

# Multiscale Postprocessor for Ensemble Streamflow Prediction for Short to Long Ranges

BABAK ALIZADEH, REZA AHMAD LIMON,<sup>a</sup> AND DONG-JUN SEO

*Department of Civil Engineering, The University of Texas at Arlington, Arlington, Texas*

HAKSU LEE

*LEN Technologies, Inc., Oak Hill, Virginia*

JAMES BROWN

*Hydrologic Solutions Limited, Southampton, United Kingdom*

(Manuscript received 30 July 2019, in final form 6 November 2019)

## ABSTRACT

A novel multiscale postprocessor for ensemble streamflow prediction, MS-EnsPost, is described and comparatively evaluated with the existing postprocessor in the National Weather Service's Hydrologic Ensemble Forecast Service, EnsPost. MS-EnsPost uses data-driven correction of magnitude-dependent bias in simulated flow, multiscale regression using observed and simulated flows over a range of temporal aggregation scales, and ensemble generation using parsimonious error modeling. For comparative evaluation, 139 basins in eight River Forecast Centers in the United States were used. Streamflow predictability in different hydroclimatological regions is assessed and characterized, and gains by MS-EnsPost over EnsPost are attributed. The ensemble mean and ensemble prediction results indicate that, compared to EnsPost, MS-EnsPost reduces the root-mean-square error and mean continuous ranked probability score of day-1 to day-7 predictions of mean daily flow by 5%–68% and by 2%–62%, respectively. The deterministic and probabilistic results indicate that for most basins the improvement by MS-EnsPost is due to both magnitude-dependent bias correction and full utilization of hydrologic memory through multiscale regression. Comparison of the continuous ranked probability skill score results with hydroclimatic indices indicates that the skill of ensemble streamflow prediction with post processing is modulated largely by the fraction of precipitation as snowfall and, for non-snow-driven basins, mean annual precipitation.

## 1. Introduction

Streamflow simulations from hydrologic models contain errors propagated from uncertain forcings, model initial conditions, parameters and structures, and human control of storage and movement of water (Ajami et al. 2007; Doherty and Welter 2010; Gupta et al. 2012; Krzysztofowicz 1999; Montanari and Brath 2004; NRC 2006; Renard et al. 2010; Schaake et al. 2007b; Seo et al. 2006; Wood and Schaake 2008). For risk-based management of water resources and water-related hazards, it is necessary to quantify the uncertainties arising from

these sources (Borgomeo et al. 2014; Butts et al. 2004; Georgakakos et al. 2004; Hall and Borgomeo 2013; Hall et al. 2020). Ensemble forecasting has emerged in recent years as the methodology of choice for modeling and communicating forecast uncertainty (Cloke and Pappenberger 2009; Demargne et al. 2014; Demeritt et al. 2010; NRC 2006; Schaake et al. 2007b). In the United States, the National Weather Service (NWS) has recently implemented the Hydrologic Ensemble Forecast Service (HEFS; Demargne et al. 2014) at all River Forecast Centers (RFC) (Lee et al. 2018) following experimental operation at selected RFCs (Hartman et al. 2015; Kim et al. 2018; Wells 2017). To reduce biases in precipitation and temperature forecasts, the HEFS uses the Meteorological Ensemble Forecast Processor (MEFP; Schaake et al. 2007a; Wu et al. 2011; NWS 2017a). To reduce and quantify hydrologic uncertainty in streamflow prediction, the HEFS employs the ensemble

<sup>a</sup>Current affiliation: Servant Engineering and Consulting, PLLC, Austin, Texas.

*Corresponding author:* Babak Alizadeh, babak.alizadeh@mavs.uta.edu

postprocessor, EnsPost (Seo et al. 2006; NWS 2017b). In the HEFS, each MEFP-processed forcing ensemble member is input to the chain of hydrologic models in the Community Hydrologic Prediction System (CHPS; Gijbers et al. 2009; Roe et al. 2010). The resulting ensemble trace of “raw” streamflow forecast may be input to the ensemble postprocessor to produce an ensemble member of postprocessed streamflow forecast. The descriptor “post” arises from the fact that post processing of streamflow ensemble forecast occurs after the generation of raw ensemble streamflow forecast.

EnsPost was developed originally for short-range forecasting of natural flows in headwater basins and models predictive hydrologic uncertainty using a combination of probability matching (PM; Hashino et al. 2002; Madadgar et al. 2014) and autoregressive (AR)-1 model with an exogenous variable, or ARX (1,1) (Bennett et al. 2014; Damon and Guillas 2002), in bivariate normal space (Krzysztofowicz 1999; Seo et al. 2006). EnsPost applies PM and ARX(1,1) at a daily scale only. In reality, the characteristic time scales of error in model-simulated flow may span a range of scales, depending on the residence time of the hydrologic processes involved and the error characteristics of the forcings and the hydrologic models used (Blöschl and Sivapalan 1995). In addition, if the flow is strongly regulated, the errors may be reducible only over a certain range of temporal scales of aggregation due to the altered residence time and storage-outflow relationships. The postprocessing approach used in this work reflects the multiscale nature of the hydrologic and atmospheric processes (Blöschl and Sivapalan 1995; Carlu et al. 2019; Kumar 2011), and of hydrologic modeling and prediction, including parameter estimation (Mizukami et al. 2017), parameter regionalization (Samaniego et al. 2010), model evaluation (Rakovec et al. 2016), and data assimilation (Li et al. 2015).

The positive impact of postprocessing raw model simulations of streamflow in ensemble streamflow forecasting has been widely reported (Kim et al. 2018, 2016; Madadgar et al. 2014). Recently, it has also been shown that EnsPost significantly increases skill in ensemble forecasts of outflow from a water supply reservoir in North Texas during significant releases, in addition to that in ensemble inflow forecasts (Limon 2019). With increasing acceptance and adoption of ensemble streamflow forecasting by the operational community, developing more potent postprocessing methods has been a very active area of research. To that end, a number of comparison studies have recently been carried out. For postprocessing of meteorological forecast, Wilks (2006) compared direct model output (Wilks 2006), rank histogram recalibration (Hamill and Colucci 1998), single-integration Model Output Statistics (MOS;

Erickson 1996), ensemble dressing (Roulston and Smith 2003), logistic regression (Hamill et al. 2004), non-homogeneous Gaussian regression (Gneiting et al. 2005), forecast assimilation (Stephenson et al. 2005), and Bayesian model averaging (Raftery et al. 2005). He concluded that logistic regression (Duan et al. 2007; Hamill et al. 2004), ensemble MOS (Gneiting et al. 2005), and ensemble dressing outperform the others. For postprocessing of hydrologic forecast, Boucher et al. (2015) compared the regression and dressing methods using synthetic data. They concluded that the techniques have similar overall performance, and that the regression and dressing methods perform better in terms of resolution and reliability, respectively. Mendoza et al. (2016) used medium-range ensemble streamflow forecasts from the System for Hydrometeorological Applications, Research and Prediction, and compared quantile mapping (Mendoza et al. 2016; Hashino et al. 2006; Piani et al. 2010; Regonda and Seo 2008; Wood and Schaake 2008; Zhu and Luo 2015), logistic regression, quantile regression (Bjørnar Bremnes 2004; Bogner et al. 2016; Coccia and Todini 2011; Koenker and Bassett 1978) and the general linear model postprocessor (GLMPP; Zhao et al. 2011). They found that no single method performed best in all situations, and that the post processors’ performance depended on factors such as soil type and land use and hydroclimatic conditions of the basin. Ye et al. (2015) developed canonical events-based GLMPP for postprocessing of streamflow during dry season. Li et al. (2016) developed and evaluated for short-term streamflow forecasting a new method for error modeling where a sequence of simple error models, instead of a single complex model, is run through different stages. Recently, Li et al. (2017) and Vannitsem et al. (2018) carried out comprehensive reviews on the application of different postprocessing techniques.

In this paper, we introduce a new multiscale postprocessor for ensemble streamflow prediction for short to long ranges. By short and long ranges, we mean up to several days and at least 1 month ahead, respectively. The proposed technique, referred to herein as MS-EnsPost, is designed to reduce magnitude-dependent biases in raw model-simulated flow, and utilize all available skill that may exist over a range of temporal scales of aggregation in simulated and observed flows. We then comparatively evaluate MS-EnsPost with EnsPost for 139 basins in the services areas of eight RFCs in the continental United States (CONUS). As part of the evaluation, we also address the following research questions:

- How does the prediction skill, as measured by MS-EnsPost in reference to climatology, compare among the

different RFCs, and among different basins within an RFC?

- How does the above skill relate to the hydroclimatology of the basin?
- How does MS-EnsPost perform relative to EnsPost in reducing systematic errors and initial condition uncertainties?

This paper is organized as follows. [Section 2](#) describes the study basins and data used. [Section 3](#) describes the methods used. [Section 4](#) presents the results. [Section 5](#) provides the conclusions and future research recommendations.

## 2. Study basins and data used

A total of 139 basins, comprising 11, 13, 7, 19, 13, 28, 42, and 6 in the service areas of the Colorado Basin RFC (CBRFC), California Nevada RFC (CNRFC), Middle Atlantic RFC (MARFC), Missouri Basin RFC (MBRFC), North Central RFC (NCRFC), Northeast RFC (NERFC), Northwest RFC (NWRFC), and West Gulf RFC (WGRFC), respectively, are used (see [Fig. 1](#)). The CBRFC, CNRFC, MARFC, and WGRFC basins were used in previous studies of EnsPost ([Regonda and Seo 2008](#); [Seo et al. 2006](#)). The other basins were selected by the respective RFCs toward improved ensemble service. The basin areas range from 91 to 30 700 km<sup>2</sup> with 48, 41, and 50 basins under 500, between 500 and 1000, and over 1000 km<sup>2</sup>, respectively. The basins cover a wide range of hydroclimatology as may be seen in mean annual precipitation, the aridity index, and the fraction of precipitation as snowfall (see [Fig. 1](#)). The aridity index  $\varphi$  is defined as ([Budyko et al. 1974](#)):

$$\varphi = \frac{\bar{E}}{\bar{P}}, \quad (1)$$

where  $\bar{E}$  and  $\bar{P}$  denote the mean potential evaporation and precipitation (mm day<sup>-1</sup>), respectively. Semiarid arid basins have  $2 < \varphi < 3$  and  $3 < \varphi < 7$ , respectively ([Sankarasubramanian and Vogel 2003](#)). The fraction of precipitation as snowfall  $f_s$  is defined as

$$f_s = \frac{\bar{P}[T < 0^\circ\text{C}]}{\bar{P}}, \quad (2)$$

where  $T$  denotes the surface air temperature (°C) and  $\bar{P}[\ ]$  denotes the mean of precipitation for which the event bracketed holds true. Basins with  $f_s > 0.4$  are considered snow-dominated and tend to be located at higher elevations ([Knowles et al. 2006](#)). In general, a semiarid or arid basin has smaller predictability of streamflow than a humid basin, and a snowfall-driven basin has larger predictability than a rainfall-driven basin ([Berghuijs et al. 2014](#); [Berghuijs and Woods 2016](#)).

The data used for this study are mean daily observed and simulated streamflow. The historical observed mean daily streamflow, referred to as QME in the NWS, is obtained from the U.S. Geological Survey. The focus of this work is on reducing and quantifying hydrologic uncertainty. As such, the model output of interest is the simulated streamflow, which reflects hydrologic uncertainty only, rather than the streamflow forecast, which reflects both meteorological and hydrologic uncertainties ([Krzysztofowicz 1999](#); [Seo et al. 2006](#)). The simulated mean daily flow, or SQME, is derived from the simulated instantaneous flow, or SQIN, generated at a 6-h interval using the operational hydrologic models, and the observed forcings of mean areal precipitation, temperature, and potential evapotranspiration. For the remainder of this paper, by daily flow, we mean mean daily flow. The hydrologic models used are the Sacramento (SAC; [Burnash et al. 1973](#)) for soil moisture accounting, unit hydrograph ([Chow et al. 1988](#)) for surface runoff routing, and SNOW-17 ([Anderson 1973](#)) for snow ablation. The MARFC uses the continuous antecedent precipitation index model (API-CONT; [Fedora and Beschta 1989](#); [Sittner et al. 1969](#)) instead of SAC. The SQIN time series were produced by the respective RFCs using the Community Hydrologic Prediction System (CHPS; [Gijbers et al. 2009](#); [Roe et al. 2010](#)) based on the RFCs' historical forcings and calibrated model parameters. The CHPS is the main operational forecasting system at the RFCs, and uses the single (RES-SNGL) and joint (RES-J) reservoir regulation models, and the SSARR reservoir regulation (SSARRESV) model for simulation of reservoir operations ([Adams 2016](#); [NWS 2008a,b](#)). Reservoir models were included in the hydrologic modeling of 20 of the 139 locations used in this work. There are about 14 additional locations impacted by reservoir regulations that are not modeled. [Limon \(2019\)](#) has shown that, for a water supply reservoir in North Texas, the magnitude of reservoir modeling uncertainty may be comparable to that of all other hydrologic uncertainties combined, and may even approach that of the meteorological uncertainty. As such, flow regulations present a large additional challenge to streamflow postprocessing. Experience thus far indicates that at least 20 years' worth of data is necessary for estimation of the EnsPost parameters ([NWS 2017b](#)). The period of record used in this work common to both QME and SQME time series ranged from 12 to 66 years, and exceeded 30 years for over 90% of the basins.

## 3. Methods used

MS-EnsPost consists of three elements: bias correction, multiscale regression, and ensemble generation. [Figure 2](#) provides a schematic of the data flow and the

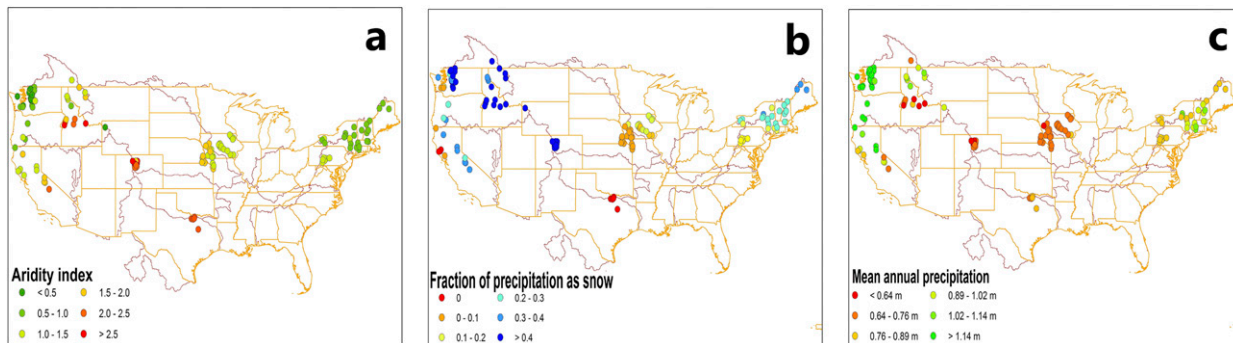


FIG. 1. Maps of 139 study basins, showing (a) the aridity index, (b) the fraction of precipitation as snowfall, and (c) mean annual precipitation.

associated processes. In this section, we describe the above elements and how MS-EnsPost is evaluated.

