Clim Dyn
DOI 10.1007/s00382-017-3794-7

A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed

Louise J. Slater[1,2] · Gabriele Villarini[1] · A. Allen Bradley[1] · Gabriel A. Vecchi[3,4,5]

Received: 30 March 2016 / Accepted: 7 July 2017

Abstract The state of Iowa in the US Midwest is regularly affected by major floods and has seen a notable increase in agricultural land cover over the twentieth century. We present a novel statistical-dynamical approach for probabilistic seasonal streamflow forecasting using land cover and General Circulation Model (GCM) precipitation forecasts. Low to high flows are modelled and forecast for the Raccoon River at Van Meter, a 8900 km^2 catchment located in central-western Iowa. Statistical model fits for each streamflow quantile (from seasonal minimum to maximum; predictands) are based on observed basin-averaged total seasonal precipitation, annual row crop (corn and soybean) production acreage, and observed precipitation from the month preceding each season (to characterize antecedent wetness conditions) (predictors). Model fits improve when including agricultural land cover and antecedent precipitation as predictors, as opposed to just precipitation. Using the dynamically-updated relationship between predictand and predictors every year, forecasts are computed from 1 to 10 months ahead of every season based on annual row crop acreage from the previous year (persistence forecast) and the monthly precipitation forecasts from eight GCMs of the North American Multi-Model Ensemble (NMME). The skill of our forecast streamflow is assessed in deterministic and probabilistic terms for all initialization months, flow quantiles, and seasons. Overall, the system produces relatively skillful streamflow forecasts from low to high flows, but the skill does not decrease uniformly with initialization time, suggesting that improvements can be gained by using different predictors for specific seasons and flow quantiles.

Keywords Seasonal forecasting · Probabilistic forecast · Streamflow forecasts · North-American Multi Model ensemble (NMME)

This paper is a contribution to the special collection on the North American Multi-Model Ensemble (NMME) seasonal prediction experiment. The special collection focuses on documenting the use of the NMME system database for research ranging from predictability studies, to multi-model prediction evaluation and diagnostics, to emerging applications of climate predictability for subseasonal to seasonal predictions. This special issue is coordinated by Annarita Martiotti (NOAA), Heather Archambault (NOAA), Jin Huang (NOAA), Ben Kirtman (University of Miami) and Gabriele Villarini (University of Iowa).

Electronic supplementary material The online version of this article (doi:10.1007/s00382-017-3794-7) contains supplementary material, which is available to authorized users.

✉ Louise J. Slater
l.slater@lboro.ac.uk

[1] IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

[2] Department of Geography, Loughborough University, Loughborough, UK

[3] NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

[4] Department of Geosciences, Princeton University, Princeton, NJ, USA

[5] Princeton Environmental Institute, Princeton University, Princeton, NJ, USA

Published online: 13 July 2017

1 Introduction

Multi-model ensemble forecasting systems are increasingly popular for improving climate forecast consistency and reliability by leveraging the uncorrelated components of random and structural errors from a range of models (e.g., Hagedorn et al. 2005). The North American Multi-Model Ensemble (NMME) is a collaborative project in which participating North-American modeling centers contribute intra-seasonal to inter-annual forecasts of different climate quantities, including precipitation and temperature. The forecasts are made publicly available as a global, 1° latitude by 1° longitude grid, and with lead times ranging from 0.5 to 11.5 months (Kirtman et al. 2014). The participating modeling centers contribute both their hindcasts dating back to the early 1980s, as well as their current real-time forecasts, to an online repository (Table 1).

Existing analyses of the NMME have largely focused on evaluating the predictive skill of participating models for precipitation, temperature, or climatological extremes in different parts of the world, including China (Ma et al. 2015a, b), the Southwestern USA (Barnston and Lyon 2016) the Southeastern USA (Infanti and Kirtman 2014), the continental USA (Wang 2014; Slater et al. 2017), and at the global scale (Mo and Lyon 2015; Becker et al. 2014), over a range of lead times. Studies have also observed NMME skill in terms of large scale climate indices, showing improvements in predictive skill in regions with El Niño-Southern Oscillation (ENSO) teleconnections (Mo and Lyon 2015; Roundy et al. 2015), using NMME data to forecast the Pacific Meridional Mode precursor (Larson and Kirtman 2014), the Atlantic Warm Pool and its teleconnections (Misra and Li 2014) or to improve ENSO predictability (Barnston et al. 2015; Ham and Kug 2015; Larson and Kirtman 2014; Lopez and Kirtman 2014).

A growing number of studies are now examining the practical applications of the NMME for driving physically-based hydrologic forecasting systems. NMME outputs are used to force the variable infiltration capacity (VIC) hydrologic model (e.g., Liang et al. 1994, 1996; Mo et al. 2012; Yuan et al. 2013, 2015; Mo and Lettenmaier 2014; Shrestha et al. 2015; Sikder et al. 2015), or the mesoscale hydrologic model (mHM) (Thober et al. 2015). The hydrologic forecasts resulting from these models are typically assessed against those obtained from the ensemble streamflow prediction (ESP) method, which uses meteorological forcings resampled from the historical distribution as well as initial land condition persistence (Twedt et al. 1977; Day 1985). Encouragingly, results show that the limited accuracy of raw NMME precipitation forecasts tends to improve when they are used inside a flow forecasting framework (Wood et al. 2016). However, physically-based NMME forecasts of streamflow, soil moisture or runoff consistently indicate that the potential forecast skill depends on both the antecedent hydrologic conditions and the forecast climate conditions. The most skillful forecasts tend to be obtained in dry or subfreezing regions, when runoff is strongly influenced by initial hydrologic conditions (e.g., during southern US summers, or northern Great Plain winters). In contrast, the least skillful forecasts are obtained in wet conditions when there is a strong coupling between rainfall and runoff, and a more limited influence of initial soil moisture conditions on streamflow (e.g., the US West Coast) (Mo et al. 2012; Yuan et al. 2013; Mo and Lettenmaier 2014; Wood et al. 2016). To varying degrees by location and season, flow forecast skill depends on both initial hydrologic conditions and climate model forecast skills. Additionally, the skill of NMME-driven forecasting systems depends on the influence of large-scale climate drivers, such as teleconnections from the El Niño-Southern Oscillation (ENSO) phenomenon (Shrestha et al. 2015), and on the temporal resolution of the forecast target, with seasonally-aggregated forecasts producing better results than those focusing on higher-frequency data in major river basins (Yuan et al. 2015; Sikder et al. 2015).

Despite the growing interest in using NMME data to force physical hydrologic models, no study has yet developed a dynamical streamflow forecasting system (using GCM forecasts) based on statistical models. Statistical models can be viewed as a complementary approach to physically-based hydrologic models, with considerable advantages arising from their simplicity, limited computational costs and flexibility. In this study we explore the skill of statistical models in forecasting seasonal streamflow (from low to high flows) at the Raccoon River at Van Meter (USGS ID 05484500), located in central-western Iowa. Our rationale for choosing this location is the frequency of hydrologic extremes in the region. Catastrophic weather events are regularly responsible for billion-dollar economic and social losses in Iowa (Smith and Matthews 2015). In 2013 alone, the State witnessed widespread flooding during the spring and drought conditions during the summer. While little can be done to prevent such events, we *can* improve preparedness by enabling Iowans to make more informed water management decisions.

A second novelty of this work is the use of agricultural land cover as a predictor in the statistical–dynamical framework. The influence of changing catchment land cover on flow distributions is still poorly understood (Rogger et al. 2017), and Iowan watersheds are a good place to test this influence as many have witnessed extensive change over the past century. Across much of the US Midwest, the conversion of perennial grasslands and forests to agricultural row crops (e.g., Frans et al. 2013) is believed to have amplified the influence of precipitation on groundwater recharge and soil water storage (Zhang and Schilling 2006;

Table 1 Summary of the characteristics of the eight NMME models

Model name	Modeling center	Available period	Ensemble size	Lead times (months)	References	Retrieved from
CCSM3 (Community Climate System Model version 3)	NCAR/COLA/RSMAS	1982–Present	6	0.5–11.5	Kirtman and Min (2009)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.COLA-RSMAS-CCSM3/
CCSM4 (Community Climate System Model version 4)	NCAR/COLA/RSMAS	1982–Present	10	0.5–11.5	Lawrence et al. (2012)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.COLA-RSMAS-CCSM4/
CanCM3 (3rd Generation)	CMC	1981–Present	10	0.5–11.5	Merryfield et al. (2013)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.CMC1-CanCM3/
CanCM4 (4th Generation)	CMC	1981–Present	10	0.5–11.5	Merryfield et al. (2013)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.CMC2-CanCM4/
GFDL2.1 (version 2.1)	NOAA/GFDL	1982–Present	10	0.5–11.5	Zhang et al. (2007); Delworth et al. (2006)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.GFDL-CM2p1-aer04/
GEOS5 (Goddard Earth Observing System Model version 5)	NASA/GMAO	1981–Present	12	0.5–8.5	Vernieres et al. (2012); Molod et al. (2012)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NASA-GMAO-062012/
FLORb01 (version 2.5)	NOAA/GFDL	1982–Present	12	0.5–11.5	Vecchi et al. (2014)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.GFDL-CM2p5-FLOR-B01
CFSv2 (operational Climate Forecast System version 2)	NOAA/NCEP	1982–Present	28 (24 used 4 incompl.)	0.5–9.5	Saha et al. (2014)	http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/.NCEP-CFSv2/

The number of ensemble members indicates the largest number of members per model. The 0.5-lead time is the shortest available lead time and refers to the forecast for a month issued at the beginning of the month itself (e.g., the 0.5 lead time forecast for January 2000 is issued at the beginning of January 2000). See the Methods for the definition of modeling center acronyms

Schilling et al. 2008; Gupta et al. 2015), thus increasing the low flows in dry seasons, and the high flows during heavy rainfall events (Villarini and Strong 2014). These effects are also amplified by the role of artificial drainage, which accelerates water residence time in depressions of water fields (Schottler et al. 2014). In watersheds like the Raccoon River basin, the fraction of the land surface cultivated as corn or soybean has grown from approximately one-third to over three quarters of the watershed (Fig. 1). Therefore, if land cover has a demonstrable impact on streamflow distributions, taking a fraction of the land out of agricultural production may potentially be used as a viable strategy to attenuate risks arising from hydrological extremes. The availability of seasonal streamflow forecasts in this region could be of considerable societal and economic use for water resources management, disaster forecasting and prevention, energy, finance and insurance, food security, policy-making and public authorities, transportation (e.g., Harrison et al. 2007), and for crop management decisions (e.g., Asseng et al. 2016). Thus, in this work, we address the following research questions:

- How well can seasonal variations in streamflow be described with a simple statistical model using just precipitation and agricultural acreage as predictors for all flow quantiles?
- How well can we forecast seasonal streamflow across a wide range of quantiles (from low to high flow), and how does the skill change as a function of lead time (i.e., initialization month) and season?

