes, which make these methods more suitable for high-dimensional structures, such as simultaneous spatiotemporal and intervariable dependency structure. The drawback of this kind of methods is that their performance depends on the representativeness of the reordering templates. For example, the Schaake shuffle may not perform well for extreme events that rarely appear in historical archives.[4]

Some comparison experiments have been conducted for these two types of methods. For example, Hemri applied the nonparametric ECC and parametric Gaussian copula approach to postprocess the streamflow with different lead times. They found both methods to be generally suitable for modeling temporal dependency of hydrological forecasts.[116] Combinations of the two types of methods have also emerged, such as the bivariate EMOS–ECC method,[176] which models intervariable dependency

**TABLE 3** | List of Representative Works on Methods to Account for Dependency Structure (Not Exhaustively)

| Types | Method | Dependency | References for Meteorological Forecasts | References for Hydrological Forecasts |
|---|---|---|---|---|
| Parametric | EMOS-based | Spatial | Scheuerer and Buermann (2014)[30] | |
| | | Temporal | Pinson et al. (2011)[166] | Hemri et al. (2015)[116] |
| | | Intervariable | Baran and Möller (2016)[170] | |
| | BMA-based | Spatial | Berrocal (2007)[83] | |
| | | Temporal | | Hemri et al. (2013)[172] |
| | | Intervariable | Möller et al. (2013)[168] | |
| | Regression-based | Temporal | | Engeland and Steinsland (2014)[171] |
| Nonparametric | Schaake shuffle | Spatial/temporal | Clark et al. (2004)[80] | |
| | Ensemble copula coupling (ECC) | Spatial/temporal | Schefzik et al. (2013)[85] | Hemri et al. (2015)[116] |
| | SimSchaake | Spatial | Schefzik (2015)[173] | |

 

via parametric bivariate EMOS and describes the spatial dependencies via nonparametric ECC. Schefzik's experiments have demonstrated that this new method performed well for postprocessing temperature and wind speed at several locations simultaneously, but problems such as multivariate ranking and sampling still need to be addressed.[176]

## CONCLUSION

Statistical postprocessors are an integral part of HEFSs. Their purpose is to calibrate the biases and quantify the uncertainty of the raw forecasts. Various statistical postprocessing methods have been developed to calibrate raw hydrometeorological forecasts in recent decades. In this paper, statistical postprocessing methods for hydrometeorological forecasts are reviewed in three aspects: (1) postprocessing of the meteorological forecasts and hydrological forecasts, (2) postprocessing for consensus forecasts from multiple models, and (3) generating ensemble members that preserve the spatiotemporal and intervariable dependency. In the future, more experiments will be needed to verify and compare these methods, especially using real datasets instead of synthetic ones, and to provide suggestions on how to choose suitable postprocessing methods for specific operational applications.[118]

In addition to the methods reviewed in this paper, we indicated several directions that need further work in statistical postprocessing for hydrometeorological forecasts. First, for non-Gaussian hydrological variables, although data transformations such as NQT have been widely applied, problems such as extrapolation to extreme values have occurred.[177] New tools such as nonparametric models,[64] a copula-based method,[91,147,148] or machine learning methods[138] may help to address these problems.

Moreover, stationarity is often assumed in traditional statistical postprocessing methods. Under such an assumption, the statistical correlation between observations and raw forecasts from simulation models in the training period is similar to that in the verification period. However, this assumption is not always valid in hydrology.[178] Ceola et al. developed a theoretical framework to deal with hydrological nonstationarity, which may help to alleviate this problem.[179]

In addition, the forecasting of extreme events is of great importance for applications such as flood warnings. However, as there are far fewer extreme events than normal events in the historical archive, it is difficult to train a statistical postprocessing model for extreme events.[118] Traditional statistical postprocessing methods, such as classical linear regression, mainly deal with the mean of the predictand rather than extreme situations.[37] Thus, new methods need to be developed to deal with extreme events.

Another aspect that needs more attention is the treatment of systematic timing and spatial errors in raw precipitation and streamflow forecasts (e.g., the timing and location of the forecasted storm peaks are systematically shifted from the observed values). Those forecasts would be skillful if the timing and spatial errors were removed. However, most postprocessors build statistical relationships between observations and raw forecasts for a particular location and a specific forecast time and do not account for the timing and spatial errors, which makes the capture of the true skill of raw forecasts unlikely. A potential way to alleviate this kind of error is to incorporate spatiotemporal neighborhood information, which means that more sophisticated postprocessing methods to correct the timing and spatial errors of the forecasts need to be developed.[99,159]

Another problem emerges when postprocessing forecasts for locations without gauge stations. Several studies have been conducted on this topic. Kleiber et al. developed the geostatistical model averaging (GMA) for precipitation forecasts.[150] Later works include locally adaptive EMOS and standardized anomaly MOS (SAMOS).[180,181] For hydrological forecasts, Skoien et al. developed the top kriging-based EMOS method to interpolate EMOS parameters at calibrated locations to uncalibrated locations.[182] More methods still need to be developed in this field, especially for hydrological forecasts.