### a. Bias correction

We define within some time period of interest the multiplicative bias  $\beta_i$  in the simulated flow valid at the  $i$ th day  $q_i^s$  as

$$\beta_i = \frac{q_i^o}{q_i^s}, \quad (3)$$

where  $q_i^o$  denotes the observed flow valid at the  $i$ th day. Model-simulated high and low flows generally have different multiplicative biases. Hydrologic models tend to simulate the physical processes that govern high flows relatively more accurately (Dunne and Black 1970; Engman and Rogowski 1974; Freeze 1972; Horton 1933; Loague and VanderKwaak 2004). In addition, most hydrologic models, Continuous API being a prime example, are calibrated to perform better for high flows than for low flows (Fowler et al. 2018; Freer et al. 1996; Gan et al. 1997; Kim et al. 2007; Krause et al. 2005; Legates and McCabe 1999; Nash and Sutcliffe 1970; Smith et al. 2014). The bias correction procedure in MS-EnsPost is designed to address this dependence. Sample estimates of  $\beta_i$  at a daily scale are very noisy due to very large variabilities in  $q_i^o$  and  $q_i^s$ . To obtain stable estimates of the magnitude-dependent bias, the proposed procedure first pairs  $q_i^o$  and  $q_i^s$ , jointly sorts them in the ascending order of  $q_i^s$ , and aggregates the resulting daily flows over different time scales. The temporal aggregation and the attendant noise cancellation greatly reduce the sampling uncertainty in the estimated bias, compared to that without aggregation. The time-aggregated flows are expressed as

$$a_{k,(j)}^o = \sum_{i=(j_k-1)L_k+1}^{j_k L_k} q_{(i)}^o, \quad (4)$$

$$a_{k,(j)}^s = \sum_{i=(j_k-1)L_k+1}^{j_k L_k} q_{(i)}^s. \quad (5)$$

In the above, the symbol  $(i)$  signifies that the variable subscripted is sorted in the ascending order of  $q_i^s$ ,  $L_k$  denotes the  $k$ th time scale of aggregation, and  $j_k$  denotes the  $j$ th aggregation window of the  $k$ th scale within the period of record, and  $a_{k,(j)}^o$  and  $a_{k,(j)}^s$  denote the sorted observed and simulated flows aggregated over the  $j$ th time window of the  $k$ th time scale, respectively, where the symbol  $(j)$  signifies that the aggregation is based on the sorted daily flow. Equations (4) and (5) pool the simulated and observed flows such that, when averaged over the respective aggregation periods, the aggregated simulated flows are similar in magnitude to the conditioning flow  $q_i^s$  in Eq. (3). The bias for  $q_i^s$  at the  $k$ th aggregation scale  $\beta_{k,i}$  is given by

$$\beta_{k,i} = \frac{a_{k,(j)}^o}{a_{k,(j)}^s}, \quad i \in [(j_k - 1)L_k + 1, j_k L_k]. \quad (6)$$

In the above, the range for the index  $i$ , which is associated with the simulated flow to be bias-corrected  $q_i^s$  identifies the time scale of aggregation associated with the bias being estimated. In this work, we used  $L_k = 2^k$  (days),  $k = 1, \dots, 14$ , for the aggregation scales, but other choices are possible. In the above, the largest aggregation scale is almost 45 years long with which one would be applying a single multiplicative bias estimate for all simulated daily flows regardless of their magnitude.

Among the total of  $K$  different temporal scales of aggregation, the best-performing scale is identified via leave-one-year-out (or similar) cross validation using a period of record of  $N$  years as described below. First, the magnitude-dependent biases are estimated at the  $K$  different scales of aggregation using an  $(N - 1)$ -yr period of observed flow and matching model simulation. The resulting biases are applied to the simulated daily flow valid on each Julian day of the withheld year.

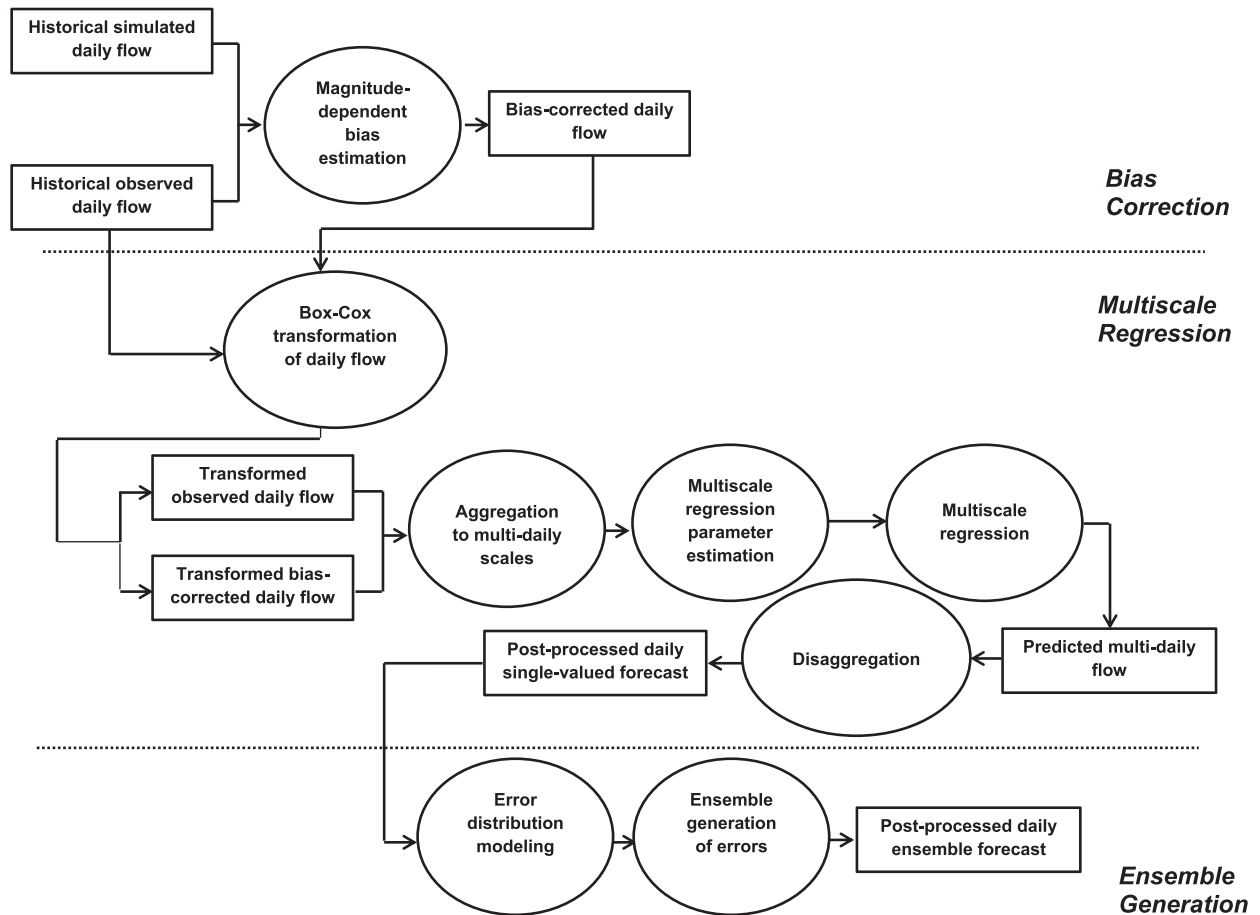


FIG. 2. Schematic of MS-EnsPost elements and dataflow.

The procedure then identifies the aggregation scale that produces the smallest RMSE over the year in the bias-corrected simulated daily flow by comparing with the verifying observed flow. Once completed for all  $N$  years, the leave-one-year-out cross validation produces a total of  $N$  different sets of magnitude-dependent biases for simulated daily flow. For a given  $q_i^s$ , the procedure then arithmetically averages the  $N$  different biases associated with the respective time windows that enclose the particular value of  $q_i^s$  in the sorted series. The procedure repeats the above steps for all possible values of  $q_i^s$ , from which the final single relationship between the simulated flow and the magnitude-dependent bias results. Figure 3 shows the schematic of the magnitude-dependent bias estimation procedure. Alizadeh (2019) provides examples of the resulting relationship of the multiplicative bias versus the magnitude of simulated daily flow.

*b. Multiscale regression*

Postprocessing generally seeks predictions at the highest possible temporal resolution. High dimensional

stochastic modeling necessary for such predictions, however, is a large challenge due to the complexity involved and large data requirements. In the multiscale regression approach used in this work, we solve instead a large number of very low-dimensional statistical modeling problems. Figure 4 illustrates the basic idea behind the approach in the context of predicting  $L_M$  day-ahead observed daily flow using the model-simulated daily flow valid over the  $L_M$  day-long prediction horizon, and the observed daily flow  $L_M - 1$  days into the past where  $M$  refers to the index for the largest aggregation scale. In this approach, rather than predicting  $q_i^o$ ,  $i = 1, \dots, L_M$ , using  $q_i^s$ ,  $i = 1, \dots, L_M$ , and  $q_i^o$ ,  $i = -(L_M - 1), \dots, 0$ , we predict  $a_{k,1}^o = \sum_{i=1}^{L_k} q_i^o$  using  $a_{k,1}^b = \sum_{i=1}^{L_k} q_i^b$  and  $a_{k,0}^o = \sum_{i=-(L_k-1)}^0 q_i^o$  for all time scales of aggregation,  $k = 1, \dots, M$ , where  $q_i^b$  denotes the bias-corrected model-simulated daily flow  $\beta_{k,i} q_i^s$  [see Eqs. (3) and (6)] and the subscripts 0 and 1 signify the current and one-step-ahead time intervals of the  $k$ th time scale, respectively. We then disaggregate the predicted multidaily flow to daily flow using the granular patterns of daily flow in the

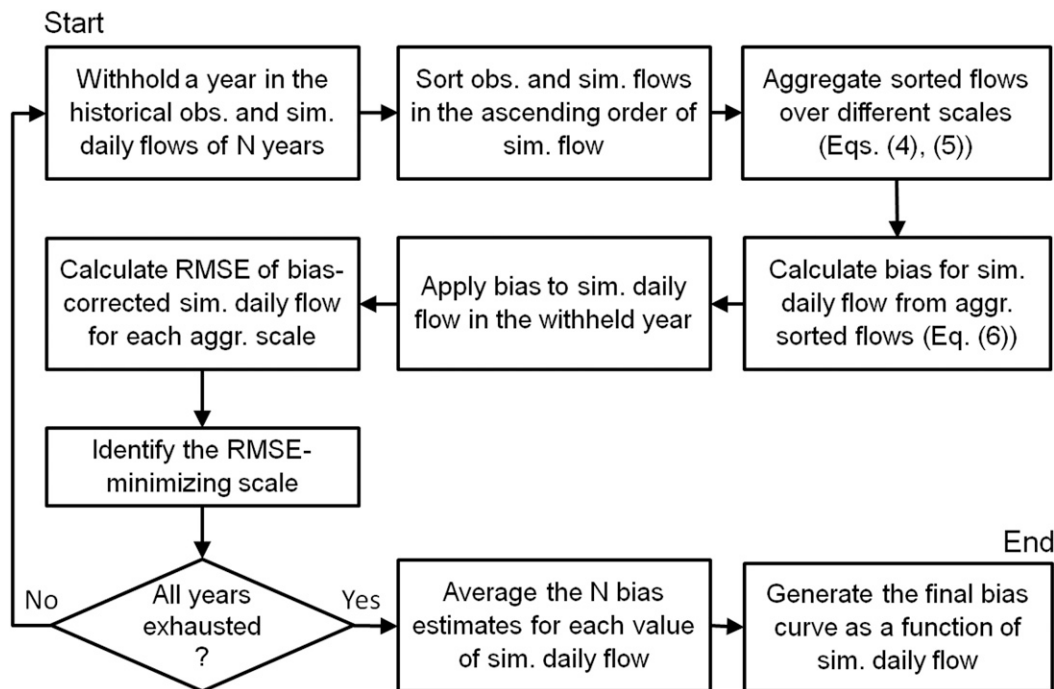


FIG. 3. Schematic of the magnitude-dependent bias correction procedure.

bias-corrected model-simulated daily flow. The above approach is motivated by the fact that the larger the temporal scale of aggregation is, the more skillful  $a_{k,1}^b$  is likely to be (Kim et al. 2018; Limon 2019). Similar approaches have also been used in postprocessing forecasts of precipitation (Kim et al. 2018; Schaake et al. 2007a) and streamflow (Regonda and Seo 2008). In this work, the prediction horizon used,  $L_M$ , is 32 days, and the aggregation scales used are 1, 2, 4, 8, 16, and 32 days (see Fig. 4). Depending on the application and the pattern of time-scale-dependent predictability, however, one may choose a different set of scales.