2 Data and methods

Our forecasting framework can be summarized as follows, with details provided in subsequent sections. We begin by fitting three statistical streamflow models from 1927 to 2016 using the observed streamflow quantiles, basin-averaged precipitation, and total agricultural row crop acreage. The three models are assessed against the observed streamflow quantiles in terms of their goodness-of-fit. We then use these models to forecast streamflow quantiles for the years 2001–2016. The models are both statistical (based on gamma regression models) and dynamical (using precipitation forecasts from eight NMME GCMs and agricultural acreage as inputs).

2.1 Data for historical model fits

As reference streamflow data, we use daily observations from the Raccoon River at Van Meter (USGS station number 05484500; Fig. 1), which has a drainage area of 3441 mi^2 (8912 km^2). For every season, we compute streamflow quantiles ranging from $Q_{0.00}$ (minimum daily streamflow) to $Q_{1.00}$ (maximum daily streamflow) with a step of 0.05. These quantile time series are used to fit the models from low to high flows at the seasonal time scale. All of our observed time series begin in 1927 because that is the year from which the harvested corn and soybean data are available. We focus on four seasons: winter (December–January–February; DJF), spring (March–April–May;

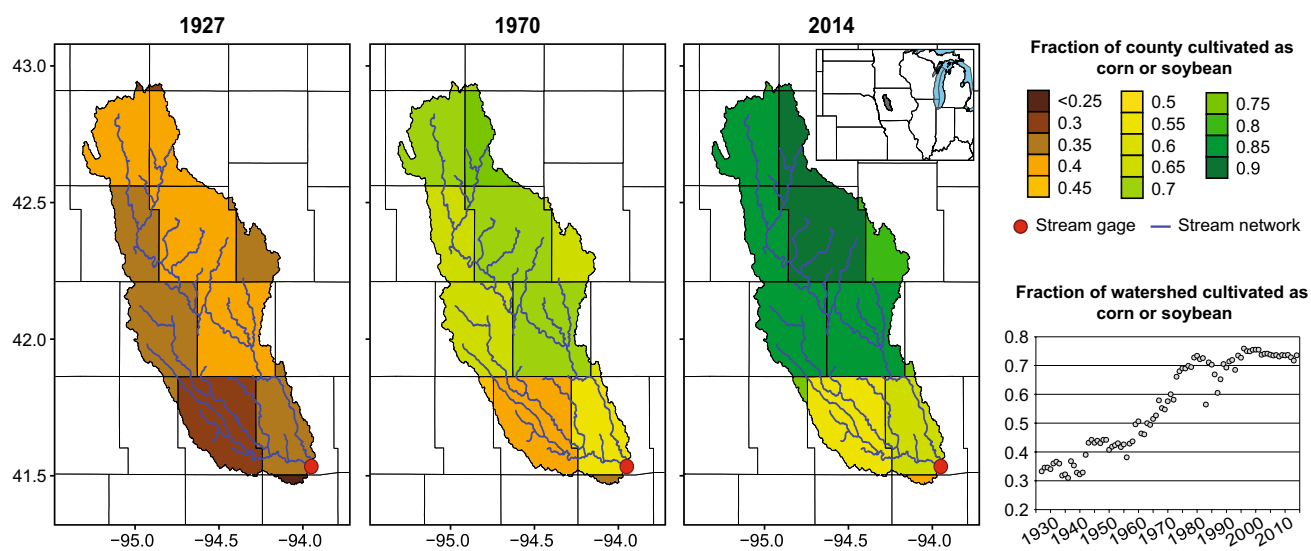

Fig. 1 Evolution of corn and soybean cultivation within the Raccoon River basin from 1927 to 2014. The three maps on the *left* indicate land cover for the years 1927, 1970 and 2014 and show the progression of corn and soybean cultivation within the basin, as a fraction of each county. The Raccoon River at Van Meter stream gage is indicated as a *red circle* at the outlet of the basin. The *inset* map in the 2014 *panel* shows the location of the watershed within the US Midwest. The time series on the right indicates the change in the fraction of the total watershed that is cultivated as corn and soybean, from 1927 to 2014

MAM), summer (June–July–August; JJA), and fall (September–October–November; SON).

To model the seasonal streamflow quantiles, we use precipitation and agricultural harvested land cover as predictors. Basin-averaged total monthly precipitation records are obtained from the PRISM Climate Group and aggregated at the seasonal scale. These data are freely available online from 1890 to the present (Daly et al. 2002), and represent the official climatological data for the US Department of Agriculture (USDA). The precipitation data are averaged over the entire catchment draining to the Van Meter stream gauge, providing basin-averaged monthly (aggregated to seasonal) precipitation time series.

Our second predictor is total harvested corn and soybean acreage at the county level from the USDA's National Agricultural Statistics Services (NASS) database (NASS and USDA 2015). Following Villarini and Strong (2014), we use these data to represent the effects of agricultural practices and land cover changes on the flow frequency distribution (see also Schilling et al. (2008), among others). We calculate the fraction of each county that is within the Raccoon River basin (Fig. 1) and multiply this fraction by the total agricultural acreage of that county. Assuming that the farmed area is uniformly distributed within each county, the total acreage within the Raccoon watershed is then obtained as the sum of computed values across all counties. We do this for every year to obtain a time series of total annual cultivated corn and soybean acreage, from 1927 to 2014. Data for 2015/2016 was not yet available at the time of writing this paper, so we use the 2014 data instead, under the assumption that farmed acreage has not changed considerably in 2 years (e.g., Villarini and Strong 2014). Figure 1 shows how the fraction of each county cultivated as corn and soybean has changed between 1927 and 2014. Overall, there has been a sharp increase in agricultural intensity over this watershed since the 1940s, when perennial vegetation was replaced with seasonal row crop (corn and soybean). This agricultural push continued into the 1970s, but has levelled out since the mid-1990s, with about three quarters of the watershed in production (e.g., Zhang and Schilling 2006; Schilling et al. 2008; Villarini and Strong 2014; Fig. 1).

2.2 Statistical model formulation

Our statistical modeling builds on and improves the methodology described in Villarini and Strong (2014). Let Y, the predictand, represent a given quantile of the seasonal streamflow time series ranging from minimum ($Q_{0.00}$) to maximum ($Q_{1.00}$) flow. For example, if Y is the spring $Q_{0.50}$, we compute the median of the daily streamflow distribution for the 3-month period ranging from March until May (MAM), for each year, from spring 1927 to 2016. The resulting time series represents Y.

For each given quantile, we model Y using a gamma distribution, which has two parameters, μ and σ. Based on the parameterization in the Generalized Additive Models for Location, Scale and Shape (GAMLSS; Rigby and Stasinopoulos 2005; Stasinopoulos et al. 2006), the expected value of Y is equal to μ and the variance to $\sigma^2\mu^2$. We describe the variability of these two parameters over time in terms of covariates (predictors): x_p, which indicates precipitation over the study area; and x_a, which is the harvested corn and soybean acreage. All the predictors are standardized relative to the 1983–2000 period (by subtracting the mean for the 1983–2000 period, and dividing by the standard deviation for the 1983–2000 period). We write three different formulations of the model, starting with Model 1 (the same as in Villarini and Strong 2014):

$$\begin{aligned} \mu_1 &= \exp\left(\alpha_1 + \beta_1 \cdot x_p + \gamma_1 \cdot x_p \cdot x_a\right), \\ \sigma_1 &= \exp\left(\kappa_1 + \eta_1 \cdot x_p + \xi_1 \cdot x_p \cdot x_a\right), \end{aligned} \tag{1}$$

where the Greek letters within the parentheses represent the coefficients to estimate. The interaction term $(x_p \cdot x_a)$ is used to model the influence of changing agricultural land cover on the streamflow distribution. For instance, increases in agricultural land cover due to deforestation and forest fragmentation are expected to amplify precipitation effects on low flows (Zhang and Schilling 2006; Schilling et al. 2008) and high flows during the heaviest rainfall events (Villarini and Strong 2014), thus strengthening the streamflow-precipitation relationship.