Finally, studies to integrate multiple ensemble forecasting techniques to account for total uncertainty of different sources are needed. Statistical postprocessing methods mainly address the uncertainty in model input and output. However, hydrometeorological forecasts include various uncertainties, such as model parameters, model structure, and model initial or boundary conditions. All these uncertainties should be considered together in order to gain a better understanding of the interactions among uncertainties throughout the forecasting process.[2,13] There have been some studies that try to address different uncertainty sources in an integrated manner, including the Bayesian Total Error Analysis (BATEA) and the Integrated Bayesian Uncertainty Estimator (IBUNE),[128,183–185] but those methods have not been implemented in a real-time forecasting system. Therefore, more studies are needed for the development of a total uncertainty approach in operational hydrometeorological forecasting.

## ACKNOWLEDGMENT

We thank two anonymous reviewers for their valuable comments that improved this manuscript substantially. This work was supported by the Natural Science Foundation of China (No. 210100025) and the Intergovernmental Key International Science and Technology Innovation Cooperation Program (No.2016YFE0102400).

## REFERENCES

1. Schaake JC, Hamill TM, Buizza R, Clark M. HEPEX: the hydrological ensemble prediction experiment. *Bull Am Meteorol Soc* 2007, 88:1541–1547.
2. Cloke HL, Pappenberger F. Ensemble flood forecasting: a review. *J Hydrol* 2009, 375:613–626.
3. Zappa M, Beven KJ, Bruen M, Cofiño AS, Kok K, Martin E, Nurmi P, Orfila B, Roulin E, Kai S. Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. *Atmos Sci Lett* 2010, 11:83–91.
4. Demargne J, Wu L, Regonda SK, Brown JD, Lee H, He M, Seo DJ, Hartman R, Herr HD, Fresch M, et al. The science of NOAA's operational hydrologic ensemble forecast service. *Bull Am Meteorol Soc* 2014, 95:79–98.
5. Schaake J, Franz K, Bradley A, Buizza R. The Hydrologic Ensemble Prediction EXperiment (HEPEX). *Hydrol Earth Syst Sci Discuss* 2006, 3:3321–3332.
6. Schaake J, Pailleux J, Thielen J, Arritt R, Hamill T, Luo L, Martin E, McCollor D, Pappenberger F. Summary of recommendations of the first workshop on postprocessing and downscaling atmospheric forecasts for hydrologic applications held at Meteo-France, Toulouse, France, 15–18 June 2009. *Atmos Sci Lett* 2010, 11:59–63.
7. van Andel SJ, Weerts A, Schaake J, Bogner K. Post-processing hydrological ensemble predictions intercomparison experiment. *Hydrol Process* 2012, 27:158–161.
8. Hamill TM, Hagedorn R, Whitaker JS. Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: precipitation. *Mon Weather Rev* 2008, 136:2620–2632.
9. Hagedorn R, Hamill TM, Whitaker JS. Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: two-meter temperatures. *Mon Weather Rev* 2008, 136:2608–2619.
10. Hamill TM, Whitaker JS, Mullen SL. Reforecasts: an important dataset for improving weather predictions. *Bull Am Meteorol Soc* 2006, 87:33–46.
11. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. *J R Stat Soc* 2007, 69:243–268.
12. Gneiting T, Katzfuss M. Probabilistic forecasting. *Annu Rev Stat Appl* 2014, 1:125–151.
13. Yuan X, Wood EF, Ma Z. A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *WIREs Water* 2015, 2:523–536.
14. Glahn B, Peroutka M, Wiedenfeld J, Wagner J, Zylstra G, Schuknecht B, Jackson B. MOS uncertainty estimates in an ensemble framework. *Mon Weather Rev* 2009, 137:246–268.
15. Klein WH, Lewis BM, Enger I. Objective prediction of five-day mean temperatures during winter. *J Atmos Sci* 1959, 16:672–682.
16. Glahn HR, Lowry DA. The use of model output statistics (MOS) in objective weather forecasting. *J Appl Meteorol* 1972, 11:1203–1211.
17. Van Den Dool HM. Searching for analogues, how long must we wait? *Tellus Ser A–Dyn Meteorol Oceanogr* 1994, 46:314–324.
18. Lorenz EN. Atmospheric predictability as revealed by naturally occurring analogues. *J Atmos Sci* 1969, 26:636–646.
19. Hamill TM, Whitaker JS. Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Mon Weather Rev* 2006, 134:3209–3229.
20. Hamill TM. Interpretation of rank histograms for verifying ensemble forecasts. *Mon Weather Rev* 2001, 129:550.
21. Hamill TM, Colucci SJ. Verification of eta RSM short-range ensemble forecasts. *Mon Weather Rev* 1997, 125:1312–1327.
22. Hashino T, Bradley AA, Schwartz SS. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol Earth Syst Sci* 2007, 3:939–950.
23. Piani C, Haerter JO, Coppola E. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor Appl Climatol* 2010, 99:187–192.
24. Schaake J, Demargne J, Hartman R, Mullusky M, Welles E, Wu L, Herr H, Fan X, Seo DJ. Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol Earth Syst Sci Discuss* 2007, 4:655–717.
25. Wu L, Seo DJ, Demargne J, Brown JD, Cong S, Schaake J. Generation of ensemble precipitation

forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. *J Hydrol* 2011, 399:281–298.

26. Krzysztofowicz R, Maranzano CJ. *Bayesian Processor of Output for Probabilistic Quantitative Precipitation Forecasting*. San Francisco: American Geophysical Union; 2006.

27. Krzysztofowicz R, Evans WB. Probabilistic forecasts from the National Digital Forecast Database. *Weather Forecast* 2008, 23:270–289.

28. Krzysztofowicz R. Bayesian processor of ensemble: concept and development. In: *19th Conference on Probability and Statistics in the Atmosphere Sciences*, New Orleans, Louisiana, 2008.

29. Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon Weather Rev* 2005, 133:1098–1118.

30. Scheuerer M, Büermann L. Spatially adaptive post-processing of ensemble forecasts for temperature. *J R Stat Soc* 2014, 63:405–422.

31. Scheuerer M, Hamill TM. Statistical post-processing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon Weather Rev* 2015, 143:4578–4596.

32. Hamill TM, Whitaker JS, Wei X. Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. *Mon Weather Rev* 2004, 132:1434–1447.

33. Wilks DS. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorol Appl* 2009, 16:361–368.

34. Messner JW, Mayr GJ, Wilks DS, Zeileis A. Extending extended logistic regression: extended vs. separate vs. ordered vs. censored. *Mon Weather Rev* 2014, 142:3003–3013.

35. Messner JW, Mayr GJ, Zeileis A. Heteroscedastic censored and truncated regression with crch. *R J* 2015, 8:1–12.

36. Bremnes JB. Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Mon Weather Rev* 2004, 132:338–347.

37. Friederichs P, Hense A. Statistical downscaling of extreme precipitation events using censored quantile regression. *Mon Weather Rev* 2007, 135:2365–2378.

38. Van Schaeybroeck B, Sp V. Ensemble post-processing using member-by-member approaches: theoretical aspects. *Q J Roy Meteor Soc* 2015, 141:807–818.

39. Fortin V, Favre A-C, Said M. Probabilistic forecasting from ensemble prediction systems: improving upon the best-member method by using a different weight and dressing kernel for each member. *Q J Roy Meteor Soc* 2006, 132:1349–1369.

40. Roulston MS, Smith LA. Combining dynamical and statistical ensembles. *Tellus Ser A—Dyn Meteorol Oceanogr* 2003, 55:16–30.

41. Wang X, Bishop CH. Improvement of ensemble reliability with a new dressing kernel. *Q J Roy Meteor Soc* 2005, 131:965–986.

42. Krzysztofowicz R, Herr HD. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *J Hydrol* 2001, 249:46–68.

43. Krzysztofowicz R, Kelly KS. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour Res* 2000, 36:3265–3277.

44. Reggiani P, Weerts AH. A Bayesian approach to decision-making under uncertainty: an application to real-time forecasting in the river Rhine. *J Hydrol* 2008, 356:56–69.

45. Reggiani P, Renner M, Weerts AH, Van Gelder PAHJM. Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resour Res* 2009, 45:335–345.

46. Herr HD, Krzysztofowicz R. Ensemble Bayesian forecasting system. Part I: theory and algorithms. *J Hydrol* 2015, 524:789–802.

47. Todini E. A model conditional processor to assess predictive uncertainty in flood forecasting. *Int J River Basin Manage* 2008, 6:123–137.