To predict the observed flow at the  $k$ th scale, we use the following linear model for the predicted, time-aggregated observed flow at the  $k$ th time scale,  $a_{k,1}^p$ , where the subscript 1 signifies that the prediction is for a single time step ahead at the  $k$ th time scale:

$$a_{k,1}^p = \varpi_k a_{k,0}^o + (1 - \varpi_k) a_{k,1}^b, \quad k = 1, \dots, M. \quad (7)$$

In the above,  $\varpi_k$  denotes the optimal weight for the time-aggregated observed flow at the  $k$ th scale  $a_{k,0}^o$ , where the subscript 0 signifies that  $a_{k,0}^o$  is valid at the current time step at the  $k$ th time scale. As may be seen in Eq. (7), one may consider multiscale regression a form of statistical fusion of observed streamflow with simulated flow over a range of time scales of aggregation. Significant improvement by multiscale regression is an indication that large uncertainties exist in the initial condition of the hydrologic models, or significant hydrologic memory exists in the surface and soil water storages of the basin. The optimal weight  $\varpi_k$  in Eq. (7) may be obtained via optimal linear (i.e., maximum likelihood) estimation (Deutsch 1965; Schweppe 1973; Gelb 1974) as

$$[\varpi_k(1 - \varpi_k)] = [\mathbf{U}^T \mathbf{R}^{-1} \mathbf{U}]^{-1} \mathbf{U}^T \mathbf{R}^{-1}. \quad (8)$$

In the above,  $\mathbf{U}$  denotes the  $(2 \times 1)$  unit vector, and  $\mathbf{R}$  denotes the error covariance matrix:

$$\mathbf{R} = \begin{bmatrix} \text{Var}(a_{k,0}^o - a_{k,1}^o) & \text{Cov}(a_{k,0}^o - a_{k,1}^o, a_{k,1}^b - a_{k,1}^o) \\ \text{Cov}(a_{k,0}^o - a_{k,1}^o, a_{k,1}^b - a_{k,1}^o) & \text{Var}(a_{k,1}^b - a_{k,1}^o) \end{bmatrix}. \quad (9)$$

The predicted daily flow for the  $i$ th day from the multiscale regression at the  $k$ th time scale  $q_{k,i}^p$  may be

obtained by disaggregating  $a_{k,1}^p$ ,  $k = 1, \dots, M$ , according to

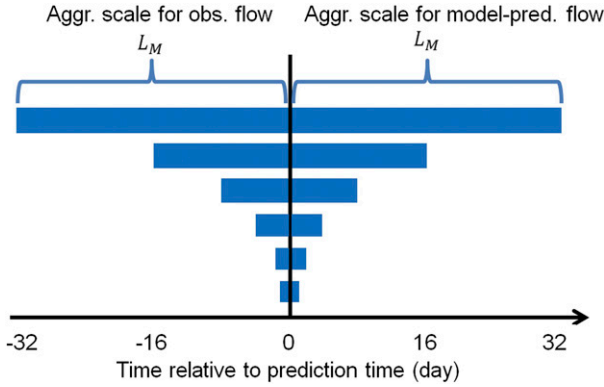


FIG. 4. Schematic of multiscale regression.

$$q_{k,i}^p = \frac{q_i^b}{a_{k,1}^b} a_{k,1}^p = \frac{a_{k,1}^p}{a_{k,1}^b} q_i^b. \quad (10)$$

Equation (10) amounts to adjusting the bias-corrected model-simulated daily flow  $q_i^b$  based on how much larger or smaller the predicted time-aggregated flow is relative to the time-aggregated bias-corrected flow, that is,  $a_{k,1}^p/a_{k,1}^b$ . Once the disaggregation process is complete for all time scales of aggregation, the final prediction of observed daily flow,  $q_i^p$ ,  $i = 1, \dots, L_M$ , is constructed from  $q_{k,i}^p$ ,  $k = 1, \dots, M$ , by choosing for each day  $i$  in the prediction horizon  $i = 1, \dots, L_M$ , the predicted daily flow  $q_{k,i}^p$  associated with the smallest  $k$ , that is, the smallest time scale of aggregation. In this way, if there are multiple predictions with overlapping prediction horizons, the procedure selects the one associated with the shortest lead time. Figure 4 shows the resulting time scales of aggregation over the prediction horizon of 32 days. Albeit heuristic, the above selection rule is based on the reasonable assumption that, the shorter the lead time is, the more skillful  $q_{k,i}^p$  is. If the period of record is too short relative to the largest time scale of aggregation, the estimation of the error covariance terms in Eq. (9) may not be possible due to small sample size. In such a case, the largest time scale may have to be reduced or dropped.

c. Error modeling and ensemble generation

The last element of MS-EnsPost models the error in the above prediction and its temporal structure for ensemble generation. This element models the time-correlated errors in  $q_i^p$ ,  $i = 1, \dots, L_M$  from multiscale regression. We define the error  $\varepsilon_i$  in the predicted daily flow  $q_i^p$  valid for the  $i$ th day in the prediction horizon as

$$\varepsilon_i = q_i^o - q_i^p, \quad \varepsilon_i \geq -q_i^p, \quad (11)$$

where  $q_i^o$  denotes the verifying observed daily flow. In EnsPost,  $q_i^o$  and  $q_i^p$  are normal quantile-transformed

(NQT) empirically (Krzysztofowicz and Kelly 2000; Krzysztofowicz and Herr 2001), which renders  $\varepsilon_i$  normal in the transformed space (Seo et al. 2006). In MS-EnsPost,  $q_i^o$  and  $q_i^p$  are Box-Cox-transformed (Box and Cox 1964) to avoid data-intensive empirical distribution modeling while rendering the error in the transformed space  $\tilde{\varepsilon}_i$  approximately normal and homoscedastic with respect to  $q_i^p$ :

$$\tilde{\varepsilon}_i = \tilde{q}_i^o - \tilde{q}_i^p = \frac{(q_i^o)^\lambda - 1}{\lambda} - \frac{(q_i^p)^\lambda - 1}{\lambda} = \frac{(q_i^o)^\lambda - (q_i^p)^\lambda}{\lambda},$$

$$\tilde{\varepsilon}_i \geq -(q_i^p)^\lambda/\lambda. \quad (12)$$

In the above,  $\lambda$  denotes the Box-Cox parameter, and  $\tilde{q}_i^o$  and  $\tilde{q}_i^p$  denote the transformed observed and predicted daily flows, respectively. The parameter  $\lambda$  is chosen such that the probability density function (PDF) of  $\tilde{\varepsilon}_i$  for any  $i$ , or  $\tilde{\varepsilon}$ , conditional on  $\{\tilde{\varepsilon} \geq q^p\}$ , or  $f_1(\tilde{\varepsilon}|\tilde{\varepsilon} \geq q^p)$ , may be approximated with truncated normal distribution (Robert 1995):

$$f_1(\tilde{\varepsilon}|\tilde{\varepsilon} \geq q^p) = N(m_{\tilde{\varepsilon}}, \sigma_{\tilde{\varepsilon}}^2; q^p), \quad (13)$$

where  $m_{\tilde{\varepsilon}}$  and  $\sigma_{\tilde{\varepsilon}}^2$  denote the mean and variance of the untruncated distribution, respectively, and  $q^p$  denotes the conditioning predicted flow in the truncated distribution. The unknown statistics  $m_{\tilde{\varepsilon}}$  and  $\sigma_{\tilde{\varepsilon}}^2$  may be solved for by equating the sample mean  $\mu_{\tilde{\varepsilon}}$  and variance  $s_{\tilde{\varepsilon}}^2$  of actual realizations of  $\tilde{\varepsilon}$  with the mean and variance of modeled  $\tilde{\varepsilon}$  as shown below:

$$\mu_{\tilde{\varepsilon}} = \int_0^\infty E[\tilde{\varepsilon}|\tilde{\varepsilon} \geq -(q^p)^\lambda/\lambda] f_2(q^p) dq^p, \quad (14)$$

$$s_{\tilde{\varepsilon}}^2 = \int_0^\infty E[\tilde{\varepsilon}^2|\tilde{\varepsilon} \geq -(q^p)^\lambda/\lambda] f_2(q^p) dq^p - \mu_{\tilde{\varepsilon}}^2, \quad (15)$$

where  $E[\cdot]$  denotes expectation, and  $f_2$  denotes the PDF of  $q^p$ . The expectations in Eqs. (14) and (15) may be expressed in terms of  $m_{\tilde{\varepsilon}}$ ,  $\sigma_{\tilde{\varepsilon}}^2$ , and  $\alpha$  as follows (Greene 2003):

$$E[\tilde{\varepsilon}|\tilde{\varepsilon} \geq -(q^p)^\lambda/\lambda] = m_{\tilde{\varepsilon}} + \sigma_{\tilde{\varepsilon}} \frac{\phi(\alpha)}{1 - \Phi(\alpha)}, \quad (16)$$

$$E[\tilde{\varepsilon}^2|\tilde{\varepsilon} \geq -(q^p)^\lambda/\lambda] = m_{\tilde{\varepsilon}}^2 + \sigma_{\tilde{\varepsilon}}^2 \left\{ 1 - \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \times \left[ \frac{\phi(\alpha)}{1 - \Phi(\alpha)} - \alpha \right] \right\}, \quad (17)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the standard normal PDF and cumulative distribution function (CDF), respectively, and the standardized lower bound  $\alpha$  is given by  $\alpha = [-(q^p)^\lambda/\lambda - m_{\tilde{\varepsilon}}]/\sigma_{\tilde{\varepsilon}}$ . Using Eqs. (14)–(17) and the empirical distribution of  $q^p$ , one may solve for  $m_{\tilde{\varepsilon}}$  and

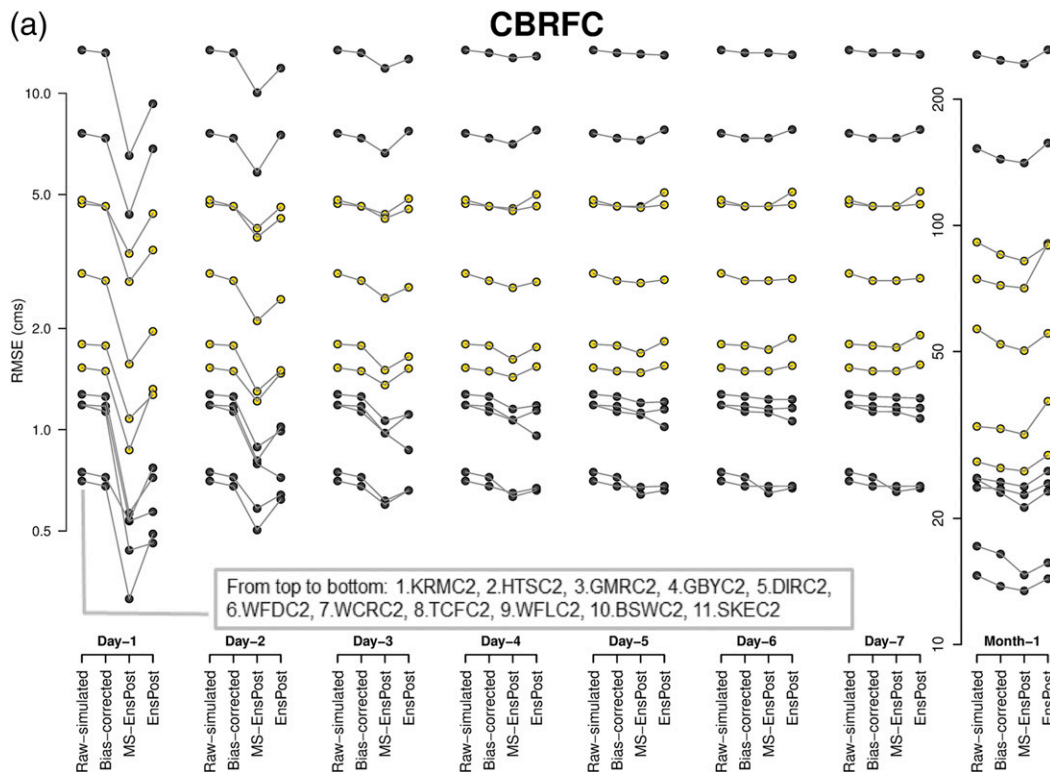


FIG. 5. (a) RMSE of the raw, bias-corrected, MS-EnsPost ensemble mean and EnsPost ensemble mean predictions for lead times of 1–7 days and 1 month for the basins in the CBRFC’s service area (yellow dots indicate basins with reservoir model included).

$\sigma_{\tilde{\varepsilon}}^2$  numerically. Once  $m_{\tilde{\varepsilon}}$  and  $\sigma_{\tilde{\varepsilon}}^2$  are estimated, an ensemble realization of  $\tilde{\varepsilon}$  conditioned on  $q_i^p$ , or  $\tilde{\varepsilon}_i(\omega)$ , may be generated by inverting the standard normal CDF (Robert 1995):

$$\tilde{\varepsilon}_i(\omega) = m_{\tilde{\varepsilon}} + \sigma_{\tilde{\varepsilon}} \Phi^{-1}\{\Phi(z_i^l) + U(\omega)[1 - \Phi(z_i^l)]\}. \quad (18)$$

In the above,  $U(\omega)$  denotes the sample realization of the  $[0,1]$  uniform random variable, the bracketed term represents  $\Phi\{[\tilde{\varepsilon}_i(\omega) - m_{\tilde{\varepsilon}}]/\sigma_{\tilde{\varepsilon}}\}$  and the normalized lower bound of  $\tilde{\varepsilon}_i$ , or  $z_i^l$ , is given by  $z_i^l = [-(q_i^p)^\lambda/\lambda - m_{\tilde{\varepsilon}}]/\sigma_{\tilde{\varepsilon}}$ . An ensemble trace of the postprocessed daily flow  $q_i^o(\omega)$  may then be obtained from

$$q_i^o(\omega) = q_i^p + \varepsilon_i(\omega) = [(q_i^p)^\lambda + \lambda \tilde{\varepsilon}_i(\omega)]^{1/\lambda}, \quad (19)$$

$$\tilde{\varepsilon}_i(\omega) \geq -(q_i^p)^\lambda/\lambda,$$

where  $q_i^o(\omega)$  and  $\varepsilon_i(\omega)$  denote the ensemble realizations of  $q_i^o$  and  $\varepsilon_i$ , respectively.

The error modeling as described above requires estimation of  $\lambda$ ,  $m_{\tilde{\varepsilon}}$ , and  $\sigma_{\tilde{\varepsilon}}^2$  that render the distribution of  $\tilde{\varepsilon}$ , approximately truncated normal given  $\tilde{q}^p$  [see Eq. (12)]. In addition, Eq. (13) assumes that  $\tilde{\varepsilon}$  is approximately homoscedastic with respect to  $\tilde{q}^p$  except near the origin

where the lower bound strongly suppresses variability. In reality, the above assumptions may not be met for all basins. In addition, there may not be enough data points over the tail ends of  $\tilde{q}^p$  to test the conditional truncated normality or homoscedasticity. In this work, we check the reasonableness of the above assumptions by examining for each basin the sample moments, normal quantile plots, histograms, and scatterplots of  $\tilde{\varepsilon}_i$  versus  $\tilde{q}_i^p$ . For those basins showing significant departures from truncated normality or homoscedasticity, we adjusted  $\lambda$  and/or  $\sigma_{\tilde{\varepsilon}}^2$  until the results passed the visual test. To improve the objectivity of the error modeling procedure, additional research is necessary. In this study, we only considered truncated normal for the conditional PDF of  $\tilde{\varepsilon}$ . Other distributions, such as truncated gamma (Chapman 1956), are also possible.