Model 2 is a simplified version of Model 1, with the same μ but a constant σ, because Villarini and Strong (2014) found that the σ parameter was not significantly dependent on the two predictors:

$$\begin{aligned} \mu_2 &= \exp\left(\alpha_2 + \beta_2 \cdot x_p + \gamma_2 \cdot x_p \cdot x_a\right), \\ \sigma_2 &= \exp(\kappa_2). \end{aligned} \tag{2}$$

Last, Model 3 includes an additional parameter, x_d, which is the observed precipitation from the month preceding each season, used as a proxy for the effects of antecedent wetness:

$$\begin{aligned} \mu_3 &= \exp\left(\alpha_3 + \beta_3 \cdot x_p + \gamma_3 \cdot x_p \cdot x_a + \partial_3 \cdot x_d\right), \\ \sigma_3 &= \exp(\kappa_3). \end{aligned} \tag{3}$$

We selected these three different model formulations to reflect the fact that there are multiple plausible ways of connecting streamflow to precipitation and agriculture, but recognize that there are many other potential formulations, and that this paper only explores one aspect of that

relationship. Other model formulations could have been considered, especially including non-linear dependencies. Here we wanted to keep the modeling simple and examine which model produces the most consistent streamflow forecasts based on the different streamflow-predictor relationships. To assess the extent to which the inclusion of x_a improves model fit, we quantify the improvement in model fit and forecasts (see Sect. 3 and Supplementary Materials).

For each seasonal streamflow quantile ranging from $Q_{0.00}$ to $Q_{1.00}$, we estimate the relevant parameters in Eqs. 1–3 using the observed streamflow, observed precipitation, and agricultural acreage for the period 1927–2015. For each seasonal streamflow quantile, we compute the residuals and assess their mean, variance, coefficient of skewness, coefficient of kurtosis, and Filliben correlation coefficient (Filliben 1975), to evaluate the goodness-of-fit of each model from a statistical standpoint. If these models are able to describe the systematic variations in the streamflow quantiles, the residuals should be white noise (Gaussian distributed with a mean of 0 and a variance of 1). We also retrieve the computed percentiles P_5, P_{25}, P_{50}, P_{75} and P_{95} of the fitted gamma distributions (i.e., a probabilistic forecast is issued for every seasonal flow quantile), and plot them to show how each model performs. We would expect the fitted quantiles to be close to the observed flow distribution (e.g., the value of the P_{95} is exceeded by roughly 5% of the observations). Last, these same values are also displayed as time series alongside the observed data to assess the goodness-of-fit of each model.

2.3 Model evaluation

Numerous statistical measures of model accuracy or performance criteria exist in the literature, such as the mean square error, the mean/median absolute error, the root mean squared error, the mean/median absolute percentage error, the mean/median relative absolute error, the Nash-Sutcliffe efficiency metric. Most of these measures have been shown to be inadequate, sensitive to outliers, or produce infinite/undefined values and misleading results (e.g., Hyndman and Koehler 2006). The mean absolute scaled error (MASE) has been proposed as a preferable measure, as it is less sensitive to outliers, less variable on small samples than other metrics, and easy to interpret (Hyndman and Koehler 2006; Franses 2016). Because of the probabilistic nature of the forecasts, we measure the accuracy of the fitted models by comparing the computed P_{50} (50th percentile) of the fitted gamma distribution as "best estimate" (e.g., Villarini and Serinaldi 2012) with the observed historical $Q_{0.5}$, using the MASE as our chosen metric for every seasonal streamflow quantile.

2.4 NMME data and streamflow forecasts (2001–2016)

We compute the forecasts over the 2001–2016 period so that we have enough observational data to verify the skill of the forecasts. Precipitation forecasts are obtained from eight NMME models (see Table 1 for details): CCSM3 and CCSM4 from the National Center for Atmospheric Research (NCAR), the Center for Ocean-Land-Atmosphere Studies (COLA) and the Rosenstiel School of Marine and Atmospheric Science from the University of Miami (RSMAS); CanCM3 and CanCM4 from Environment Canada's Meteorological Service of Canada - Canadian Meteorological Center (CMC); GFDL2.1 and FLORb01 from the National Oceanic and Atmospheric Administration (NOAA)'s Geophysical Fluid Dynamics Laboratory (GFDL); GEOS5 from the National Aeronautics and Space Administration (NASA)'s Global Modeling and Assimilation Office (GMAO); and CFSv2 from NOAA's National Centers for Environmental Prediction (NCEP). Every modeling center contributes the model forecasts to the NMME through the International Research Institute (IRI) for Climate and Society web archive. Each of the eight NMME models has between 6 and 24 members, but we use only the mean of each model's members in this study.

Monthly precipitation forecasts are available at the beginning of each month with a temporal horizon of between 0.5 and up to 11.5 months—so for instance, the precipitation forecast initialized at the beginning of October 2013 provides monthly precipitation forecasts ranging from October 2013 to September 2014. We downloaded the data for each of the eight NMME models, and extracted the areal-averaged forecast issued within the boundaries of the Raccoon River basin for every 1 of the 94 model members (see Table 1). We computed the ensemble forecast as the mean of all the members. This approach is recognized as a simple but effective method for obtaining an enhanced multi-model forecast, wherein the skill of the means is often as good as or better than that of the best single ensemble member (e.g., Becker et al. 2014; Slater et al. 2017; Ma et al. 2015a). However, more sophisticated approaches can also be utilized to heighten the skill of the input precipitation data (e.g., Scheuerer and Büermann 2014; Wanders and Wood 2016; Hodyss et al. 2016; Bogner et al. 2017).

The time series were then aggregated to obtain the total seasonal precipitation ensemble forecast for every lead time. For example, the summer forecast issued in June is the sum of the 0.5 month lead forecast for June, the 1.5-lead forecast for July, and the 2.5-lead forecast for August. Thus, a seasonal forecast can be issued up to 9.5 months in advance for an NMME model issuing 11.5 lead times. Last, the seasonal forecasts are transformed as standardized anomalies (by subtracting the mean and dividing by the standard deviation computed over the 1983–2000 period),

so that they are on the same scale as the observed data (Villarini and Strong 2014). All anomalies are computed with respect to the baseline period 1983–2000, i.e., the year from which all NMME forecasts are available, and the last year of the model fit (before forecasts are issued). The raw NMME precipitation forecasts tend to have relatively low skill and are much less variable than the observed data (Fig. 2). There is no clear difference between the precipitation forecasts issued at different lead times, in terms of variability or magnitude. Overall, the summer forecast tends to be slightly better and more consistent than the other seasons (i.e., the scatter is a little closer to the 1:1 line; Fig. 2). Both observed precipitation and agricultural acreage are also transformed as standardized anomalies so that they are on comparable scales for the model fitting and forecasting.

To compute a streamflow forecast for a given year, streamflow quantile, initialization month, and season, we begin by training the model by fitting observed values of x_a, x_p and x_d for all years preceding the forecast year. The corresponding values of α, β, γ, δ, κ, η, and ξ (where relevant: see Eqs. 1–3) are retrieved for the given period (e.g., 1927–2000 for the forecast year 2001; 1927–2001 for the forecast year 2002), consistent with the retroactive validation method described in Mason and Baddour (2008). To compute the streamflow forecasts, we use the precipitation forecasts from the NMME archive (x_p), and the harvested soybean and crop acreage from the previous year (x_a, persistence forecast). This approach is similar to those described in Jain and Lall (2001), Sankarasubramanian and Lall (2003), and Towler et al. (2010). For x_d, if we are predicting streamflow at the shortest lead time (the same month), we use observed precipitation from the previous month, if it is already available. If the observed precipitation is not yet available, and for all lead times beyond the 0.5 month lead, we retrieve the NMME forecast value (transformed as standardized anomaly) for the month preceding the given season.

The observed values of x_a, x_p and x_d for all years preceding and excluding the forecast year are used to compute the μ and σ parameters for the forecast year. The gamlss.dist package (Stasinopoulos et al. 2016) takes the μ and σ parameters as input to compute the predicted streamflow distribution for the given forecast quantile. For example, if we are predicting the $Q_{0.50}$ for 2001, our probabilistic forecast produces a range of probable percentiles wherein the forecast value may lie (we retain the percentiles P_5, P_{25}, P_{50},