48. Todini E. From HUP to MCP: analogies and extended performances. *J Hydrol* 2013, 477:33–42.

49. Coccia G, Todini E. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrol Earth Syst Sci* 2011, 15:3253–3274.

50. Wang QJ, Robertson DE, Chiew FHS. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour Res* 2009, 45:1–18.

51. Wang QJ, Robertson DE. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resour Res* 2011, 47:1–19.

52. Zhao T, Wang QJ, Bennett JC, Robertson DE, Shao Q, Zhao J. Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. *J Hydrol* 2015, 528:329–340.

53. Zhao L, Duan Q, Schaake J, Ye A, Xia J. A hydrologic post-processor for ensemble streamflow predictions. *Adv Geosci* 2011, 29:51–59.

54. Ye A, Duan Q, Schaake J, Xu J, Deng X, Di Z, Miao C, Gong W. Post-processing of ensemble forecasts in low-flow period. *Hydrol Process* 2015, 29:2438–2453.

55. Ye A, Duan Q, Yuan X, Wood EF, Schaake J. Hydrologic post-processing of MOPEX streamflow simulations. *J Hydrol* 2014, 508:147–156.

56. Vrugt JA, Gupta HV, Nualláin B, Bouten W. Real-time data assimilation for operational ensemble streamflow forecasting. *J Hydrometeorol* 2006, 7:548–565.
57. Vrugt JA, Robinson BA. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour Res* 2007, 43:1–15.
58. Seo DJ, Herr HD, Schaake JC. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol Earth Syst Sci Discuss* 2006, 3:1987–2035.
59. Renner M, Werner MGF, Rademacher S, Sprokkereef E. Verification of ensemble flow forecasts for the river Rhine. *J Hydrol* 2009, 376:463–475.
60. Hopson TM, Webster PJ. A 1-10-day ensemble forecasting scheme for the major river basins of Bangladesh: forecasting severe floods of 2003–07. *J Hydrometeorol* 2010, 11:618–641.
61. Bogner K, Kalas M. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmos Sci Lett* 2008, 9:95–102.
62. Bogner K, Pappenberger F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resour Res* 2011, 47:1772–1780.
63. Bogner K, Liechti K, Zappa M. Post-processing of stream flows in Switzerland with an emphasis on low flows and floods. *Water* 2016, 8:115–115.
64. Brown JD, Seo D-J. A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *J Hydrometeorol* 2010, 11:642–665.
65. Brown JD, Seo DJ. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrol Process* 2013, 27:83–105.
66. Ajami NK, Duan Q, Gao X, Sorooshian S. Multimodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. *J Hydrometeorol* 2006, 7:755–768.
67. Mylne KR, Evans RE, Clark RT. Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting. *Q J Roy Meteor Soc* 2002, 128:361–384.
68. Atger F. The skill of ensemble prediction systems. *Mon Weather Rev* 1999, 127:1941–1953.
69. Ziehmann C. Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus Ser A—Dyn Meteorol Oceanogr* 2000, 52:280–299.
70. Hagedorn R, Doblas-Reyes FJ, Palmer TN. The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus Ser A—Dyn Meteorol Oceanogr* 2005, 57:219–233.
71. Georgakakos KP, Seo DJ, Gupta H, Schaake J, Butts MB. Characterizing streamflow simulation uncertainty through multimodel ensembles. *J Hydrol* 2004, 298:222–241.
72. Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* 1999, 285:1548–1550.
73. Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow TE, Bachiochi DR, Williford CE, Gadgil S, Surendran S. Multimodel ensemble forecasts for weather and seasonal climate. *J Climate* 2000, 13:4196–4216.
74. Krishnamurti TN, Kumar V, Simon A, Bhardwaj A, Ghosh T, Ross R. A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Rev Geophys* 2016, 54:336–377.
75. Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* 2005, 133:1155–1174.
76. Duan Q, Ajami NK, Gao X, Sorooshian S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv Water Res* 2007, 30:1371–1386.
77. Sloughter JM, Raftery AE, Gneiting T, Fraley C. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon Weather Rev* 2007, 135:3209–3220.
78. Shamseldin AY, O'Connor KM. A non-linear neural network technique for updating of river flow forecasts. *Hydrol Earth Syst Sci* 2001, 5:577–597.
79. Shamseldin AY, O'Connor KM, Liang GC. Methods for combining the outputs of different rainfall–runoff models. *J Hydrol* 1997, 197:203–229.
80. Clark M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R. The Schaake shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J Hydrometeorol* 2004, 5:243–262.
81. Scheuerer M, Büermann L. Spatially adaptive post-processing of ensemble forecasts for temperature. *J R Stat Soc* 2013, 63:405–422.
82. Feldmann K, Scheuerer M, Thorarinsdottir TL. Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon Weather Rev* 2015, 143:955–971.
83. Berrocal VJ, Raftery AE, Gneiting T. Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon Weather Rev* 2007, 135:1386–1402.