To capture the distributional characteristics of multi-daily flow, it is necessary to model the temporal dependence of  $\tilde{\varepsilon}_i$ . Due to the large number of basins involved, basin-specific modeling of error time series (Box and Jenkins 1976) was beyond the scope of this study. Instead, we modeled the error,  $\tilde{\varepsilon}_i = \tilde{q}_i^o - \tilde{q}_i^p$ , with AR(1) as a first-order approximation for all basins. The use of  $\tilde{q}_i^b$  rather than  $\tilde{q}_i^p$  in the above is motivated by the fact that  $\tilde{q}_i^b$  is not



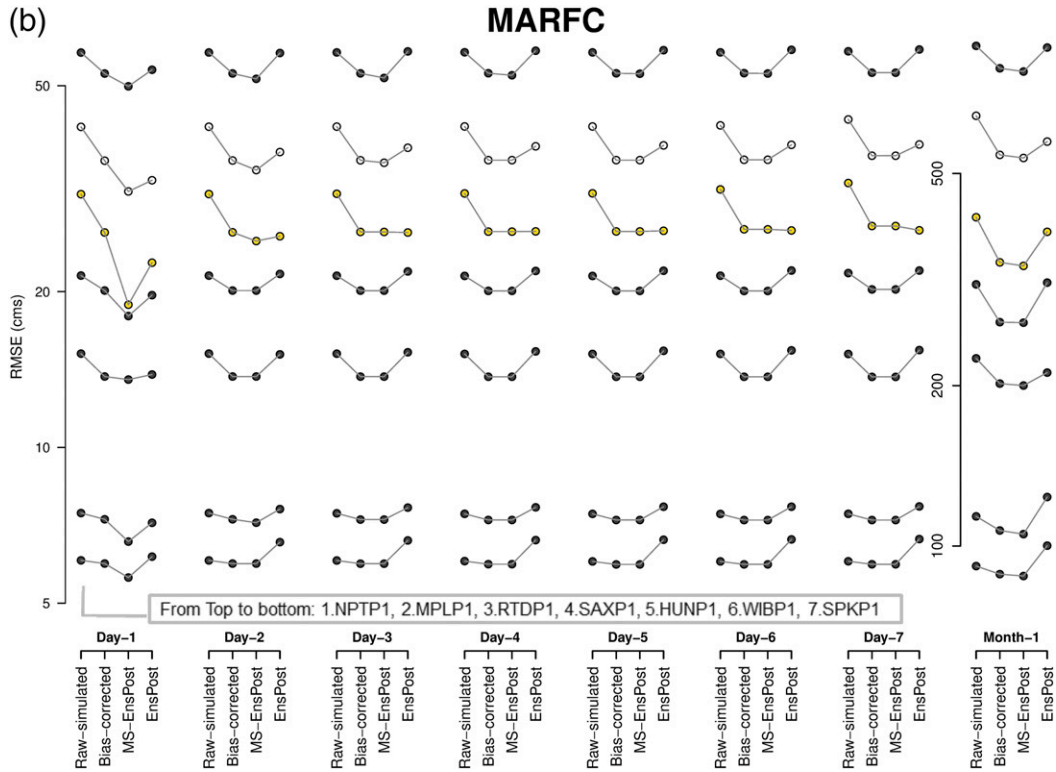


FIG. 5b. As in (a), but for the MARFC basins.

lead-time dependent, and hence greatly simplifies the modeling. The impact of this simplification appears acceptably small for all but a small number of MBRFC basins (see section 4b). To assess the adequacy of AR(1), we carried out structure identification for a small number of basins in the MARFC, NCRFC, NERFC, and NWRFCs’ service areas. The results indicate that the error structures are generally more complex than AR(1), and contain both autoregressive and moving-average components of higher order. This is not very surprising given the very widely varying hydroclimatology of the basins and goodness of the hydrologic modeling. The simplifying choice of AR(1) in this work is additionally motivated by its use in EnsPost which facilitates direct comparison between MS-EnsPost and EnsPost. Though limited in sample size, the above findings suggest that additional improvement in ensemble prediction of multidaily flow may be possible with improved modeling of temporal dependence of prediction error.

d. Evaluation

For comparative evaluation of MS-EnsPost, we carried out both single-valued and ensemble verification of MS-EnsPost via leave-two-years-out cross validation using the Ensemble Verification System (EVS; Brown et al. 2010). The leave-one-year-out cross validation results

are similar. In single-valued verification, we evaluate the raw and bias-corrected predictions, and ensemble mean predictions from postprocessing with EnsPost and MS-EnsPost. In ensemble verification, we evaluate the ensemble predictions from EnsPost and MS-EnsPost, and assess their skill in reference to sample climatology of historical observed flow. In both, we consider predictions of daily flow with lead times of 1–32 days, and of monthly flow with a lead time of one month. For single-valued predictions, we use root-mean-square error (RMSE) as the primary measure of performance:

$$RMSE(k) = \sqrt{\frac{1}{n(k)} \sum_{i=1}^{n(k)} [q_i^p(k) - q_i^o(k)]^2}, \quad (20)$$

where  $q_i^p(k)$  denotes the  $i$ th day- $k$  prediction of daily flow;  $q_i^o(k)$  denotes the verifying observed daily flow; and  $n(k)$  denotes the total number of day- $k$  daily flow predictions. For ensemble predictions, we use the mean continuous ranked probability score (CRPS), its decomposition, and continuous ranked probability skill score (CRPSS) as primary measures (Brown and Seo 2010; Kim et al. 2018). The CRPS represents the integral squared difference between the CDF of the predicted variable  $F_Y(q)$  and that of the verifying observed variable  $F_X(q)$  (i.e., a step function):

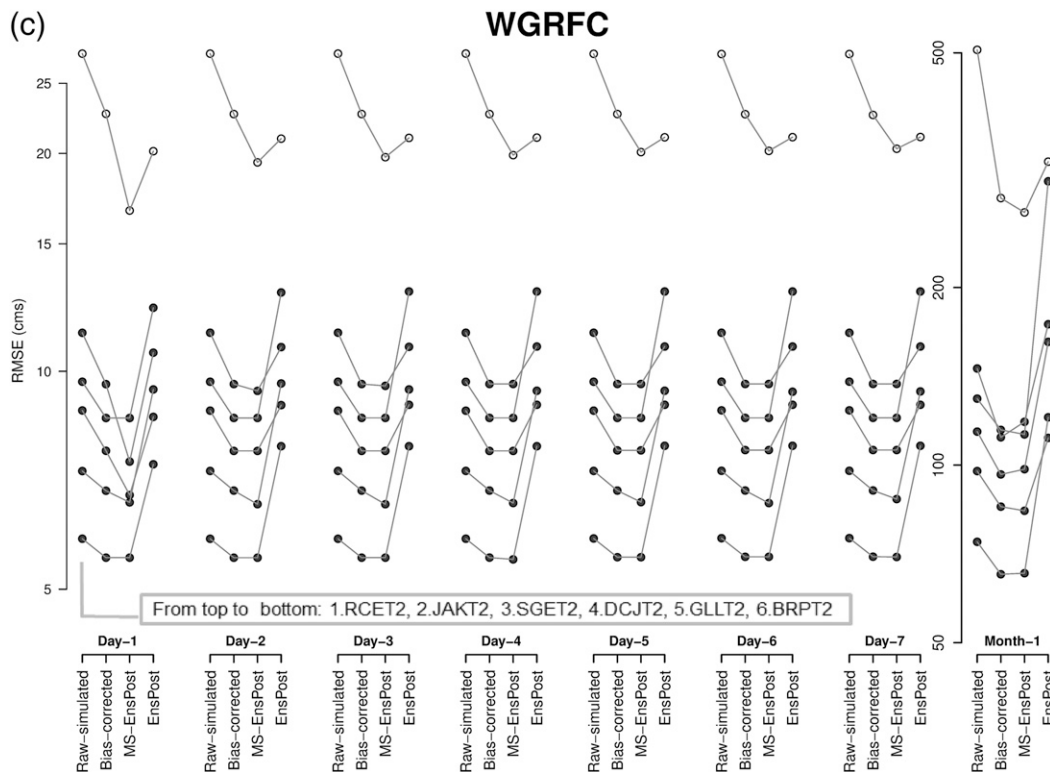


FIG. 5c. As in (a), but for the WGRFC basins.

$$CRPS = \int [F_Y(q) - F_X(q)]^2 dq. \tag{21}$$

The mean CRPS ( $\overline{CRPS}$ ) is the average of the CRPS values from the individual pairs of ensemble forecasts and observations and reflects the overall quality of an ensemble forecasting system (the smaller, the better). In this work, mean CRPS is evaluated following Hersbach (2000). The mean CRPS can be decomposed into mean reliability ( $\overline{REL}$ ), mean resolution ( $\overline{RES}$ ), and mean uncertainty ( $\overline{UNC}$ ), or into mean REL and mean potential CRPS ( $\overline{CRPS}_{POT}$ ) (Hersbach 2000):

$$\overline{CRPS} = \overline{REL} - \overline{RES} + \overline{UNC} = \overline{REL} + \overline{CRPS}_{POT}. \tag{22}$$

Smaller  $\overline{REL}$  indicates more reliable ensembles (desirable) and larger  $\overline{RES}$  means better resolution (desirable). The  $\overline{RES}$  component ( $= \overline{UNC} - \overline{CRPS}_{POT}$ ) is positive if the ensemble forecast is better than the climatological ensemble forecast (Hersbach 2000). The  $\overline{UNC}$  component reflects climatological uncertainties in the observations and does not relate to forecast attributes. The  $\overline{CRPS}_{POT}$  ( $= \overline{CRPS} - \overline{REL}$ ) represents the  $\overline{CRPS}$  for a perfectly reliable forecast (Hersbach 2000).

The CRPSS measures this skill relative to climatology, that is, historical traces of observed daily flow valid at the same time of the year as the subject forecast:

$$\overline{CRPSS} = \frac{\overline{CRPS}_{clim} - \overline{CRPS}}{\overline{CRPS}_{clim}} \tag{23}$$

Perfect and skill-less ensemble forecasts have CRPSS of unity and zero, respectively.

#### 4. Results

This section presents the comparative evaluation results for single-valued and ensemble predictions, and assesses the predictability of streamflow as measured from the ensemble prediction results for different hydroclimatological regions.

##### a. Single-valued streamflow prediction

Figure 5 shows the RMSE of the raw, bias-corrected, MS-EnsPost ensemble mean, and EnsPost ensemble mean streamflow predictions for lead times of 1–7 days and 1 month for the basins in the CBRFC, MARFC, and WGRFCs’ service areas. For similar plots for all other RFCs, the reader is referred to Alizadeh (2019). The above three RFCs are chosen to represent the regions of

the largest, medium, and the most limited predictability among the groups of basins considered in this work. In the figure, we connect the RMSE values for each basin to help assess the relative performance among the four different predictions for each basin. A reduction in RMSE by the bias-corrected prediction over the raw is an indication that significant magnitude-dependent biases exist in the raw model-simulated flow due to parametric or structural errors in the hydrologic models, biases in the forcings, or flow regulations. A reduction in RMSE by the MS-EnsPost ensemble mean prediction over the bias-corrected is due to multiscale regression, and indicates that significant uncertainties exist in the initial conditions of the hydrologic models, or significant hydrologic memory exists in the surface and soil water storages of the basin. Due to the temporal aggregation, the monthly results (rightmost columns in Fig. 5) amplify the relative performance of the bias correction components of MS-EnsPost and EnsPost. The monthly results are hence more reflective of the bias correction operation, which impacts over the entire forecast horizon, than the multiscale regression operation, which impacts over the range of hydrologic memory only. The results for all 139 basins indicate that, MS-EnsPost reduces the RMSE of the raw model predictions of daily flow by 5%–74% and when compared to the EnsPost predictions, by 5%–68%, and that MS-EnsPost is superior to EnsPost for 1-month-ahead streamflow prediction for all basins examined in this work. Below we summarize the main RMSE results for each of the 3 RFCs. The results for the other RFCs and the discussion may be found in Alizadeh (2019).

For CBRFC (Fig. 5a), the MS-EnsPost ensemble mean prediction improves over the raw and EnsPost ensemble mean predictions for all basins. The yellow circles in the figure indicate that the basin has flow regulations that are modeled with CHPS. In general, both bias correction and multiscale regression contributes to the improvement by MS-EnsPost. For a number of basins, the reduction in RMSE due to multiscale regression persists to Day 4 and beyond, a reflection of the longer hydrologic memory present in the Upper Colorado River basin owing to the snowmelt. For MARFC (Fig. 5b), bias correction generally contributes more to the RMSE reduction by MS-EnsPost than multiscale regression. In Fig. 5, the empty circles indicate that the basin has unmodeled regulated flow. The largest improvement by MS-EnsPost over EnsPost is in the day-1 prediction for RTDP1 (third from the top) which is downstream of Raystown Dam on the Raystown Branch of the Juniata River. Overall, the impact of multiscale regression is rather modest and wears off within the first two days of lead time, an indication

that the hydrologic memory in the Juniata River basin in Pennsylvania is relatively short. For the WGRFC basins (Fig. 5c), the improvement by MS-EnsPost over EnsPost is particularly large. These basins are located in the semiarid western part of the Upper Trinity River basin (Kim et al. 2018). As such, they have short memory in surface and soil water storages, and their streams are ephemeral despite relatively large basin size (441 ~ 1764 km<sup>2</sup>). Because EnsPost does not model intermittency of streamflow, its results are particularly poor for the WGRFC basins. MS-EnsPost, on the other hand, is able to address intermittency to a significant extent by aggregating flow which reduces or removes zero flows at sufficiently large temporal scales. Overall, the reduction in RMSE due to multiscale regression is rather short-lived. The monthly results (rightmost panel in Fig. 5c) for JAKT2 and DCJT2 (second and fourth from the top, respectively) are unexpected in that multiscale regression in MS-EnsPost slightly increased RMSE over magnitude-dependent bias correction alone. The above observation indicates that statistical assimilation of observed streamflow up to a month in aggregation scale does not add skill due to the weak hydrologic memory in streamflow in these basins.