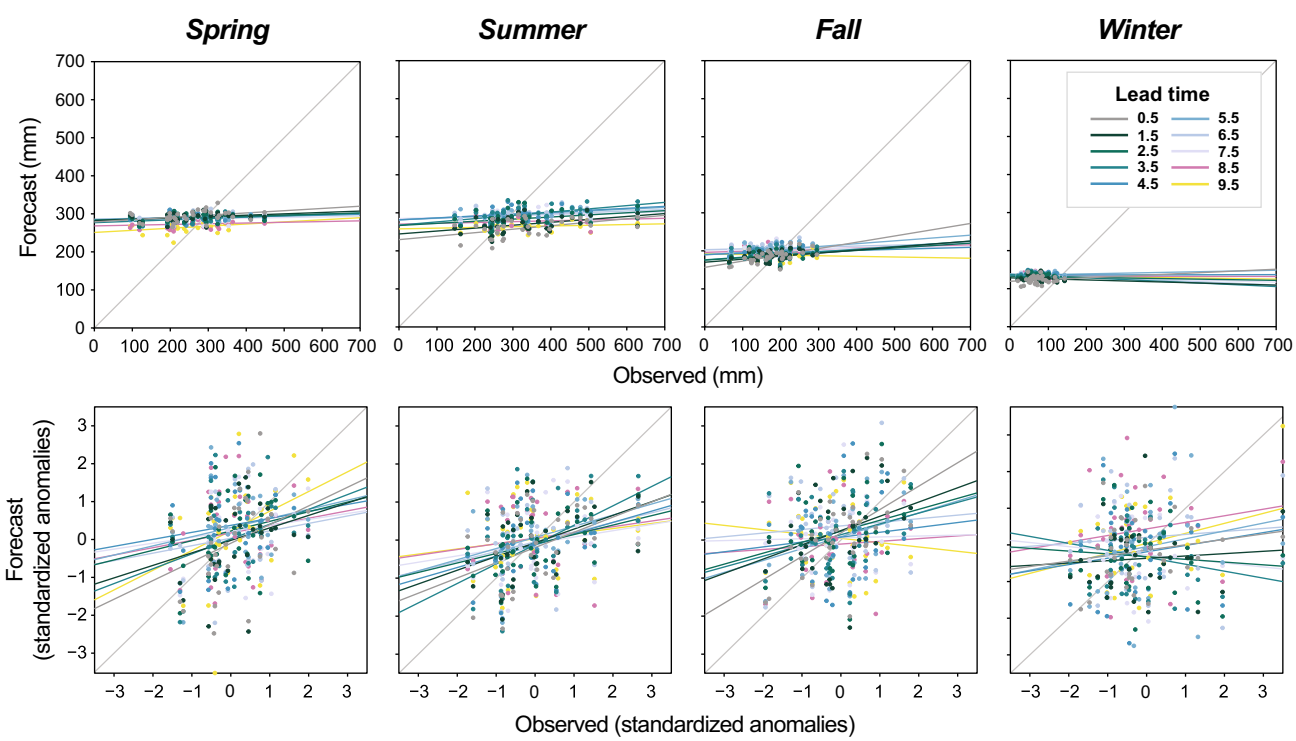

Fig. 2 Scatterplots of observed basin-averaged precipitation for the Raccoon River watershed, versus NMME precipitation (ensemble mean) at different lead times, before and after standardizing the data. The *top row* indicates raw forecasted (NMME) versus raw observed (PRISM) data. The bias in the data is clear and varies seasonally and by lead time (i.e. number of months ahead of the season). The *bottom row* indicates standardized anomalies for both observed and forecasted data. The values (1983–2016) are transformed by subtracting the mean and dividing by the standard deviation computed over the 1983–2000 period. Different lead times of the NMME forecasts are shown using a *color* spectrum (*top right inset panel*)

P_{75} and P_{95} to visualize and evaluate the forecast). We store the probabilistic forecast output and the μ and σ parameters for each model, initialization month, season, forecast year, and predicted quantile.

Our streamflow forecasts are verified both deterministically and probabilistically. The forecast accuracy is based on the comparison between the median of the forecast distribution (P_{50}, for each initialization month) and the observations, for every seasonal streamflow quantile. The MASE score, a scaled error measure, is used to compare the forecast against the average one-step naïve forecast, where values smaller than 1 indicate that the model performs better, on average, than the one-step naïve forecast. We also compute the correlation coefficient to quantify the amount of variability in the observations that is explained by the forecasts.

3 Results

3.1 Model fit and diagnostics

The three models produce good fits to the observed historical data from 1927 to 2016, with correlation coefficients between the observed quantiles and the predicted percentile P_{50} generally exceeding 0.6 across all streamflow quantiles and seasons (Supplementary Fig. 1). Model 3, which includes both agricultural land cover and antecedent precipitation as predictors, tends to produce the best fits, with higher correlation coefficients ($R > 0.7$) and lower root mean square error, especially in the fall and summer when antecedent moisture may affect the relationship between precipitation and streamflow (Supplementary Fig. 1).

The time series indicate how closely the observed seasonal streamflow quantiles ($Q_{0.05}$, $Q_{0.5}$ and $Q_{1;}$ gray circles) lie with respect to the modelled probabilistic distribution (P_5–P_{95}; colored ribbons) for the entire period (Fig. 3 and Supplementary Fig. 2). Model 3 reproduces the observed inter-annual variability relatively well for all seasons. Some periods like the spring/winter low flows ($Q_{0.05}$) in the 1970s–1980s are less well fit, possibly because the inter-annual variability of x_a is much higher and so less well captured by the persistence forecast. The progressive improvement in model fit over time is possibly related to the increased role of row crop acreage (Fig. 1), as this is the only variable that changes notably over the time period. The width of the fitted distributions is generally narrow, and the goodness of these fits is supported by high

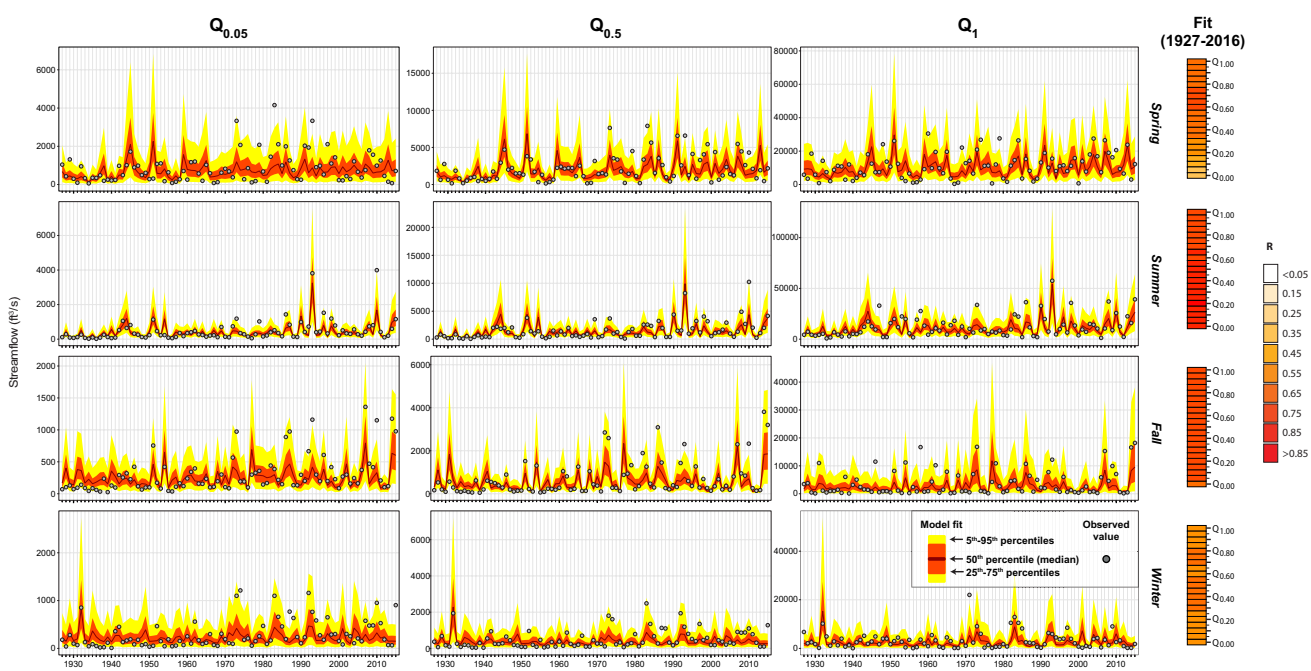

Fig. 3 Time series indicating the fit of Model 3 against the observed values (1927–2016; see Supplementary materials for a comparison of the three models). For every season (*rows*), five percentiles of the predicted streamflow distribution are shown (P_5, P_{25}, P_{50}, P_{75} and P_{95}) within each plot (*color ribbons*), for three quantiles (*columns*): low flow ($Q_{0.05}$), median flow ($Q_{0.50}$), and maximum seasonal flow (Q_1). The *dark red line* represents the median (P_{50}) of the predicted distribution, the *orange region* the area between P_{25} and P_{75}, and the *yellow region* the area between P_5 and P_{95}. The *gray circles* indicate the observed values. To the right of the time series, seasonal *color bars* indicate the fit (as measured by the correlation coefficient R) of the observed versus the fitted flow quantiles (1927–2016), ranging from white ($R < 0.05$) to red ($R > 0.85$)

correlation coefficients: R > 0.85 in much of fall and summer; R > 0.5 in much of spring and winter, with the best fits often in the high flow quantiles (Fig. 3; Supplementary Fig. 1). The model fit is the closest to the observed values in the summer months, when the flow is most variable, and the least well-fit in the spring low flows (suggesting that we may be missing important predictors reflecting antecedent conditions; Fig. 3 and Supplementary Fig. 1). Ideally, the observed values should lie close to the median of the predicted distribution (dark red line). However, since our forecasts are probabilistic (i.e., provide the whole probability distribution of flow forecasts, for each streamflow quantile) it is to be expected that 10% of observed values will lie above and below the 95th and 5th percentiles (white area), 40% between the 5th–25th and 75th–95th percentiles (the yellow ribbon), and remaining 50% between the 25th and 75th percentiles (orange ribbon).

Statistically, we assess the goodness-of-fit of the three gamma models to determine which model produces the best fit for every season and flow quantile, and to detect where the uncertainties lie (Figs. 4, 5). Overall, the residuals (observed minus predicted values) are relatively well-distributed for all three models: the mean and coefficient of skewness tend to be close to zero (indicating that the errors have zero bias and that their distribution is symmetric), the variance close to one (indicating that the models are not over- or under-dispersed), the coefficient of kurtosis close to three (so they are neither leptokurtic or platykurtic), and the Filliben correlation coefficients close to one (supporting the Gaussianity of the distribution of the residuals) (Fig. 4). Model 3 does tend to out-perform the other two models, although there are notable differences by season and streamflow quantile. In the spring, summer, and winter, relatively little difference can be seen among models.