 

84. Berrocal VJ, Raftery AE, Gneiting T. Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann Appl Stat* 2008, 2:1170–1193.
85. Schefzik R, Thorarinsdottir TL, Gneiting T. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat Sci* 2013, 28:616–640.
86. Wilks DS. Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorol Appl* 2006, 13:243–256.
87. Hamill TM, Colucci SJ. Evaluation of eta RSM ensemble probabilistic precipitation forecasts. *Mon Weather Rev* 1998, 126:711–724.
88. Wood AW, Schaake JC. Correcting errors in streamflow forecast ensemble mean and spread. *J Hydrometeorol* 2008, 9:132–148.
89. Zhu Y, Luo Y. Precipitation calibration based on frequency matching method. *Weather Forecast* 2014, 30:1109–1124.
90. Verkade JS, Brown JD, Reggiani P, Weerts AH. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J Hydrol* 2013, 501:73–91.
91. Madadgar S, Moradkhani H, Garen D. Towards improved post-processing of hydrologic forecast ensembles. *Hydrol Process* 2014, 28:104–122.
92. Zhao T, Bennett JC, Wang Q, Schepen A, Wood AW, Robertson DE, Ramos M-H. How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J Climate* 2017, 30:3185–3196.
93. Messner JW, Mayr GJ. Probabilistic forecasts using analogs in the idealized Lorenz96 setting. *Mon Weather Rev* 2011, 139:1960–1971.
94. Krzysztofowicz R. Bayesian processor of output: probability of precipitation occurrence. In: *AGU Fall Meeting*, San Francisco, 2006.
95. Reggiani P, Weerts AH. Probabilistic quantitative precipitation forecast for flood prediction: an application. *J Hydrometeorol* 2008, 9:76–95.
96. Van der Waerden BL. Order tests for the two-sample problem and their power I. *Indagat Math* 1952, 14:453–458.
97. Van der Waerden BL. Order tests for the two-sample problem and their power II. *Indagat Math* 1953, 15:303–310.
98. Van der Waerden BL. Order tests for the two-sample problem and their power III. *Indagat Math* 1953, 15:311–316.
99. Hamill TM. Evaluation of the "Bayesian Processor of Forecasts" algorithm using GFS surface-temperature reforecasts. Available at: http://www.cdc.noaa.gov/people/tom.hamill/ (Accessed September 01, 2016).
100. Liu Y, Duan Q, Zhao L, Ye A, Tao Y, Miao C, Mu X, Schaake JC. Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin. *Hydrol Process* 2013, 27:57–74.
101. Ye A, Deng X, Ma F, Duan Q, Zhou Z, Du C. Integrating weather and climate predictions for seamless hydrologic ensemble forecasting: a case study in the Yalong river basin. *J Hydrol* 2017, 547:196–207.
102. Krzysztofowicz R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour Res* 1999, 35:2739–2750.
103. Todini E, Coccia G, Ortiz E. On the proper use of ensembles for predictive uncertainty assessment. In: *EGU General Assembly Conference*, Vienna, Austria, 2015.
104. Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. *Biometrika* 2000, 87:954–959.
105. Wang QJ, Shrestha DL, Robertson DE, Pokhrel P. A log-sinh transformation for data normalization and variance stabilization. *Water Resour Res* 2012, 48:1–7.
106. Robertson DE, Shrestha DL, Wang QJ. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrol Earth Syst Sci* 2013, 17:3587–3603.
107. Shrestha DL, Robertson DE, Bennett JC, Wang QJ. Improving precipitation forecasts by generating ensembles through postprocessing. *Mon Weather Rev* 2015, 143:3642–3663.
108. Antolik MS. An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J Hydrol* 2000, 239:306–337.
109. Jensenius Jr J. The use of grid-binary variables as predictors for statistical weather forecasting. In: *Preprints Twelfth Conference on Probability and Statistics in the Atmospheric Sciences*, American Meteorological Society, Toronto, 1992.
110. Erickson M. Medium-range prediction of PoP and max/min in the era of ensemble model output. In: *Conference on Weather Analysis and Forecasting*, American Meteorological Society, 1996.
111. Coelho CAS, Pezzulli S, Balmaseda M, Doblasreyes FJ, Stephenson DB. Forecast calibration and combination: a simple Bayesian approach for ENSO. *J Climate* 2004, 17:1504–1516.
112. Stephenson DB, Coelho CAS, Doblas-Reyes FJ, Balmaseda M. Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. *Tellus Ser A—Dyn Meteorol Oceanogr* 2005, 57:253–264.
113. Crochet P. Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in Iceland. *Meteorol Appl* 2004, 11:173–187.