#### b. Ensemble streamflow prediction

In this section, we comparatively evaluate the MS-EnsPost ensemble streamflow predictions with the EnsPost. To facilitate comparison for a large number of basins, we use “worm” plots in which the mean CRPS of the MS-EnsPost predictions ( $y$  axis) versus the EnsPost predictions ( $x$  axis) are dot-plotted and connected for lead times of 1–7 days to form a worm for each basin. Figure 6a shows the worm plots in log–log scale for all study basins for each RFC. The lower and upper ends of each worm are associated with day-1 and -7 predictions for that basin, respectively. If MS-EnsPost improves over EnsPost for 7-day-ahead prediction, the worms would stretch downward from the diagonal. The longer the downward stretch is, the larger the improvement by MS-EnsPost over EnsPost. If MS-EnsPost does not improve over EnsPost, the worms would lie along the diagonal. Figure 6b shows the mean CRPS scatterplots of 1-month-ahead MS-EnsPost predictions of monthly flow versus the EnsPost.

Figure 6a shows that, for most basins, MS-EnsPost significantly improves over EnsPost. For most MARFC basins, however, little improvement is seen. For some MBRFC basin, MS-EnsPost performed worse than EnsPost for day-1 and -2 predictions. For RTDP1 of MARFC (the 3rd worm from the top), MS-EnsPost clearly improves over EnsPost. Recall in the single-valued prediction results that MS-EnsPost generally showed significant improvement over EnsPost for regulated

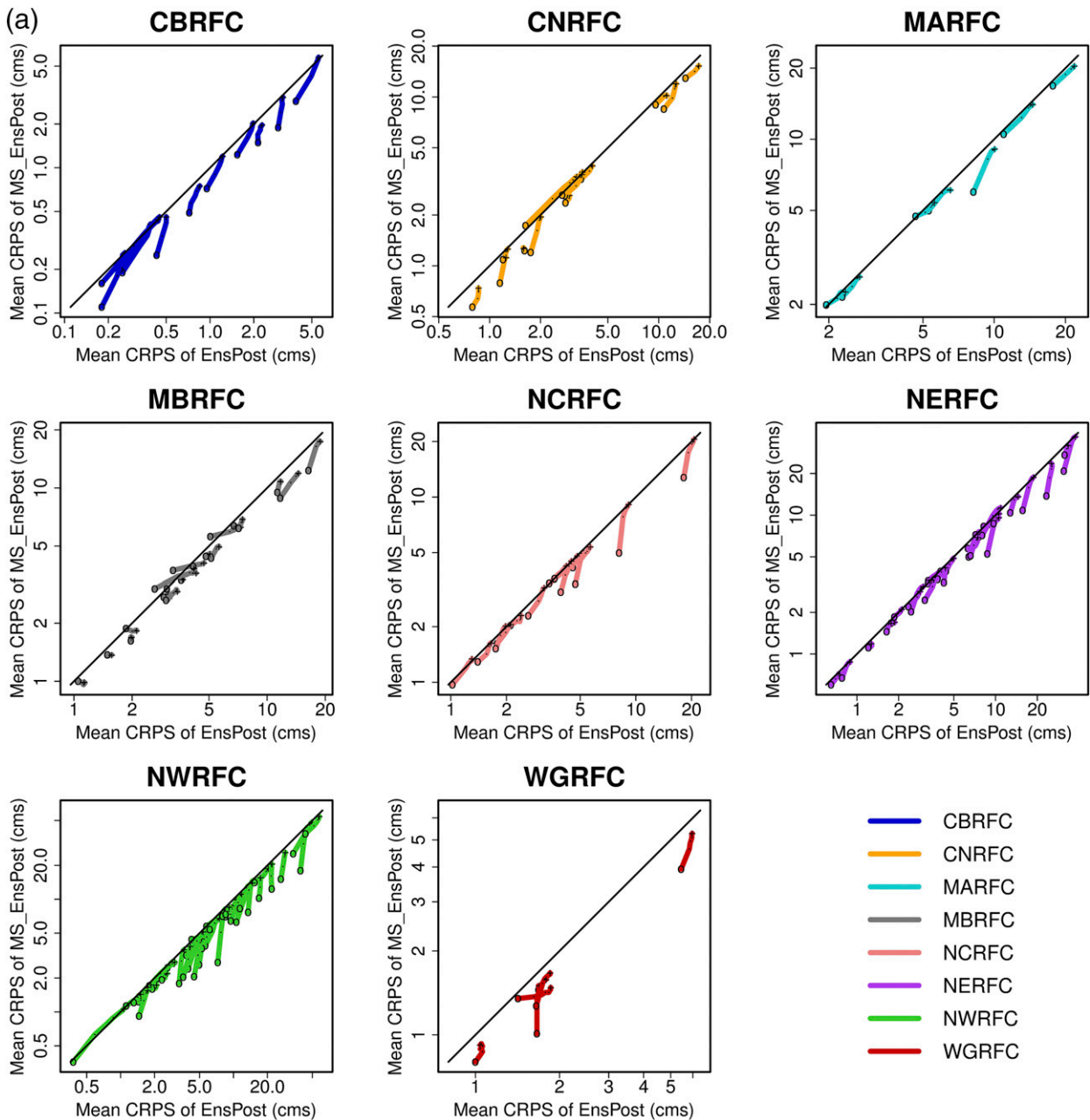


FIG. 6. (a) Worm plots (see text for explanation) of mean CRPS of ensemble predictions of daily flow from MS-EnsPost and EnsPost for lead times of 1–7 days.

flows. The MBRFC results were unexpected in that MS-EnsPost was clearly superior to EnsPost in ensemble mean prediction. A closer examination indicates that, for the four MBRFC basins, the EnsPost ensemble predictions are superior to the MS-EnsPost for the first one or two days of lead time. The above result was traced to reduced reliability (Alizadeh 2019) due to the use of the error statistics of  $\hat{q}_i^b$  rather than  $\hat{q}_i^p$  for simplicity (see section 3c). The above suggests that

lead-time-dependent prescription of the error statistics may be necessary for some basins. The positive impact of bias correction may be seen in Fig. 6b for all MBRFC basins; the 1-month-ahead MS-EnsPost predictions of monthly flow are clearly superior to the EnsPost. Note also in Fig. 6b that the improvement by MS-EnsPost over EnsPost is larger for the smaller MBRFC basins. This is due to the fact that bias correction, rather than multiscale regression, is largely responsible for the

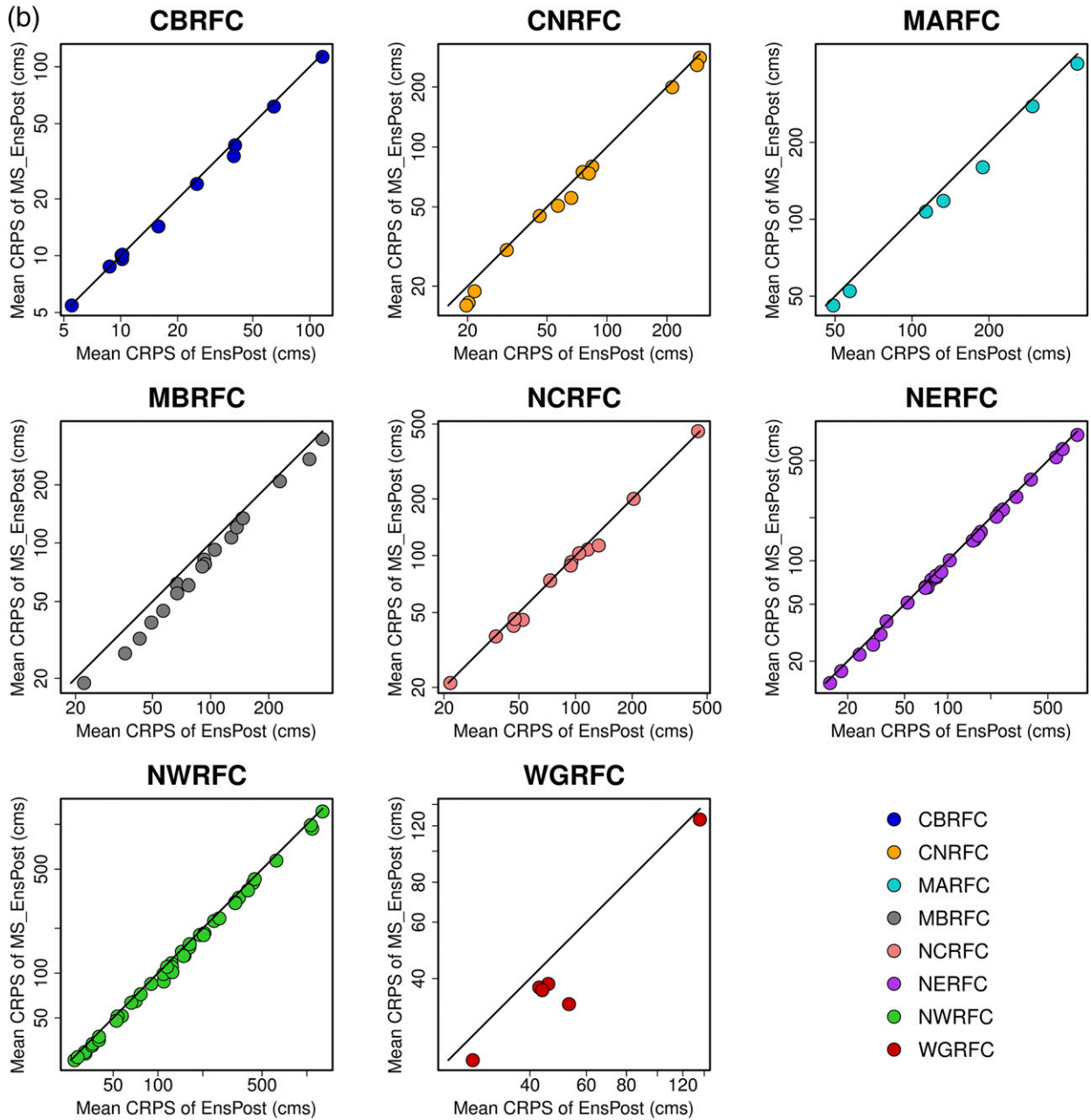


FIG. 6b. As in (a), but for 1-month-ahead predictions of monthly flow.

improvement by MS-EnsPost which produces a large positive cumulative impact for prediction of monthly flow.

Overall, the CBRFC, NCRFC, and NWRFC basins show particularly large improvement by MS-EnsPost over EnsPost. For many CBRFC and NWRFC basins, streamflow is fed by snowmelt, which increases hydrologic memory. The CBRFC and NWRFC results indicate that multiscale regression in MS-EnsPost is able to utilize effectively the predictability present in the model-simulated and observed flows over a range of

temporal scales of aggregation. For the NCRFC basins, on the other hand, the significant improvement by MS-EnsPost is found to be due more to bias correction than multiscale regression (Alizadeh 2019). For the NERFC basins, MS-EnsPost shows significantly larger improvement over EnsPost for larger basins. For the CNRFC basins, MS-EnsPost significantly improves over EnsPost for some basins whereas MS-EnsPost and EnsPost perform similarly for the others. The improvement is found to be generally smaller for the coastal

basins (Alizadeh 2019). For the WGRFC basins, MS-EnsPost significantly improves over EnsPost. It indicates that bias correction and multiscale regression are effective in addressing the flow-magnitude-dependent biases in raw model-predicted flow and intermittency of streamflow in the semiarid region.

Decomposition of the mean CRPS [Eq. (22); see Alizadeh (2019) for examples] indicates that, for most basins, the reduction in mean CRPS by MS-EnsPost over EnsPost is due mostly to improved resolution, rather than improved reliability. This is not very surprising in that the EnsPost uses empirical probability-matching based on NQT whereas MS-EnsPost relies on approximate distribution modeling via the Box–Cox transformation. If the historical record is long enough to model the tails of the distributions with accuracy, one may expect the ensemble traces sampled from the empirically modeled distributions to be more reliable. To scrutinize reliability of MS-EnsPost ensemble predictions, we also examined the reliability diagrams (Brown and Seo 2010; Jolliffe and Stephenson 2012; Wilks 2006) and Brier scores (Brier 1950) for a wide range of thresholds [see Alizadeh (2019) for examples]. They indicate that the MS-EnsPost ensembles are generally as reliable as the EnsPost ensembles for the 90th percentile or larger thresholds, but less so for the 50th percentiles or smaller thresholds. Not surprisingly, Box–Cox transform is not as effective as normal quantile transform particularly for low flows. For flood and water supply forecasting, performance for larger flows is much more important than that for smaller flows. As such, some deterioration in reliability at lower thresholds may not be a large concern in most applications.

### c. Streamflow predictability

The skill in postprocessed ensemble predictions is bounded by the predictability of streamflow explainable by the forcings (Baldwin et al. 2003; Bengtsson and Hodges 2006; Gebregiorgis and Hossain 2011; Li and Ding 2011; Simmons et al. 1995), hydrologic and reservoir models (Hou et al. 2009; Mahanama et al. 2012; Maurer and Lettenmaier 2004; Schlosser and Milly 2002), and statistical assimilation of streamflow via multiscale regression (Bogner et al. 2016; Sharma et al. 2018) used. In this section, we assess and characterize the predictability of streamflow in different hydroclimatological regions based on the ensemble prediction results presented above, and attribute the gains by MS-EnsPost over EnsPost by assessing the predictability through a skill score (Hou et al. 2009; Westra and Sharma 2010). Figure 7 shows the CRPSS of the MS-EnsPost ensemble predictions for all seasons for lead times of 1–32 days. The reference forecast is the sample climatology of

historical observed flow. To assess seasonal variations, we also examined the wet-versus-dry seasonal results. They showed that, except for the CBRFC basins, the CRPSS does not differ much between the two seasons. As such, we only present the combined results which are necessarily more reflective of the wet season. For the CBRFC basins, the CRPSS is significantly lower for the dry season due to the fact that highly persistent low-flow conditions may be predicted very well with climatology. In Fig. 7, the vertical spread in the CRPSS curves represents the variations in predictability of streamflow among the different basins within each RFC's service area. It is readily seen that the CBRFC basins, all of which are in the Upper Colorado River basin, exhibit the smallest variations. The largest variations are observed with the NWRFC basins which encompass the coastal, mountain and intermountain regions of the Pacific Northwest. For each RFC, there are a small number of basins with conspicuously lower CRPSS. They are generally associated with regulated flows which increase hydrologic uncertainty. Because these basins do not represent natural flows, they are treated separately in the analysis below.