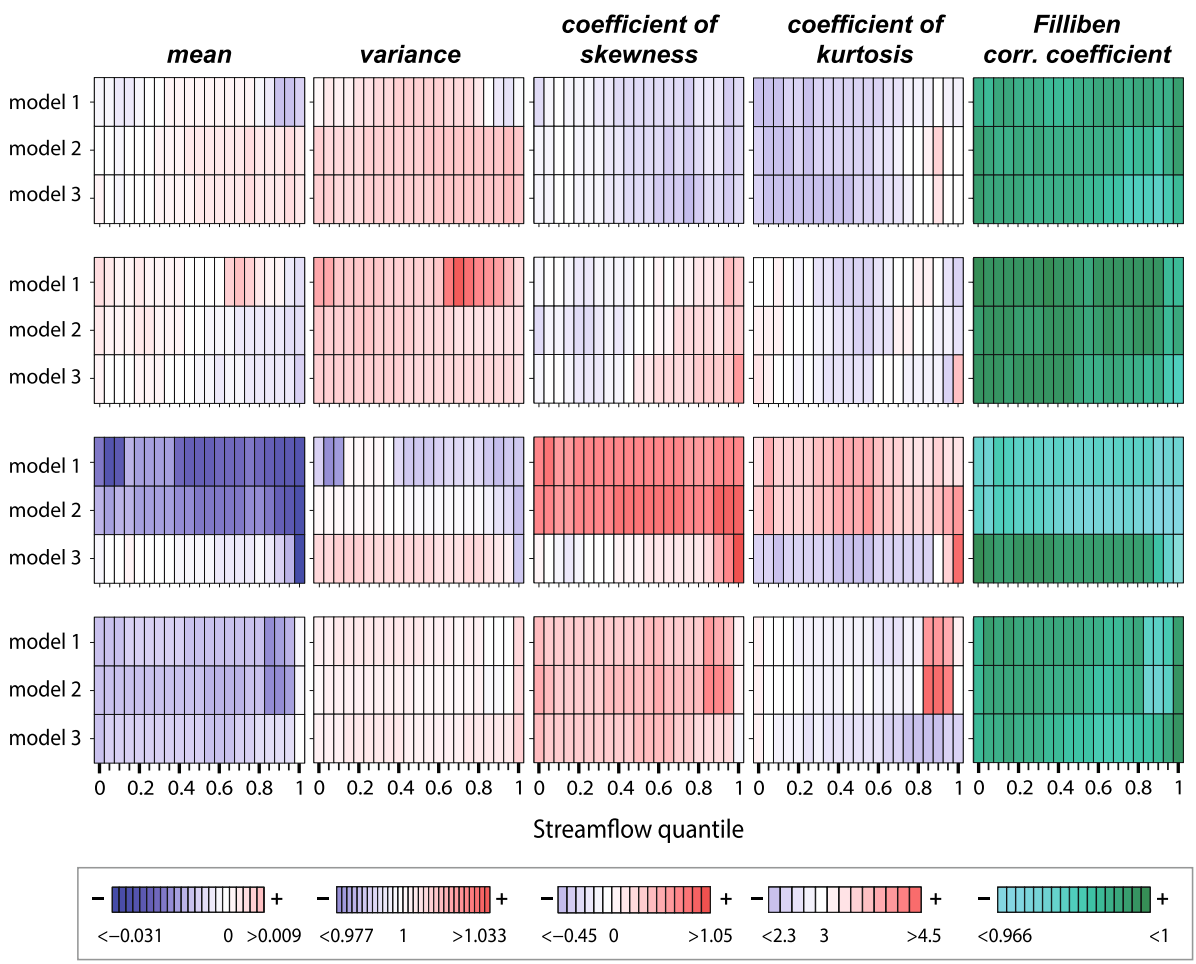

Fig. 4 Summary statistics for the three model residuals (1927–2016). Mean, variance, coefficient of skewness, coefficient of kurtosis, and Filliben correlation coefficient are displayed for all of the seasonal forecasts, for every model. The first four metrics are displayed with a scale that is centered on the best fit of the indicator (*white*), ranging from smaller (*blue*) to higher (*red*) values compared to the target. The Filliben correlation coefficient ranges from low (*turquoise*) to high (*dark green*: the target is 1). The full range of forecast streamflow quantiles is shown on the x-axis, from zero to one

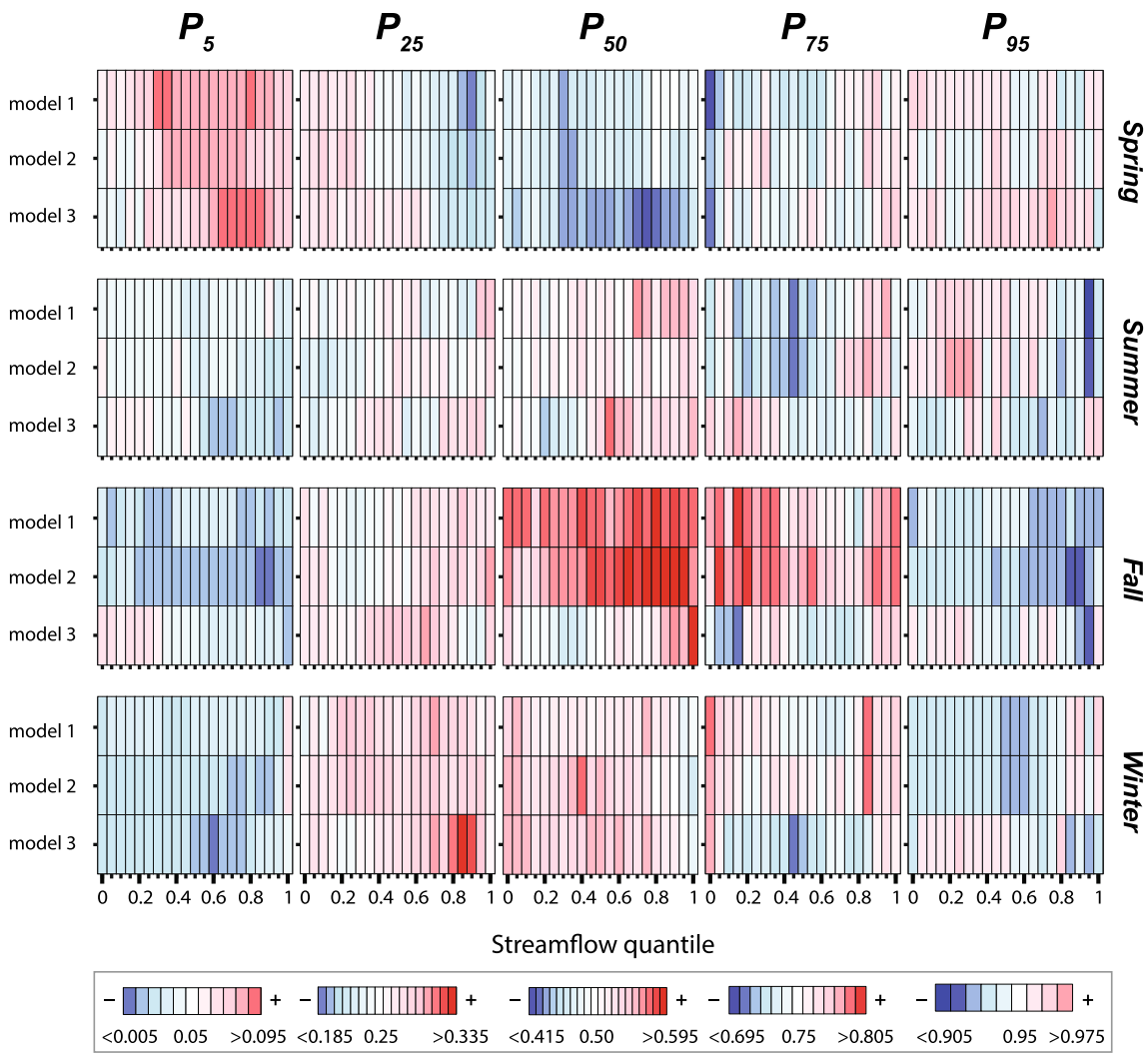

Fig. 5 Probability coverage of five fitted percentiles (P_5, P_{25}, P_{50}, P_{75} and P_{95}), for all streamflow quantiles (Q_0–Q_1) for the three models (1927–2016). The full range of streamflow quantiles are shown on the x-axes (Q_0–Q_1), and the three models on the y-axes. Colors indicate the departure of the fitted distribution (P_5–P_{95}) from the observed streamflow quantile (obtained from the flow-frequency distribution of daily streamflow during the given season). *Shades of white* indicate a close fit; *blue shades* that the fit is beneath the observed value; *red shades* that it is above. The bins are all 0.01 quantile in width, and are centered so that the central quantile is white (0.045–0.055; 0.245–0.255; 0.495–0.505; 0.745–0.755; 0.945–0.955)

In the fall, however, Models 1 and 2 tend to over-predict (the mean of residuals tends to be negative; Fig. 4), suggesting that the inclusion of antecedent wetness conditions in Model 3, x_d, plays a crucial part in the improvement of seasonal fits in the fall.

Despite model similarity, seasonal differences can also be observed across the residual diagnostics (Fig. 4). The summer and spring, which generally witness higher flows than fall and winter, have better results in terms of the mean, skewness, kurtosis, and Filliben coefficient, but also greater variability. Their coefficient of kurtosis is flatter than might be expected from a Gaussian distribution, with more evenly-distributed residuals in the summer months. The model residuals tend to be positively skewed in fall and winter, suggesting a slight tendency to over-predict, as can be seen with the mean. The winter months have the lowest flows and the lowest variance, while fall tends to produce the least accurate fits of all seasons.