 

114. Baran S, Nemoda D. Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting. *Environmetrics* 2016, 27:280–292.
115. Scheuerer M. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Q J Roy Meteor Soc* 2014, 140:1086–1096.
116. Hemri S, Lisniak D, Klein B. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resour Res* 2015, 51:7436–7451.
117. Hamill TM, Whitaker JS, Wei X. Ensemble re-forecasting: improving medium-range forecast skill using retrospective forecasts. *Bull Am Meteorol Soc* 2004, 132:3825–3830.
118. Williams RM, Ferro CAT, Kwasniok F. A comparison of ensemble post-processing methods for extreme events TL-140. *Q J Roy Meteor Soc* 2014, 140:1112–1120.
119. Messner JW, Mayr GJ, Zeileis A, Wilks DS, Messner JW, Mayr GJ, Zeileis A, Wilks DS. Heteroscedastic extended logistic regression for postprocessing of ensemble guidance. *Mon Weather Rev* 2014, 142:448–456.
120. Fundel F, Zappa M. Hydrological ensemble forecasting in mesoscale catchments: sensitivity to initial conditions and value of reforecasts. *Water Resour Res* 2011, 47:1–15.
121. Koenker R, Bassett G. Regression quantiles. *Econometrica* 1978, 46:33–50.
122. López López P, Verkade JS, Weerts AH, Solomatine DP. Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn river: a comparison. *Hydrol Earth Syst Sci* 2014, 18:3411–3428.
123. Weerts AH, Winsemius HC, Verkade JS. Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrol Earth Syst Sci* 2011, 15:255–265.
124. Vrugt JA, Diks CGH, Gupta HV, Bouten W, Verstraten JM. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resour Res* 2005, 41:407–412.
125. Johnson C, Bowler N. On the reliability and calibration of ensemble forecasts. *Mon Weather Rev* 2009, 137:1717–1720.
126. Boucher MA, Perreault L, Fo A, Favre AC. Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. *Hydrol Process* 2015, 29:1141–1155.
127. Vannitsem S. A unified linear model output statistics scheme for both deterministic and ensemble forecasts. *Q J Roy Meteor Soc* 2009, 133:937–948.
128. Roulin E, Vannitsem S. Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. *Hydrol Process* 2015, 29:1434–1449.
129. Pagano TC, Shrestha DL, Wang QJ, Robertson D, Hapuarachchi P. Ensemble dressing for hydrological applications. *Hydrol Process* 2013, 27:106–116.
130. Wilks DS, Hamill TM. Comparison of ensemble-MOS methods using GFS reforecasts. *Mon Weather Rev* 2007, 135:2379–2390.
131. Van Schaeybroeck B, Vannitsem S. Post-processing through linear regression. *Nonlinear Process Geophys* 2011, 18:147–160.
132. Duan Q, Schaake J, Andreassian V, Franks S, Goteti G, Gupta H, Gusev Y, Habets F, Hall A, Hay L. Model parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *J Hydrol* 2006, 320:3–17.
133. Brown J, Ramos M-H, Voisin N. Intercomparison of streamflow postprocessing techniques: first results of a HEPEX community experiment. In: *EGU General Assembly Conference Abstracts*, 2013.
134. Mendoza P, Wood A, Clark E, Nijssen B, Clark M, Ramos M-H, Voisin N. Improving medium-range ensemble streamflow forecasts through statistical post-processing. In: *AGU Fall Meeting Abstracts*, 2016.
135. Zalachori I, Ramos MH, On RG, Mathevet T, Gailhard J. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Adv Sci Res* 2012, 8:135–141.
136. Kang TH, Kim YO, Hong IP. Comparison of pre- and post-processors for ensemble streamflow prediction. *Atmos Sci Lett* 2010, 11:153–159.
137. Yuan X, Wood EF. Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resour Res* 2012, 48 :W12519. https://doi.org/10.1029/2012WR012256
138. Dogulu N, López López P, Solomatine DP, Weerts AH, Shrestha DL. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrol Earth Syst Sci* 2015, 19:3181–3201.
139. Ebert EE. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon Weather Rev* 2001, 129:2461–2480.
140. Dempster AP, Laird NM, Rubin DB. Maximum likelihood for incomplete data via the EM algorithm. *J R Stat Soc* 1977, 39B:1–39.


 

141. Schölzel C, Hense A. Probabilistic assessment of regional climate change in Southwest Germany by ensemble dressing. *Clim Dyn* 2011, 36:2003–2014.

142. Hamill TM. Comments on "Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging”. *Mon Weather Rev* 2007, 135:4226–4230.