MS-EnsPost seeks two effects in the CRPSS results: an increase in the limiting CRPSS,  $\text{CRPSS}(|\infty|)$ , from bias correction at very large lead times and an increase in CRPSS from multiscale regression at shorter lead times above  $\text{CRPSS}(|\infty|)$ . The first and second attributes above are referred to herein as the limiting CRPSS and the hydrologic memory scale (Schlosser and Milly 2002), respectively. The larger the limiting CRPSS, the more skillful the bias-corrected ensemble prediction relative to climatology. The larger the hydrologic memory scale, the larger the increase in CRPSS due to multiscale regression. The hydrologic memory scale  $L_{\text{hm}}$  (days), which represents the predictability of streamflow due to the surface and soil water storages in the basin (Kumar 2011), is defined as

$$L_{\text{hm}} = \int_0^{\infty} \rho_{\text{CRPSS}}(|\tau|) d\tau, \quad (24)$$

where  $\rho_{\text{CRPSS}}(|\tau|)$  denotes the normalized CRPSS at lead time  $\tau$  (days). The normalization renders CRPSS to approach zero at large lead times. One may hence consider  $\rho_{\text{CRPSS}}(|\tau|)$  as a correlogram with nugget effect (Norouzi et al. 2018).

Figure 8 shows the resulting pairs of  $L_{\text{hm}}$  and  $\text{CRPSS}(|\infty|)$  for all basins in each RFC as obtained from the MS-EnsPost and EnsPost predictions. For each basin, an arrow connects the EnsPost result to the matching MS-EnsPost result. If MS-EnsPost increases the limiting CRPSS, the arrow would point upward.

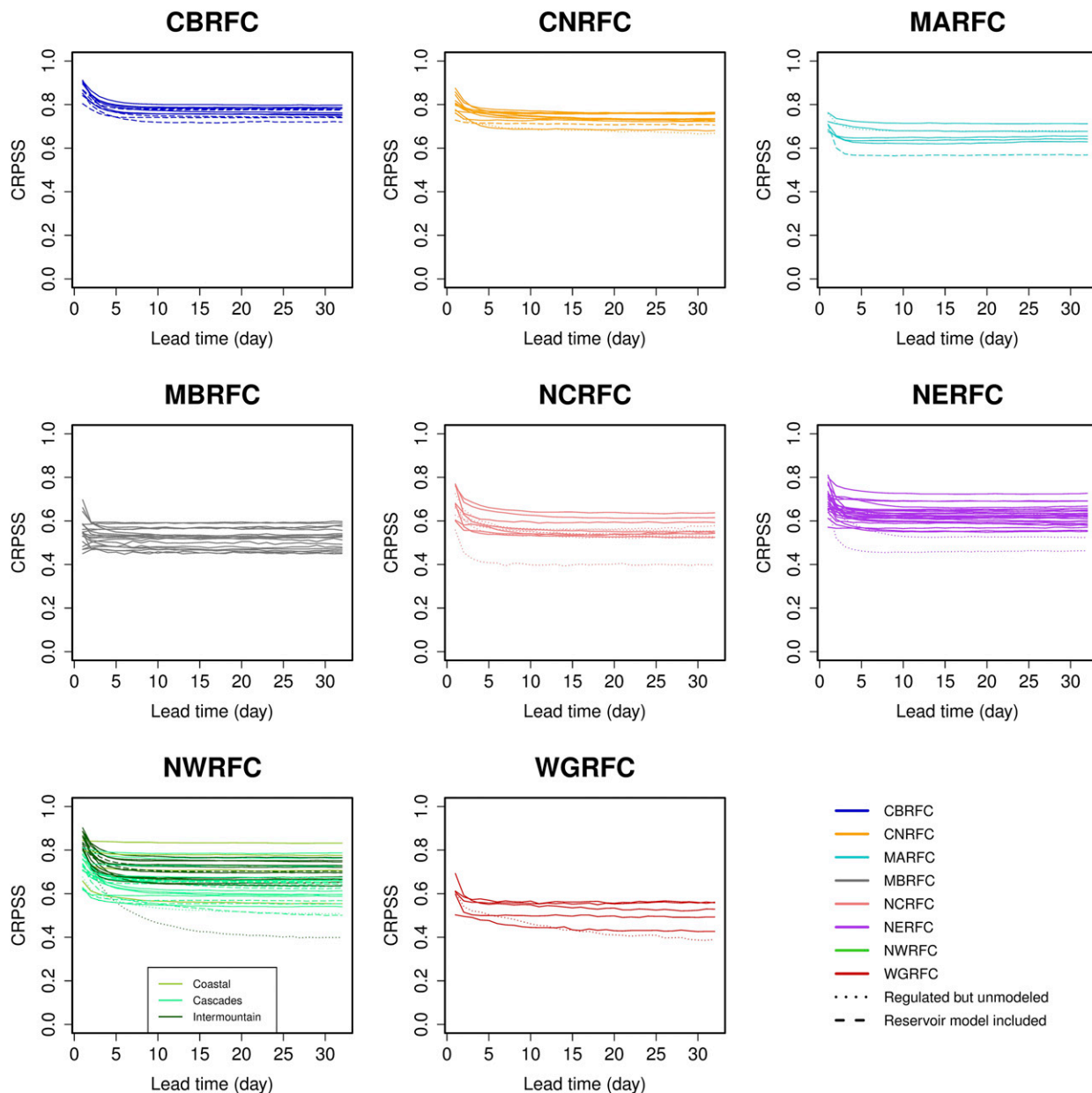


FIG. 7. CRPSS of ensemble predictions of daily flow from MS-EnsPost vs lead time. Reference is sample climatology of historical observed flow.

If MS-EnsPost increases the hydrologic memory scale, the arrow would point to the right. The longer the arrow is, the larger the improvement or deterioration is. Accordingly, lengthy arrows pointing in the upper-right direction would indicate MS-EnsPost clearly improving over EnsPost. It is seen that, for a number of basins, MS-EnsPost improves limiting CRPSS but reduces the hydrologic memory scale, resulting in arrows pointing in the upper-left direction. For these basins, the limiting CRPSS for EnsPost is significantly smaller than that for MS-EnsPost whereas the CRPSS for day-1 prediction is

as large as or larger than that for MS-EnsPost. In such a case, EnsPost prediction is likely to yield a larger hydrologic memory than the MS-EnsPost. Such “inflated” hydrologic memory for EnsPost is an artifact of significantly smaller limiting CRPSS and hence, by itself, is not a very useful indicator of predictive skill. Accordingly, one may consider MS-EnsPost inferior to EnsPost only if the arrow is pointing in the lower-left direction.

Figure 8 shows that MS-EnsPost outperforms or performs comparably to EnsPost for all basins, increases limiting CRPSS for almost all basins, and provides significant

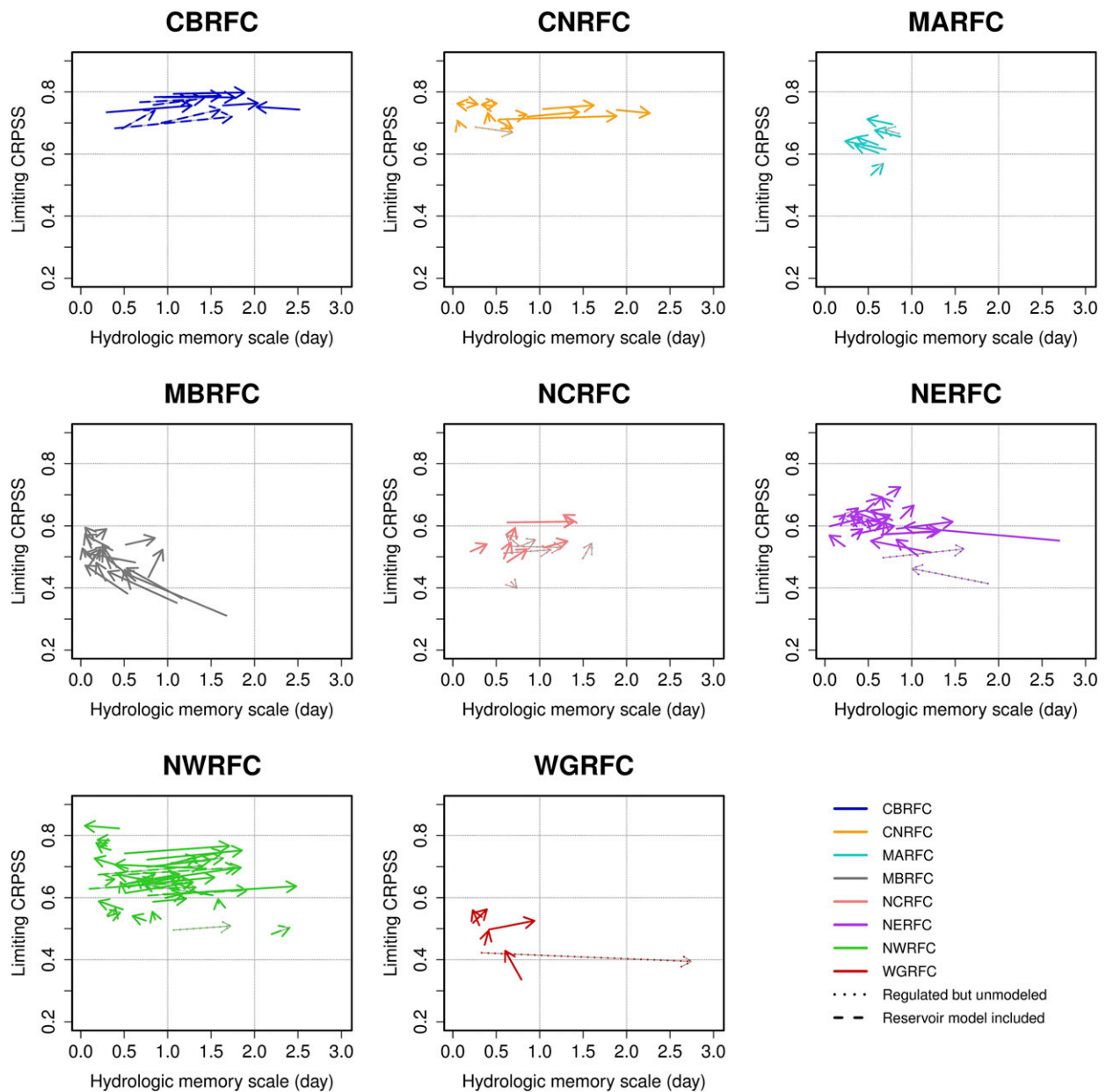


FIG. 8. Changes in limiting CRPSS and hydrologic memory scale from those of EnsPost to those of MS-EnsPost (see text for explanation).

additional skill via multiscale regression particularly for the CBRFC, CNRFC, NCRFC, and NWRFC basins. From Fig. 8, a number of postulations may also be made. The significant increase in  $L_{hm}$  by MS-EnsPost for the CBRFC, CNRFC, NCRFC, and NWRFC basins suggests that there exists significant multiscale hydrologic memory to be exploited for operational hydrologic forecasting via data assimilation. The significant increase in  $CRPSS(|\infty|)$  by MS-EnsPost for the MBRFC and NERFC basins suggests that there may exist significant room for improving calibration, hydrologic modeling, and input forcings to reduce hydrologic uncertainties. The WGRFC basin

results, on the other hand, suggest limited room for improving predictive skill within the existing modeling and forecasting process, and point to improving model physics as well as soil moisture sensing and its assimilation.

The relative importance of  $CRPSS(|\infty|)$  versus  $L_{hm}$  in assessing predictability necessarily varies with the application at hand. For long-range predictions,  $CRPSS(|\infty|)$  would be more important whereas, for short-range predictions,  $L_{hm}$  may be just as important. Hence, it is not readily possible to translate uniquely the two summary attributes into a single measure. One may consider, however, the relative positions of the  $[L_{hm}, CRPSS(|\infty|)]$  pairs



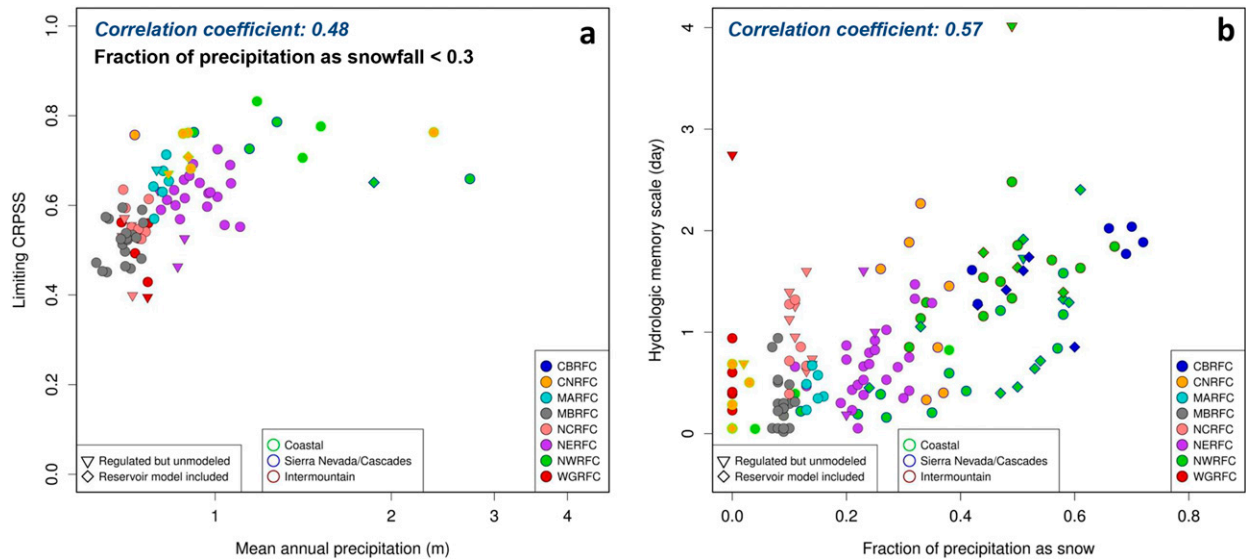


FIG. 9. (a) Limiting CRPSS vs mean annual precipitation for non-snow-driven basins and (b) hydrologic memory scale vs fraction of precipitation as snowfall.

for MS-EnsPost (i.e., the tips of the arrows) within the  $x$ - $y$  plot in Fig. 8, and approximately rank the groups of basins in different RFCs in terms of the collective strength of predictability as measured through MS-EnsPost. The figure indicates that the CBRFC, CNRFC, NWRFC basins are the most predictable, followed by the NERFC, MARFC, and NCRFC basins, and that the MBRFC and WGRFC basins are the least predictable. The above order reflects what may be garnered visually from Fig. 7, and generally follows the decreasing order of the fraction of precipitation as snowfall  $f_s$  and mean annual precipitation (see Fig. 1). To illustrate, Fig. 9 shows CRPSS( $|\infty$ ) versus annual mean precipitation for basins with  $f_s < 0.3$  (Fig. 9a) and  $L_{hm}$  versus  $f_s$  for all basins (Fig. 9b). Though the scatters are large, CRPSS( $|\infty$ ) for non-snow-dominated basins relates well with mean annual precipitation except for the few very wet coastal basins, and  $L_{hm}$  relate positively with  $f_s$  for all basins.