When comparing across streamflow quantiles, we find that the low quantiles tend to be slightly better fitted than the high quantiles (Fig. 4), especially in Model 3 (e.g., fall mean or skewness), suggesting that x_d is most helpful in fitting low flows. This finding is consistent with the idea that antecedent wetness contributes to the forecast skill in dry conditions (e.g., Mo and Lettenmaier 2014). However, the majority of models display little difference in model skill among streamflow quantiles, suggesting that the high flows

tend to be equally well predicted to the low flows, which is very promising.

Additionally, we evaluate the goodness-of-fit of the three models by assessing the distance between the predicted and observed distributions through probability coverage (Fig. 5). As an example, the 50th percentile of the fitted probabilistic distribution (P_{50}) tends to exceed about 40% (under-predicting, blue shades) to 60% (over-predicting, red shades) of the observed values. So in spring, the P_{50} tends to be slightly too low (i.e., on the time series, a majority of observed values lie above the fitted red line). Ideally, shades of white indicate that the predicted percentiles are close to the observed streamflow quantiles. Overall, the majority of fitted flows tend to lie within an acceptable range (approximately ± 0.1) from the corresponding observed streamflow quantile. There is not one model that consistently outperforms the others across all seasons, although Model 2 (with constant σ) performs better (i.e., is slightly whiter) than the others in spring and summer, and Model 3 (with x_d) performs best in fall, in keeping with the residual diagnostics shown in Fig. 4, and the time series in Supplementary Fig. 2. Surprisingly, though one might assume that the median of the fitted distribution would show the smallest errors, some of the worst fits are for the P_{50}. Among the four seasons, the summer months tend to be the best fit, and fall the least well-fit. However, among streamflow quantiles, the goodness-of-fit does not vary notably from low to high flows.

3.2 Model forecast skill

The time series in Fig. 6 indicate how the observed seasonal streamflow values lie with respect to the historical fit (1980–2000, same as Fig. 3) and forecast (2001–2015) of Model 3, for three streamflow quantiles ($Q_{0.05}$, $Q_{0.5}$ and Q_1). Both the historical fit and forecast values are shown on the same figure, to allow some comparison between the fit and forecast. Clearly, the skill of the forecast streamflow does not decrease uniformly with initialization time (i.e., number of months ahead of the forecast; Fig. 6) and is somewhat erratic (Supplementary Figs. 3–4). For certain initialization months (e.g., 0.5 months ahead of the fall season, or 1.5 months ahead of the winter season), the forecast (2001–2015) flow distributions perform almost as well as the fitted distributions (1927–2000) (Supplementary Figs. 1 and 3), despite the uncertainty resulting from the NMME precipitation forecasts. In fall, we find relatively consistent high skill for Model 3's fit and forecast (measured in terms of the correlation coefficient, R) across all streamflow quantiles (Figs. 3, 7). In the spring, summer, and fall, the high streamflow quantiles tend to be better forecast than the low flows at the shortest initialization times, showing great promise for the prediction of seasonal streamflow maxima.

To what extent does the inclusion of antecedent wetness (x_d) and agricultural land cover (x_a) actually improve the model forecasts? Model 3 reveals that there is some improvement in model forecasts resulting from the inclusion of x_d (higher R and lower RMSE in Supplementary Figs. 3–4): this is particularly true at the longer lead times in summer and fall, but not consistent across initialization months. The inclusion of x_a clearly improves the summer low-flow forecasts up to 4.5 months ahead of the forecast season (Supplementary Fig. 3). However, the spring low-flow forecasts are actually *better* without the inclusion of x_d, suggesting that forecast skill may be improved by selectively retaining only certain predictors in the seasons and flow quantiles where they are most relevant.

We also assess the MASE score of our models, as it is a more robust measure of skill (Hyndman and Koehler 2006; Franses 2016) than other metrics, and paints a slightly different picture than the correlation coefficient. MASE indicates that forecast accuracy is most reliable in the spring and summer months (low values, in red, Fig. 7 and Supplementary Fig. 5), *i.e.*, when the precipitation input forecasts are also most consistent. Interestingly, the MASE skill of the high-flow forecasts tends to increase with initialization time in the spring, and decreases with initialization time in the fall, when depleted water reserves and heightened evapotranspiration possibly decouple the rainfall-runoff relationship (Fig. 7 and Supplementary Fig. 5). The low flow forecasts, on the other hand, are the least skillful in the spring, when sub-freezing conditions trap moisture in the ground, and best in the summer and fall, when streamflow is slightly more responsive to precipitation. In other terms, our findings suggest that the added value of the NMME forecasts is most important at longer lead times in wet conditions (*i.e.*, spring), when the rainfall-runoff coupling is at its highest and the initial conditions have lesser impact on the forecast (Yuan et al. 2013), and is least important in dry or subfreezing conditions when rain and runoff are most strongly decoupled. The inclusion of antecedent wetness conditions in Model 3 thus improves model predictions in the fall, following the summer months (Supplementary Figs. 3 and 5).

Last, we assess whether the skill of our streamflow forecasts does actually improve over that of the initial input precipitation forecasts from the NMME multimodel ensemble, as found in Wood et al. (2016). Results indicate that when the initial precipitation forecast skill is high, the resulting streamflow forecast skill tends to remain equally high, with some exceptions, e.g., summer short initialization months (Fig. 7). In some cases, the streamflow forecasts do actually improve over the input precipitation skill, such as for some high-flows in the fall and spring, or some low flows in summer and winter. This non-uniform manner in which the precipitation forecast skill propagates through

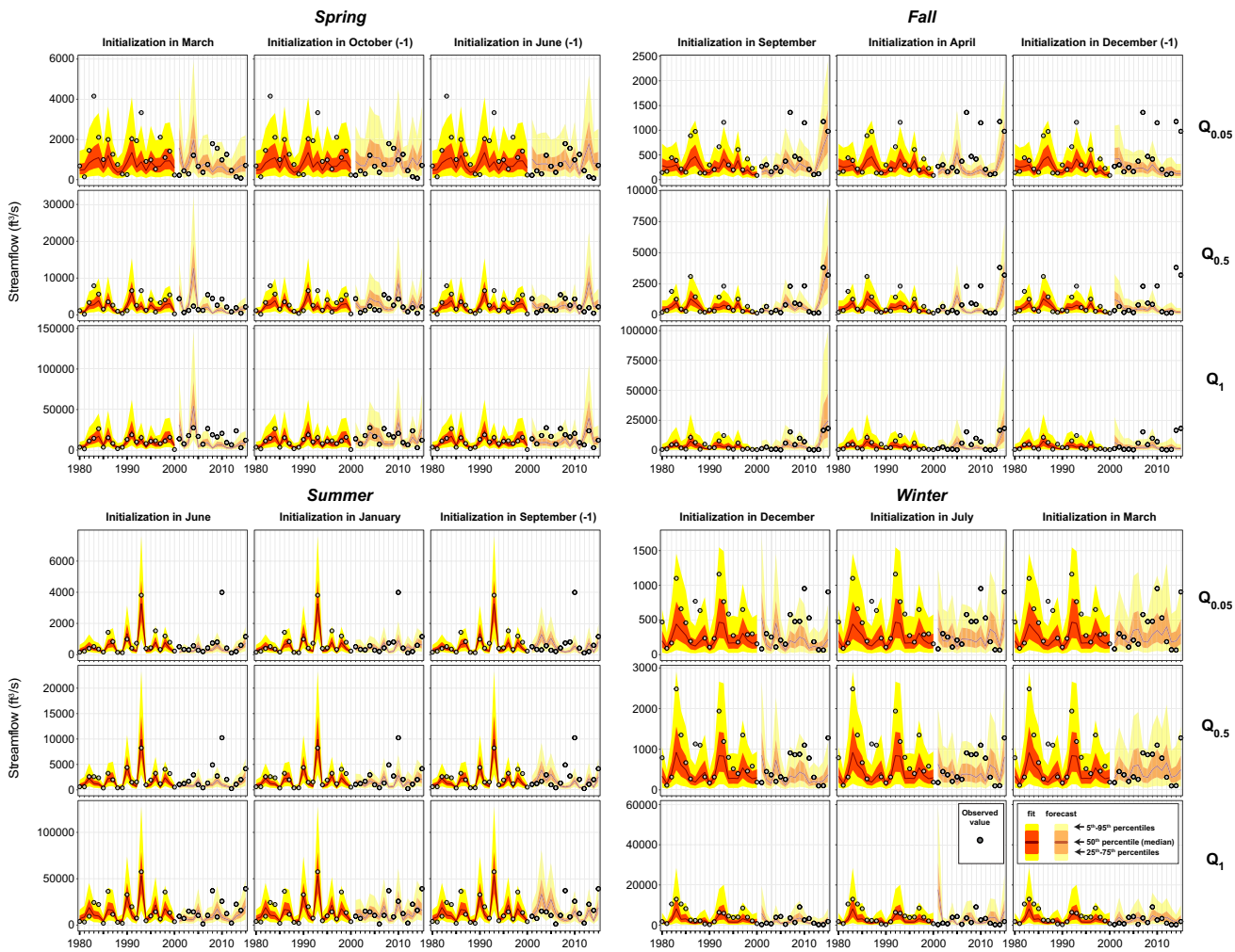

Fig. 6 Time series showing the model fit (1980–2000) and forecast (2001–2015) against the observed values. For every season, and three different initialization times (0.5, 5.5 and 9.5 months ahead of the season), five percentiles of the predicted streamflow distribution are shown (P_5, P_{25}, P_{50}, P_{75} and P_{95}) for three quantiles (*rows*): low flow ($Q_{0.05}$), median flow ($Q_{0.50}$), and maximum seasonal flow (Q_1). The symbology is the same as in Fig. 3. Fitted values are only shown for 1980–2000, even though the models have been fit to the entire period 1927–2016, in order to highlight the detail in the models. Forecast values are shown for 2001–2015; the year 2016 is shown in Fig. 8

the different streamflow quantiles allows us to identify the various strengths and weaknesses of our model, where further improvements can be made to better capture specific processes such as the effect of snowmelt on low to median flows in spring. Thus, future improvements in seasonal flow forecasting may arise from improvements in the input precipitation forecasts and/or in model formulation.