143. Hamill TM, Engle E, Myrick D, Peroutka M, Finan C, Scheuerer M. The US national blend of models statistical post-processing of probability of precipitation and deterministic precipitation amount. *Mon Weather Rev* 2017, 145: 3441–3463.

144. Fraley C, Raftery AE, Gneiting T. Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon Weather Rev* 2010, 138:190–202.

145. Parrish MA, Moradkhani H, Dechant CM. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. *Water Resour Res* 2012, 48:151–153.

146. Rings J, Vrugt JA, Schoups G, Huisman JA, Vereecken H. Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. *Water Resour Res* 2012, 48:5634–5634.

147. Klein B, Meissner D, Kobialka H-U, Reggiani P. Predictive uncertainty estimation of hydrological multimodel ensembles using pair-copula construction. *Water* 2016, 8:1–22.

148. Madadgar S, Hamid M. Improved Bayesian multimodeling: integration of copulas and Bayesian model averaging. *J Hydrol* 2014, 6:2108–2123.

149. Kleiber W, Raftery AE, Baars J, Gneiting T, Mass CF, Grimit E. Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Mon Weather Rev* 2011, 139:2630–2649.

150. Kleiber W, Raftery AE, Gneiting T. Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *J Am Stat Assoc* 2011, 106:1291–1303.

151. Marty R, Fortin V, Kuswanto H, Favre AC, Parent E. Combining the Bayesian processor of output with Bayesian model averaging for reliable ensemble forecasting. *J R Stat Soc* 2014, 64:75–92.

152. Luo L, Wood EF. Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the Eastern United States. *J Hydrometeorol* 2008, 9:866–884.

153. Luo L, Wood EF, Pan M. Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *J Geophys Res Atmos* 2007, 112:185–194.

154. Li H, Luo L, Wood EF. Seasonal hydrologic predictions of low-flow conditions over eastern USA during the 2007 drought. *Atmos Sci Lett* 2008, 9:61–66.

155. Yoon JH, Mo K, Wood EF. Dynamic-Model-based seasonal prediction of meteorological drought over the contiguous United States. *J Hydrometeorol* 2012, 13:463–482.

156. Wang QJ, Schepen A, Robertson DE. Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging. *J Climate* 2011, 25:5524–5537.

157. Schepen A, Wang QJ. Toward accurate and reliable forecasts of Australian seasonal rainfall by calibrating and merging multiple coupled GCMs. *Mon Weather Rev* 2013, 141:4554–4563.

158. Schepen A, Wang QJ, Everingham Y. Calibration, bridging and merging to improve GCM seasonal temperature forecasts in Australia. *Mon Weather Rev* 2016, 144:2421–2441.

159. Schepen A, Wang QJ. Model averaging methods to merge operational statistical and dynamic seasonal streamflow forecasts in Australia. *Water Resour Res* 2015, 51:1797–1812.

160. Schepen A, Wang QJ, Robertson DE. Seasonal forecasts of Australian rainfall through calibration and bridging of coupled GCM outputs. *Mon Weather Rev* 2014, 142:1758–1770.

161. Johnson C, Swinbank R. Medium-range multimodel ensemble combination and calibration. *Q J Roy Meteor Soc* 2010, 135:777–794.

162. Bishop CH, Shanley KT. Bayesian model averaging’s problematic treatment of extreme weather and a paradigm shift that fixes it. *Mon Weather Rev* 2008, 136:4641–4652.

163. Hodyss D, Satterfield E, McLay J, Hamill TM, Scheuerer M. Inaccuracies with multi-model postprocessing methods involving weighted, regression-corrected forecasts. *Mon Weather Rev* 2015, 144:1649–1668.

164. Diks CGH, Vrugt JA. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environ Res Risk Assess* 2010, 24:809–820.

165. Schoelzel C, Friederichs P. Multivariate non-normally distributed random variables in climate research—introduction to the copula approach. *Nonlinear Process Geophys* 2008, 15:761–772.

166. Pinson P, Girard R. Evaluating scenarios of short-term wind power generation. *Wind Energy* 2011, 12:51–62.

167. Gel Y, Raftery AE, Gneiting T. Calibrated probabilistic mesoscale weather field forecasting. *J Am Stat Assoc* 2004, 99:575–583.

168. Möller A, Lenkoski A, Thorarinsdottir TL. Multivariate probabilistic forecasting using ensemble Bayesian

 

model averaging and copulas. *Q J Roy Meteor Soc* 2013, 139:982–991.