### 5. Conclusions and future research recommendations

We describe a novel multiscale postprocessor for ensemble streamflow prediction, MS-EnsPost, and compare with the existing postprocessor, EnsPost, in the NWS’s HEFS for 139 basins in the service areas of eight RFCs in the CONUS. MS-EnsPost uses data-driven correction of magnitude-dependent biases in model-simulated flow, multiscale regression to utilize observed and simulated flows over a range of temporal scales of aggregation, and ensemble generation based on

parsimonious error modeling. Comparative evaluation of raw and bias-corrected single-valued predictions and ensemble mean predictions from MS-EnsPost and EnsPost shows that MS-EnsPost reduces the RMSE of day-1 to day-7 predictions of daily flow from EnsPost on average by 28%, and that, for most basins, the improvement is due to both bias correction and multiscale regression. Comparative evaluation of ensemble predictions from MS-EnsPost and EnsPost shows that MS-EnsPost reduces the mean CRPS of day-1 to day-7 predictions of daily flow from EnsPost on average by 18%, and that the improvement is due mostly to improved resolution than reliability. Examination of the CRPSS of ensemble predictions indicate that, for most basins, the improvement by MS-EnsPost over EnsPost is due to both magnitude-dependent bias correction and multiscale regression, which utilizes hydrologic memory more effectively. Comparison of the CRPSS with hydroclimatic indices indicates that the skill in ensemble streamflow predictions from MS-EnsPost is modulated by the fraction of precipitation as snowfall and, for non-snowfall-driven basins, mean annual precipitation.

In addition to improving performance, the development of MS-EnsPost is motivated for reducing data requirement. Streamflow responses have changed or are changing significantly in many parts of the world due to urbanization and climate change (Milly et al. 2008). Changing conditions force statistical post processors a difficult trade-off between accounting for nonstationarities by dividing the period of record or modeling trends, which would significantly increase sampling uncertainties, versus keeping sampling uncertainties smaller but at the expense

of introducing biases due to nonstationarities. Owing to the parsimony, one may expect MS-EnsPost to require significantly less data than EnsPost. We are currently assessing the data requirement for MS-EnsPost for possible application under nonstationarity, and the results will be reported in the near future.

*Acknowledgments.* This material is based upon work supported by the NOAA Climate Program Office under Grant NA15OAR4310109, the NWS COMET Program under UCAR Subaward No. SUBAWD000020, and the NSF under Grant CyberSEES-1442735. These supports are gratefully acknowledged. We thank John Lhotak of CBRFC, Brett Whitin and Ark Henkel of CNRFC, Seann Reed of MARFC, Lisa Holts of MBRFC, Andrea Holz of NCRFC, Erick Boehmler of NERFC, Brad Gillies of NWRFC, Andrew Philpott, Frank Bell, Paul McKee, Kris Lander and Mark Null of WGRFC for providing data and help during the course of this work.

#### REFERENCES

- Adams, T. E., III, 2016: Flood forecasting in the United States NOAA/National Weather Service. *Flood Forecasting: A Global Perspective*, Academic Press, 249–310, <https://doi.org/10.1016/B978-0-12-801884-2.00010-4>.
- Ajami, N. K., Q. Duan, and S. Sorooshian, 2007: An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour. Res.*, **43**, W01403, <https://doi.org/10.1029/2005WR004745>.
- Alizadeh, B., 2019: Improving post processing of ensemble streamflow forecast for short-to-long ranges: a multiscale approach, PhD dissertation, Dept. of Civil Engineering, The University of Texas at Arlington, 125 pp., <https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/28663/ALIZADEH-DISSERTATION-2019.pdf?sequence=1>.
- Anderson, E. A., 1973: National Weather Service River Forecast System—Snow accumulation and ablation model. NOAA Tech. Memo. NWS HYDRO-17, 87 pp., <https://www.wcc.nrcs.usda.gov/ftpref/wntsc/H&H/snow/AndersonHYDRO17.pdf>.
- Baldwin, M. P., D. B. Stephenson, D. W. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640, <https://doi.org/10.1126/science.1087143>.
- Bengtsson, L., and K. I. Hodges, 2006: A note on atmospheric predictability. *Tellus*, **58A**, 154–157, <https://doi.org/10.1111/j.1600-0870.2006.00156.x>.
- Bennett, C., R. Stewart, and J. Lu, 2014: Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, **7**, 2938–2960, <https://doi.org/10.3390/en7052938>.
- Berghuijs, W. R., and R. A. Woods, 2016: A simple framework to quantitatively describe monthly precipitation and temperature climatology. *Int. J. Climatol.*, **36**, 3161–3174, <https://doi.org/10.1002/joc.4544>.
- , M. Sivapalan, R. A. Woods, and H. H. Savenije, 2014: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resour. Res.*, **50**, 5638–5661, <https://doi.org/10.1002/2014WR015692>.
- Björnar Bremnes, J., 2004: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. *Mon. Wea. Rev.*, **132**, 338–347, [https://doi.org/10.1175/1520-0493\(2004\)132<0338:PFOPIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2).
- Blöschl, G., and M. Sivapalan, 1995: Scale issues in hydrological modelling: a review. *Hydrol. Processes*, **9**, 251–290, <https://doi.org/10.1002/hyp.3360090305>.
- Bogner, K., K. Liechti, and M. Zappa, 2016: Post-processing of stream flows in Switzerland with an emphasis on low flows and floods. *Water*, **8**, 115, <https://doi.org/10.3390/w8040115>.
- Borgomeo, E., J. W. Hall, F. Fung, G. Watts, K. Colquhoun, and C. Lambert, 2014: Risk-based water resources planning: Incorporating probabilistic nonstationary climate uncertainties. *Water Resour. Res.*, **50**, 6850–6873, <https://doi.org/10.1002/2014WR015558>.
- Boucher, M. A., L. Perreault, F. Anctil, and A. C. Favre, 2015: Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. *Hydrol. Processes*, **29**, 1141–1155, <https://doi.org/10.1002/hyp.10234>.
- Box, G. E., and D. Cox, 1964: An analysis of transformations. *J. Roy. Stat. Soc. B*, **26**, 211–252.
- , and G. M. Jenkins, 1976: *Time Series Analysis: Forecasting and Control*. Holden Day, 575 pp.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Brown, J. D., and D.-J. Seo, 2010: A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *J. Hydrometeorol.*, **11**, 642–665, <https://doi.org/10.1175/2009JHM1188.1>.
- , J. Demargne, D.-J. Seo, and Y. Liu, 2010: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environ. Modell. Software*, **25**, 854–872, <https://doi.org/10.1016/j.envsoft.2010.01.009>.
- Budyko, M. I., D. H. Miller, and D. H. Miller, 1974: *Climate and Life*. D. H. Miller, Ed., International Geophysics Series, Vol. 18, Academic Press, 508 pp.
- Burnash, R. J., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual models for digital computers. Joint Federal and State River Forecast Center, U.S. National Weather Service, and California Department of Water Resources Tech. Rep., 204 pp.
- Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen, 2004: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J. Hydrol.*, **298**, 242–266, <https://doi.org/10.1016/j.jhydrol.2004.03.042>.
- Carlu, M., F. Ginelli, V. Lucarini, and A. Politi, 2019: Lyapunov analysis of multiscale dynamics: the slow bundle of the two-scale Lorenz 96 model. *Nonlinear Processes Geophys.*, **26**, 73–89, <https://doi.org/10.5194/npg-26-73-2019>.
- Chapman, D. G., 1956: Estimating the parameters of a truncated gamma distribution. *Ann. Math. Stat.*, **27**, 498–506, <https://doi.org/10.1214/aoms/1177728272>.
- Chow, V. T., D. R. Maidment, and L. W. Mays, 1988: Unit hydrograph. *Applied Hydrology*, McGraw-Hill, 100–118.
- Cloke, H., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, <https://doi.org/10.1016/j.jhydrol.2009.06.005>.
- Coccia, G., and E. Todini, 2011: Recent developments in predictive uncertainty assessment based on the model conditional processor

- approach. *Hydrol. Earth Syst. Sci.*, **15**, 3253–3274, <https://doi.org/10.5194/hess-15-3253-2011>.
- Damon, J., and S. Guillas, 2002: The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**, 759–774, <https://doi.org/10.1002/env.527>.
- Demargne, J., and Coauthors, 2014: The science of NOAA's operational hydrologic ensemble forecast service. *Bull. Amer. Meteor. Soc.*, **95**, 79–98, <https://doi.org/10.1175/BAMS-D-12-00081.1>.
- Demeritt, D., S. Nobert, H. Cloke, and F. Pappenberger, 2010: Challenges in communicating and using ensembles in operational flood forecasting. *Meteor. Appl.*, **17**, 209–222, <https://doi.org/10.1002/met.194>.
- Deutsch, R., 1965: *Estimation Theory*. Prentice-Hall, 269 pp.
- Doherty, J., and D. Welter, 2010: A short exploration of structural noise. *Water Resour. Res.*, **46**, W05525, <https://doi.org/10.1029/2009WR008377>.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian, 2007: Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.*, **30**, 1371–1386, <https://doi.org/10.1016/j.advwatres.2006.11.014>.
- Dunne, T., and R. D. Black, 1970: Partial area contributions to storm runoff in a small New England watershed. *Water Resour. Res.*, **6**, 1296–1311, <https://doi.org/10.1029/WR006i005p01296>.
- Engman, E. T., and A. S. Rogowski, 1974: A partial area model for storm flow synthesis. *Water Resour. Res.*, **10**, 464–472, <https://doi.org/10.1029/WR010i003p00464>.
- Erickson, M., 1996: Medium-range prediction of PoP and Max/Min in the era of ensemble model output. Preprints, *Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J35–J38.
- Fedora, M., and R. Beschta, 1989: Storm runoff simulation using an antecedent precipitation index (API) model. *J. Hydrol.*, **112**, 121–133, [https://doi.org/10.1016/0022-1694\(89\)90184-4](https://doi.org/10.1016/0022-1694(89)90184-4).
- Fowler, K., M. Peel, A. Western, and L. Zhang, 2018: Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resour. Res.*, **54**, 3392–3408, <https://doi.org/10.1029/2017WR022466>.
- Freer, J., K. Beven, and B. Ambrose, 1996: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resour. Res.*, **32**, 2161–2173, <https://doi.org/10.1029/95WR03723>.
- Freeze, R. A., 1972: Role of subsurface flow in generating surface runoff: 2. Upstream source areas. *Water Resour. Res.*, **8**, 1272–1283, <https://doi.org/10.1029/WR008i005p01272>.
- Gan, T. Y., E. M. Dlamini, and G. F. Biftu, 1997: Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *J. Hydrol.*, **192**, 81–103, [https://doi.org/10.1016/S0022-1694\(96\)03114-9](https://doi.org/10.1016/S0022-1694(96)03114-9).
- Gebregiorgis, A., and F. Hossain, 2011: How much can a priori hydrologic model predictability help in optimal merging of satellite precipitation products? *J. Hydrometeorol.*, **12**, 1287–1298, <https://doi.org/10.1175/JHM-D-10-05023.1>.
- Gelb, A., 1974: *Applied Optimal Estimation*. MIT Press, 374 pp.
- Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *J. Hydrol.*, **298**, 222–241, <https://doi.org/10.1016/j.jhydrol.2004.03.037>.
- Gijssbers, P., L. Cajina, C. Dietz, J. Roe, and E. Welles, 2009: CHPS - An NWS development to enter the interoperability era. *2009 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract IN11A-1041.
- Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.
- Greene, W. H., 2003: *Econometric Analysis*. 5th ed. Prentice Hall, 1026 pp.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye, 2012: Towards a comprehensive assessment of model structural adequacy. *Water Resour. Res.*, **48**, W08301, <https://doi.org/10.1029/2011WR011044>.
- Hall, J. W., and E. Borgomeo, 2013: Risk-based principles for defining and managing water security. *Philos. Trans. Roy. Soc. London*, **371A**, 20120407, <https://doi.org/10.1098/rsta.2012.0407>.
- , and Coauthors, 2020: Risk-based water resources planning in practice: a blueprint for the water industry in England. *Water Environ. J.*, <https://doi.org/10.1111/wej.12479>, in press.
- Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.*, **126**, 711–724, [https://doi.org/10.1175/1520-0493\(1998\)126<0711:EOEREP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0711:EOEREP>2.0.CO;2).
- , J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, [https://doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2).
- Hartman, R., M. Fresch, and E. Wells, 2015: National Weather Service (NWS) implementation of the Hydrologic Ensemble Forecast Service. *2015 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H52A-01.
- Hashino, T., A. A. Bradley, and S. S. Schwartz, 2002: Verification of probabilistic streamflow forecasts. IIHR Rep. 427, University of Iowa, 125 pp., <https://www.iuhr.uiowa.edu/wp-content/uploads/2013/06/IIHR427.pdf>.
- , —, and —, 2006: Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol. Earth Syst. Sci.*, **11**, 939–950, <https://doi.org/10.5194/hess-11-939-2007>.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Horton, R. E., 1933: The role of infiltration in the hydrologic cycle. *Trans. Amer. Geophys. Union*, **14**, 446–460, <https://doi.org/10.1029/TR014i001p00446>.
- Hou, D., K. Mitchell, Z. Toth, D. Lohmann, and H. Wei, 2009: The effect of large-scale atmospheric uncertainty on streamflow predictability. *J. Hydrometeorol.*, **10**, 717–733, <https://doi.org/10.1175/2008JHM1064.1>.
- Jolliffe, I. T., and D. B. Stephenson, 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, 292 pp.
- Kim, S., and Coauthors, 2016: Integrating ensemble forecasts of precipitation and streamflow into decision support for reservoir operations in north central Texas. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H53I-08.
- , and Coauthors, 2018: Assessing the skill of medium-range ensemble precipitation and streamflow forecasts from the Hydrologic Ensemble Forecast Service (HEFS) for the Upper Trinity River Basin in North Texas. *J. Hydrometeorol.*, **19**, 1467–1483, <https://doi.org/10.1175/JHM-D-18-0027.1>.
- Kim, S. M., B. L. Benham, K. M. Brannan, R. W. Zeckoski, and J. Doherty, 2007: Comparison of hydrologic calibration of HSPF using automatic and manual methods. *Water Resour. Res.*, **43**, W01402, <https://doi.org/10.1029/2006WR004883>.
- Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in snowfall versus rainfall in the western United States. *J. Climate*, **19**, 4545–4559, <https://doi.org/10.1175/JCLI3850.1>.