3.3 Visualizing probabilistic forecasts for 2016

Here we illustrate how probabilistic flow forecasts can be visualized and interpreted over 1 year, by comparing our model predictions (background blue-white-red distribution) against the observed value (dashed yellow line) and the observed seasonal average for 2001–2015 (horizontal black line).

The predictions from Model 3 for winter 2016 (*i.e.*, December 2015 through February 2016) show slightly above-average streamflows: the median of the predicted distribution (white bar centered on P_{45}–P_{55}) generally lies above the observed seasonal average for 2001–2015 (see the top row in Fig. 8). However, the observed value is considerably higher than the seasonal average: although we correctly predicted wetter-than-average conditions, were unable to correctly predict the magnitude of flows. In the spring months, we correctly predicted above-average low/median flows ($Q_{0.05}$/$Q_{0.50}$), and high (Q_1) flows roughly on par with the historical average. In the summer months, our forecasts lay below the seasonal average; again we correctly predicted lower-than-average streamflows. In the fall months, however, our model incorrectly predicted much lower-than average streamflows.

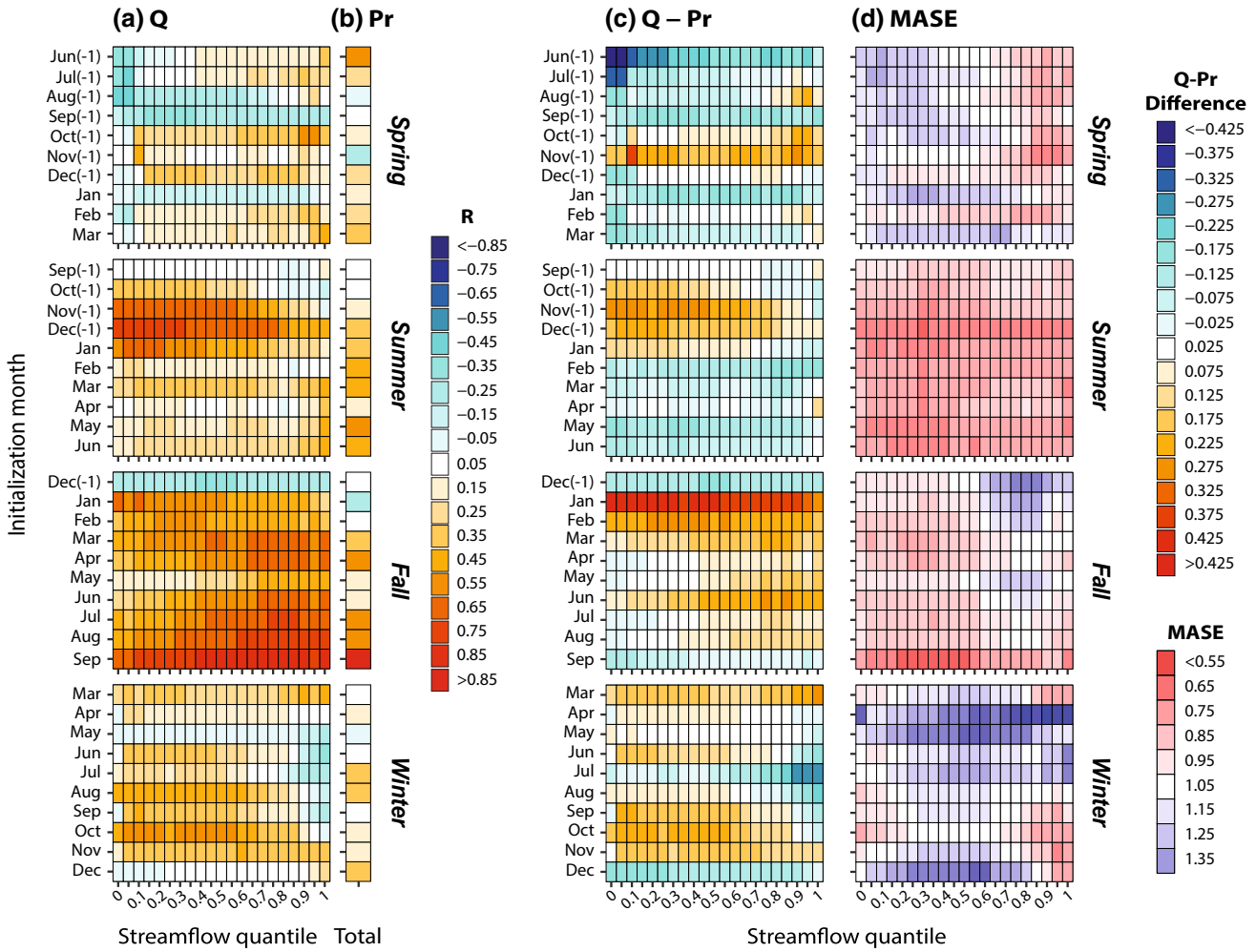

Fig. 7 Model 3 forecast skill evaluation (2001–2016). **a** Correlation coefficient (R) for Model 3 streamflow forecast (*Q*); **b** R for the NMME precipitation forecast (*Pr*); **c** Difference between *Q* and *Pr*; and **d** MASE skill for the Model 3. The skill is shown for all streamflow quantiles (x-axis) and initialization months (lead time ahead of each season; y-axis), for all four seasons (*rows*). In panels a, b, and d, the *red colors* indicate positive skill (strong R or low MASE test results supporting the notion that the forecasts are neither too narrow nor too wide). In panel c, colors range from negative (*blue*: where precipitation skill exceeds streamflow forecast skill) to positive (*red*: where the streamflow forecast skill exceeds the initial precipitation skill)

Overall, Fig. 8 reveals that our model correctly predicted whether the winter, spring and summer flow would be above- or below-average, but incorrectly forecasted fall streamflow as below-average. Most interestingly, this visualization reveals that there is not a decrease in forecast accuracy with initialization time (i.e., forecasts issued further ahead of the season do not uniformly decrease in skill), suggesting that one could consider the whole outlook (all initialization months) to gain a better overview of the seasonal flow forecast.

Given the recent improvements in seasonal forecast skill, how likely are water managers to use such probabilistic forecasts? Over a decade ago, Rayner et al. (2005) highlighted some of the barriers to their uptake, including a conservative management approach to risk, mismatch of forecast temporal/spatial scales with management needs, and barriers to implementation/interpretation. Our approach allows for greater spatial/temporal flexibility than previous methods, but is already somewhat complex, even when showing just three of the 21 flow quantiles (Fig. 8). These points raise a number of questions: how can we make the information more accessible and useable? What is the appetite for detailed flow regime quantiles when they are so uncertain, and are other approaches more suited? Our findings suggest that there is considerable potential for skillful dynamical forecasting of seasonal streamflow quantiles, by further improving the predictors, models, and forecast ensembling schemes.