169. Baran S, Möller A. Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. *Environmetrics* 2015, 26:120–132.
170. Baran S, Möller A. Bivariate ensemble model output statistics approach for joint forecasting of wind speed and temperature. *Meteorol Atmos Phys* 2015, 140:1–14.
171. Engeland K, Steinsland I. Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water Resour Res* 2014, 50:182–197.
172. Hemri S, Fundel F, Zappa M. Simultaneous calibration of ensemble river flow predictions over an entire range of lead times. *Water Resour Res* 2013, 49:6744–6755.
173. Schefzik R. A similarity-based implementation of the Schaake shuffle. *Mon Weather Rev* 2015, 144:1909–1921.
174. Scheuerer M, Hamill TM, Whitin B, He M, Henkel A. A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation. *Water Resour Res* 2017, 53:3029–3046.
175. Wilks DS. Multivariate ensemble model output statistics using empirical copulas. *Q J Roy Meteor Soc* 2015, 141:945–952.
176. Schefzik R. Combining parametric low-dimensional ensemble postprocessing with reordering methods. *Q J Roy Meteor Soc* 2016, 142:2463–2477.
177. Bogner K, Pappenberger F, Cloke HL. Technical note: the normal quantile transformation and its application in a flood forecasting system. *Hydrol Earth Syst Sci* 2012, 16:1085–1094.
178. Milly PC, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ. Stationarity is dead: whither water management? *Science* 2008, 319:573–574.
179. Ceola S, Montanari A, Koutsoyiannis D. Toward a theoretical framework for integrated modeling of hydrological change. *WIREs Water* 2014, 1:427–438.
180. Scheuerer M, König G. Gridded, locally calibrated, probabilistic temperature forecasts based on ensemble model output statistics. *Q J Roy Meteor Soc* 2014, 140:2582–2590.
181. Stauffer R, Umlauf N, Messner JW, Mayr GJ, Zeileis A. Ensemble postprocessing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies. *Mon Weather Rev* 2017, 145:955–969.
182. Skøien JO, Bogner K, Salamon P, Smith P, Pappenberger F. Regionalization of post-processed ensemble runoff forecasts. *Proc Int Assoc Hydrol Sci* 2016, 373:109–114.
183. Kavetski D, Kuczera G, Franks SW. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resour Res* 2006, 42:W03408. https://doi.org/10.1029/2005WR004376
184. Ajami NK. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour Res* 2007, 43:1–19.
185. Demirel MC, Booij MJ, Hoekstra AY. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resour Res* 2013, 49:4035–4053.
186. Berrocal VJ, Gel Y, Raftery AE, Gneiting T. ProbForecastGOP: probabilistic weather forecast using the GOP method. 2012. Available at: https://CRAN.R-project.org/package=ProbForecastGOP (Accessed February 01, 2017).
187. Siegert S. SpecsVerification: forecast verification routines for ensemble forecasts of weather and climate. 2017. Available at: https://CRAN.R-project.org/package=SpecsVerification (Accessed February 01, 2017).
188. Yuen R, Gneiting T, Thorarinsdottir T, Fraley C. ensembleMOS: Ensemble model output statistics. 2013. Available at: https://CRAN.R-project.org/package=ensembleMOS (Accessed February 01, 2017).
189. NCAR. verification: Weather forecast verification utilities. 2015. Available at: https://CRAN.R-project.org/package=verification (Accessed February 01, 2017).
190. Koenker R. quantreg: Quantile regression. 2016. Available at: https://CRAN.R-project.org/package=quantreg (Accessed February 01, 2017).
191. Fraley C, Raftery AE, Sloughter JM, Gneiting T. ensembleBMA: probabilistic forecasting using ensembles and Bayesian model averaging. 2016. Available at: https://CRAN.R-project.org/package=ensembleBMA (Accessed February 01, 2017).
192. Cannon AJ. Quantile regression neural networks: implementation in R and application to precipitation downscaling. *Comput Geosci* 2011, 37:1277–1284.

## APPENDIX

Several commonly used postprocessing-related R packages are listed in the following table, which may be helpful for interested readers.[35,186–192]

List of Commonly Used Postprocessing -Related R Packages

| R Package | Method | References |
|---|---|---|
| ProbForecastGOP | Geostatistical output perturbation (GOP) | Berrocal et al. (2012)[186] |
| ensembleBMA | Ensemble BMA | Fraley et al. (2016)[191] |
| ensembleMOS | Ensemble MOS | Yuen et al. (2013)[188] |
| crch | Heteroscedastic censored and truncated regression (including logistic regression) | Messner et al. (2015)[35] |
| quantreg | Quantile regression | Koenker (2016)[190] |
| qrnn | Quantile regression neural networks | Cannon (2011)[192] |
| verification | Weather forecast verification utilities | NCAR—Research Applications Laboratory (2015)[189] |
| SpecsVerification | Verification routines for ensemble forecasts of weather and climate | Siegert (2017)[187] |