- Koenker, R., and G. Bassett, 1978: Regression quantiles. *Econometrica*, **46**, 33–50, <https://doi.org/10.2307/1913643>.
- Krause, P., D. Boyle, and F. Bäse, 2005: Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.*, **5**, 89–97, <https://doi.org/10.5194/adgeo-5-89-2005>.
- Krzysztofowicz, R., 1999: Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.*, **35**, 2739–2750, <https://doi.org/10.1029/1999WR900099>.
- , and K. S. Kelly, 2000: Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.*, **36**, 3265–3277, <https://doi.org/10.1029/2000WR900108>.
- , and H. D. Herr, 2001: Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *J. Hydrol.*, **249**, 46–68, [https://doi.org/10.1016/S0022-1694\(01\)00412-7](https://doi.org/10.1016/S0022-1694(01)00412-7).
- Kumar, P., 2011: Typology of hydrologic predictability. *Water Resour. Res.*, **47**, W00H05, <https://doi.org/10.1029/2010WR009769>.
- Lee, H., Y. Liu, J. Brown, J. Ward, A. Maestre, M. A. Fresch, H. Herr, and E. Wells, 2018: Validation of ensemble streamflow forecasts from the Hydrologic Ensemble Forecast Service (HEFS). *2018 Fall Meeting*, Washington, D.C., Amer. Geophys. Union, H31A-05.
- Legates, D. R., and G. J. McCabe Jr., 1999: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, **35**, 233–241, <https://doi.org/10.1029/1998WR900018>.
- Li, J., and R. Ding, 2011: Temporal–spatial distribution of atmospheric predictability limit by local dynamical analogs. *Mon. Wea. Rev.*, **139**, 3265–3283, <https://doi.org/10.1175/MWR-D-10-05020.1>.
- Li, M., Q. J. Wang, J. C. Bennett, and D. E. Robertson, 2016: Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting. *Hydrol. Earth Syst. Sci.*, **20**, 3561–3579, <https://doi.org/10.5194/hess-20-3561-2016>.
- Li, W., Q. Duan, C. Miao, A. Ye, W. Gong, and Z. Di, 2017: A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdiscip. Rev.: Water*, **4**, e1246, <https://doi.org/10.1002/wat2.1246>.
- Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara, 2015: A multiscale variational data assimilation scheme: Formulation and illustration. *Mon. Wea. Rev.*, **143**, 3804–3822, <https://doi.org/10.1175/MWR-D-14-00384.1>.
- Limon, R., 2019: Improving multi-reservoir water supply system operation using ensemble forecasting and global sensitivity analysis. Ph.D. dissertation, The University of Texas at Arlington, 164 pp., <http://hdl.handle.net/10106/28115>.
- Loague, K., and J. E. VanderKwaak, 2004: Physics-based hydrologic response simulation: Platinum bridge, 1958 Edsel, or useful tool. *Hydrol. Processes*, **18**, 2949–2956, <https://doi.org/10.1002/hyp.5737>.
- Madadgar, S., H. Moradkhani, and D. Garen, 2014: Towards improved post-processing of hydrologic forecast ensembles. *Hydrol. Processes*, **28**, 104–122, <https://doi.org/10.1002/hyp.9562>.
- Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle, 2012: Soil moisture, snow, and seasonal streamflow forecasts in the United States. *J. Hydrometeorol.*, **13**, 189–203, <https://doi.org/10.1175/JHM-D-11-046.1>.
- Maurer, E. P., and D. P. Lettenmaier, 2004: Potential effects of long-lead hydrologic predictability on Missouri River mainstem reservoirs. *J. Climate*, **17**, 174–186, [https://doi.org/10.1175/1520-0442\(2004\)017<0174:PEOLHP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0174:PEOLHP>2.0.CO;2).
- Mendoza, P., A. Wood, E. Clark, B. Nijssen, M. Clark, M.-H. Ramos, and N. Voisin, 2016: Improving medium-range ensemble streamflow forecasts through statistical post-processing. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H51F-1547.
- Milly, P. C., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319**, 573–574, <https://doi.org/10.1126/science.1151915>.
- Mizukami, N., and Coauthors, 2017: Towards seamless large-domain parameter estimation for hydrologic models. *Water Resour. Res.*, **53**, 8020–8040, <https://doi.org/10.1002/2017WR020401>.
- Montanari, A., and A. Brath, 2004: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.*, **40**, W01106, <https://doi.org/10.1029/2003WR002540>.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- NRC, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. The National Academies Press, 124 pp.
- NWS, 2008a: Joint Reservoir Regulation (RES-J) Model. Accessed 30 January 2017, [http://www.nws.noaa.gov/oh/hrl/general/chps/Models/Joint\\_Reservoir\\_Regulation.pdf](http://www.nws.noaa.gov/oh/hrl/general/chps/Models/Joint_Reservoir_Regulation.pdf).
- , 2008b: Single Reservoir Regulation (RES-SNGL) Model. Accessed 30 January 2017, [http://www.nws.noaa.gov/oh/hrl/general/chps/Models/Single\\_Reservoir\\_Regulation.pdf](http://www.nws.noaa.gov/oh/hrl/general/chps/Models/Single_Reservoir_Regulation.pdf).
- , 2017a: Meteorological Ensemble Forecast Processor (MEFP) User’s Manual. 163 pp., accessed 27 June 2019, [https://vlab.ncep.noaa.gov/documents/207461/1893026/MEFP\\_Users\\_Manual.pdf](https://vlab.ncep.noaa.gov/documents/207461/1893026/MEFP_Users_Manual.pdf).
- , 2017b: Ensemble Postprocessor (EnsPost) User’s Manual. 69 pp., accessed 27 June 2019, [https://vlab.ncep.noaa.gov/documents/207461/1893026/EnsPost\\_Users\\_Manual.pdf](https://vlab.ncep.noaa.gov/documents/207461/1893026/EnsPost_Users_Manual.pdf).
- Norouzi, A., H. Habibi, B. Nazari, S. J. Noh, D.-J. Seo, and Y. Zhang, 2019: Toward parsimonious modeling of frequency of areal runoff from heavy-to-extreme precipitation in large urban areas under changing conditions: A derived moment approach. *Stoch. Environ. Res. Risk Assess.*, **33**, 1263–1281, <https://doi.org/10.1007/s00477-019-01698-8>.
- Piani, C., J. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, <https://doi.org/10.1175/MWR2906.1>.
- Rakovec, O., and Coauthors, 2016: Multiscale and multivariate evaluation of water fluxes and states over European river basins. *J. Hydrometeorol.*, **17**, 287–307, <https://doi.org/10.1175/JHM-D-15-0054.1>.
- Regonda, S., and D.-J. Seo, 2008: Statistical post processing streamflow ensembles to improve reliability over a wide range of time scales. *2nd CPPA PIs Meeting*, Silver Spring, MD, NOAA, [http://vintage.joss.ucar.edu/joss\\_psg/meetings/Meetings\\_2008/CPPA/poster/RegondaPoster.pdf](http://vintage.joss.ucar.edu/joss_psg/meetings/Meetings_2008/CPPA/poster/RegondaPoster.pdf).
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks, 2010: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.*, **46**, W05521, <https://doi.org/10.1029/2009WR008328>.

- Robert, C. P., 1995: Simulation of truncated normal variables. *Stat. Comput.*, **5**, 121–125, <https://doi.org/10.1007/BF00143942>.
- Roe, J., and Coauthors, 2010: NOAA's Community Hydrologic Prediction System. *Second Joint Federal Interagency Conf.*, Las Vegas, NV, Advisory Committee on Water Information, 12 pp., <https://training.weather.gov/nwstc/CHPS/roe.pdf>.
- Roulston, M. S., and L. A. Smith, 2003: Combining dynamical and statistical ensembles. *Tellus*, **55A**, 16–30, <https://doi.org/10.3402/tellusa.v55i1.12082>.
- Samaniego, L., R. Kumar, and S. Attinger, 2010: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.*, **46**, W05523, <https://doi.org/10.1029/2008WR007327>.
- Sankarasubramanian, A., and R. M. Vogel, 2003: Hydroclimatology of the continental United States. *Geophys. Res. Lett.*, **30**, 1363, <https://doi.org/10.1029/2002GL015937>.
- Schaake, J., J. Demargne, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, and D.-J. Seo, 2007a: Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrol. Earth Syst. Sci.*, **4**, 655–717, <https://doi.org/10.5194/hessd-4-655-2007>.
- Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark, 2007b: HEPEX: The Hydrological Ensemble Prediction Experiment. *Bull. Amer. Meteor. Soc.*, **88**, 1541–1548, <https://doi.org/10.1175/BAMS-88-10-1541>.
- Schlosser, C. A., and P. C. D. Milly, 2002: A model-based investigation of soil moisture predictability and associated climate predictability. *J. Hydrometeorol.*, **3**, 483–501, [https://doi.org/10.1175/1525-7541\(2002\)003<0483:AMBIOS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0483:AMBIOS>2.0.CO;2).
- Schweppe, F. C., 1973: *Uncertain Dynamic Systems*. Prentice Hall, 563 pp.
- Seo, D.-J., H. Herr, and J. Schaake, 2006: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol. Earth Syst. Sci. Discuss.*, **3**, 1987–2035, <https://doi.org/10.5194/hessd-3-1987-2006>.
- Sharma, S., R. Siddique, S. Reed, P. Ahnert, P. A. Mendoza Zúñiga, and A. Mejia, 2018: Relative effects of statistical preprocessing and postprocessing on a regional hydrological ensemble prediction system. *Hydrol. Earth Syst. Sci.*, **22**, 1831–1849, <https://doi.org/10.5194/hess-22-1831-2018>.
- Simmons, A., R. Mureau, and T. Petroliaqis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771, <https://doi.org/10.1002/qj.49712152711>.
- Sittner, W. T., C. E. Schauss, and J. C. Monro, 1969: Continuous hydrograph synthesis with an API-type hydrologic model. *Water Resour. Res.*, **5**, 1007–1022, <https://doi.org/10.1029/WR005i005p01007>.
- Smith, T., L. Marshall, and B. McGlynn, 2014: Calibrating hydrologic models in flow-corrected time. *Water Resour. Res.*, **50**, 748–753, <https://doi.org/10.1002/2013WR014635>.
- Stephenson, D., C. Coelho, F. Doblas-Reyes, and M. Balmaseda, 2005: Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus*, **57A**, 253–264, <https://doi.org/10.3402/tellusa.v57i3.14664>.
- Vannitsem, S., D. S. Wilks, and J. Messner, Eds., 2018: *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 362 pp.
- Wells, E., 2017: National Weather Service (NWS) Hydrologic Ensemble Forecast Service. UTA Water Management Workshop, Arlington, TX, The University of Texas at Arlington, 27 pp., [https://ispuw.uta.edu/nsf/downloads/2017\\_Workshop/Presentations/1\\_3\\_NWS\\_Wells.pdf](https://ispuw.uta.edu/nsf/downloads/2017_Workshop/Presentations/1_3_NWS_Wells.pdf).
- Westra, S., and A. Sharma, 2010: An upper limit to seasonal rainfall predictability? *J. Climate*, **23**, 3332–3351, <https://doi.org/10.1175/2010JCLI3212.1>.
- Wilks, D. S., 2006: Comparison of ensemble-MOS methods in the Lorenz'96 setting. *Meteor. Appl.*, **13**, 243–256, <https://doi.org/10.1017/S1350482706002192>.
- Wood, A. W., and J. C. Schaake, 2008: Correcting errors in streamflow forecast ensemble mean and spread. *J. Hydrometeorol.*, **9**, 132–148, <https://doi.org/10.1175/2007JHM862.1>.
- Wu, L., D.-J. Seo, J. Demargne, J. Brown, S. Cong, and J. Schaake, 2011: Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast via meta-Gaussian distribution-based models. *J. Hydrol.*, **399**, 281–298, <https://doi.org/10.1016/j.jhydrol.2011.01.013>.
- Ye, A., Q. Duan, J. Schaake, J. Xu, X. Deng, Z. Di, C. Miao, and W. Gong, 2015: Post-processing of ensemble forecasts in low-flow period. *Hydrol. Processes*, **29**, 2438–2453, <https://doi.org/10.1002/hyp.10374>.
- Zhao, L., Q. Duan, J. Schaake, A. Ye, and J. Xia, 2011: A hydrologic post-processor for ensemble streamflow predictions. *Adv. Geosci.*, **29**, 51–59, <https://doi.org/10.5194/adgeo-29-51-2011>.
- Zhu, Y., and Y. Luo, 2015: Precipitation calibration based on the frequency-matching method. *Wea. Forecasting*, **30**, 1109–1124, <https://doi.org/10.1175/WAF-D-13-00049.1>.