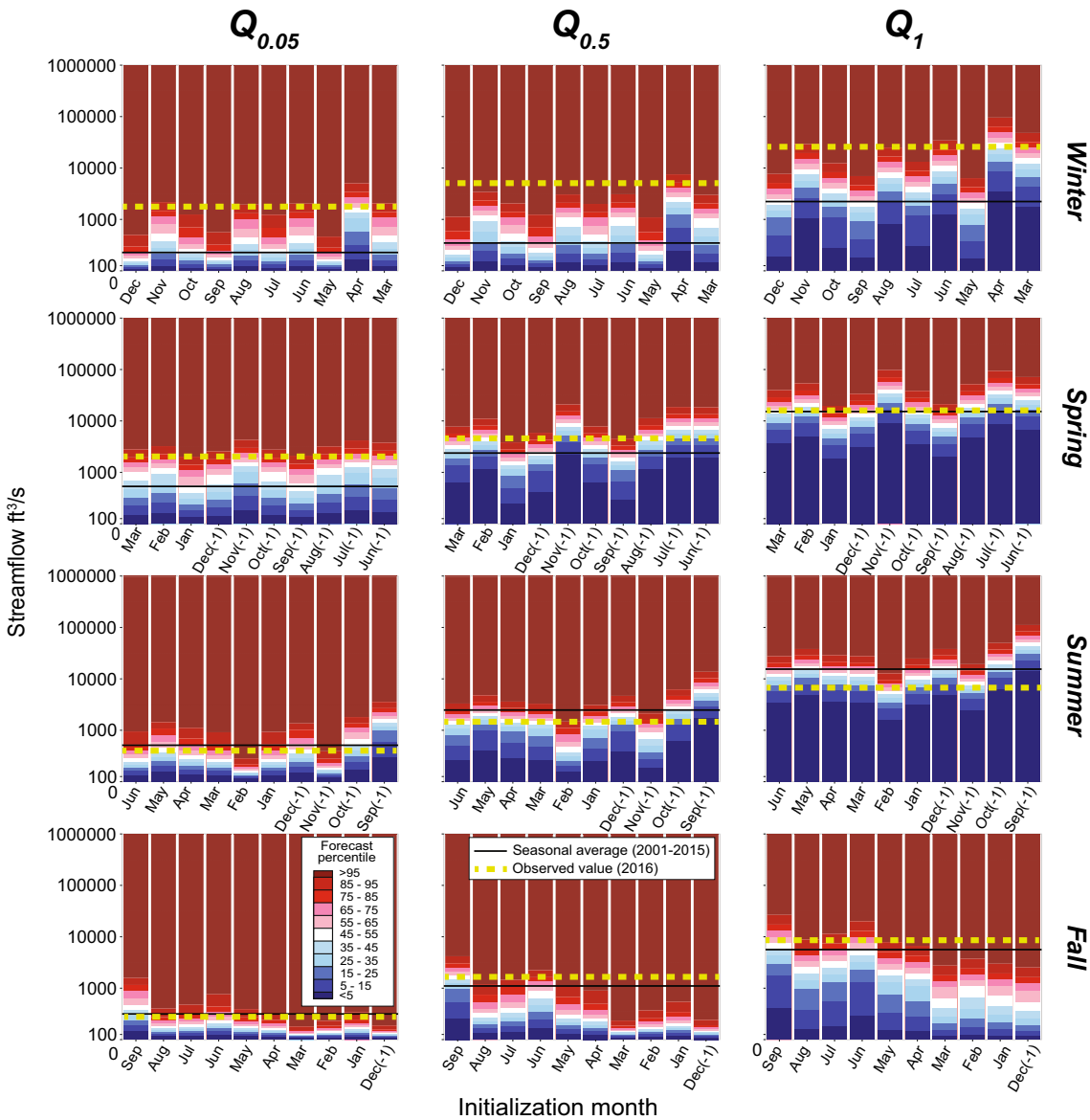

Fig. 8 Seasonal forecast for the Raccoon River in 2016 (Model 3). The x-axes indicate the initialization month of the seasonal forecast, ranging from the shortest lead time (*left*) to the longest lead time (*right*). For example, complete forecasts for the fall season (September, October, November), are issued in December of the previous year (11.5 months ahead), or January (10.5 months ahead). The y-axes indicate percentiles of the forecast flow distribution with a *color* gradient ranging from low (*blue shades*), median (*white shades*), to high (*red shades*). *Darker shades* thus indicate the extremes of the probabilistic forecast. On each sub-panel, the *horizontal black line* indicates the seasonal average flow for the period 2001–2015, while the *horizontal yellow line* indicates the observed flow for 2016. By comparing our forecast distribution against these two lines, we can determine whether above- or below-average conditions were correctly predicted

4 Summary and conclusions

How good are the statistical models that we have developed in describing and forecasting seasonal streamflow variations for the full range of flow quantiles? Results indicate that the three gamma models do describe the inter-annual flow variability accurately when they are fitted with observed precipitation and agricultural acreage. All three models have well-distributed residuals across all streamflow quantiles, especially in the summer and spring (the most variable seasons), suggesting that the high flows tend to be as well predicted as the low flows. There is not one model that consistently outperforms the others across all seasons, although Model 3 (which includes a proxy for antecedent wetness) produces a markedly better fit than the other models in the fall and for the low flows, when initial

conditions have the strongest effect on the flow distribution. These findings indicate that the relative importance of each predictor varies seasonally and by flow quantile, suggesting that forecast skill can be improved by leveraging the strengths of different predictors when they are most useful.

In terms of seasonal forecasts, we find that the forecast flow distributions (based on NMME precipitation forecasts) perform similarly to the historical fitted distributions across flow quantiles, but for certain initialization months only. Most surprisingly, the forecast skill does not decrease uniformly with lead-time, and some of the most skillful forecasts are issued 1.5, 5.5, or 6.5 months (for winter, summer and fall respectively) ahead of the predicted season. The skill comes predominantly from the precipitation forecasts, suggesting that improvements will arise principally from more sophisticated combinations of the individual NMME members.

Overall, seasonal flow forecast accuracy is improved by the inclusion of antecedent precipitation, but agricultural land cover only has minimal effect on forecast skill. This lack of improvement may be related to the quality of the land cover forecast (agricultural data are published with a certain latency, so we use agricultural coverage from the previous year); real-time data or other land cover categories may allow for a more precise evaluation of changes in the flow distribution.

Though our probabilistic seasonal streamflow forecasts are generated with simple statistical models, they may have considerable practical applications. Real-time forecasts of soil moisture and streamflow based on hydrologic macroscale models such as NMME-VIC are already being linked with impact models to predict reservoir inflow, crop yield, and wild fire (Yuan et al. 2015). This methodology is sufficiently general to be applicable across the central United States, where streamflow is affected by both climate and agricultural practices. Real-time applications of the model could include, for example, predicting the impacts of changing land cover on flow. Such outputs would enable legislators to envisage a potential trade-off between taking agricultural land out of production and minimizing the risks arising from extreme floods. More generally, real-time forecasts would provide basic information allowing decision-makers to prepare for and mitigate the disruptions arising from these catastrophic events. Similarly, long-term forecasts issued over annual to decadal timeframes with more elaborate multi-model ensembles would also enable management decisions in terms of, e.g., water resources or crop legislation.

Acknowledgements The authors thank the NMME program partners and acknowledge the help of NCEP, IRI and NCAR personnel in creating, updating and maintaining the NMME archive, with the support of NOAA, NSF, NASA and DOE. Two anonymous Reviewers are also thanked for comments that helped improve the quality of the manuscript. This study was supported in part by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections Program, Grant #NA15OAR4310073, by the Broad Agency Announcement (BAA) Program and the Engineer Research and Development Center (ERDC)–Cold Regions Research and Engineering Laboratory (CRREL) under Contract No. W913E5-16-C-0002, and by Grant/Cooperative Agreement Number G11 AP20079 from the United States Geological Survey.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Asseng S, McIntosh PC, Thomas G, Ebert EE, Khimashia N (2016) Is a 10-day rainfall forecast of value in dry-land wheat cropping?. Agric For Meteorol 216:170–176. doi:10.1016/j.agrformet.2015.10.012

Barnston AG, Lyon B (2016) Does the NMME capture a recent decadal shift toward increasing drought occurrence in the southwestern US? J Clim 29(2):561–581. doi:10.1175/JCLI-D-15-0311.1

Barnston AG, Tippett MK, Van den Dool HM, Unger DA (2015). Toward an improved multi-model ENSO prediction. J Appl Meteorol Climatol 54(7):1579–1595. doi:10.1175/JAMC-D-14-0188.1

Becker E, Van den Dool H, Zhang Q (2014). Predictability and Forecast Skill in NMME. J Clim 27(15):5891–5906. doi:10.1175/JCLI-D-13-00597.1

Bogner K, Liechti K, Zappa M (2017) Technical Note: combining quantile forecasts and predictive distributions of stream-flows. Hydrol Earth Syst Sci Discuss. doi:10.5194/hess-2017-291

Daly C, Gibson WP, Taylor GH, Johnson GL, Pasteris P (2002) A knowledge-based approach to the statistical mapping of climate. Clim Res 22(9):99–113

Day GN (1985) Extended Streamflow Forecasting Using NWSRFS. J Water Resour Plann Manage 111(2):157–170. doi:10.1061/(ASCE)0733-9496(1985)111:2(157)

Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Coke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. J Clim 19(5):643–674. doi:10.1175/JCLI3629.1

Filliben JJ (1975) The probability plot correlation coefficient test for normality. Technometrics 17(1):111. doi:10.2307/1268008

Frans C, Istanbulluoglu E, Mishra V et al (2013) Are climatic or land cover changes the dominant cause of runoff trends in the Upper Mississippi River Basin?. Geophys Res Lett 40:1104–1110. doi:10.1002/grl.50262

Franses PH (2016). A note on the mean absolute scaled error. Int J Forecast 32(1):20–22. doi:10.1016/j.ijforecast.2015.03.008

Gupta SC, Kessler AC, Brown MK, Zvomuya F (2015) Climate and agricultural land use change impacts on streamflow in the

upper midwestern United States. Water Resour Res 6:446. doi:10.1016/0022-1694(68)90080-2

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. Tellus Ser A Dyn Meteorol Oceanogr 57(3), 219–233. doi:10.1111/j.1600-0870.2005.00103.x

Ham Y-G, Kug J-S (2015) Improvement of ENSO simulation based on intermodel diversity. J Clim 28(3):998–1015. doi:10.1175/JCLI-D-14-00376.1

Harrison M, Troccoli A, Coughlan M, Williams JB (2007) Seasonal forecasts in decision making. In: Troccoli A, Harrison M, Andersen D.L.T., Mason SJ (eds) Seasonal climate: forecasting and managing risk. Springer, Dordrecht, 13–44. doi:10.1007/978-1-4020-6992-5_2

Hodyss D, Satterfield E, McLay J, Hamill TM, Scheuerer M (2016) Inaccuracies with multimodel postprocessing methods involving weighted, regression-corrected forecasts. Mon Weather Rev 144:1649–1668

Hyndman RJ, Koehler AB (2006). Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688. doi:10.1016/j.ijforecast.2006.03.001

Infanti JM, Kirtman BP (2014) Southeastern U.S. rainfall prediction in the North American Multi-Model Ensemble. J Hydrometeorol 15(2):529–550. doi:10.1175/JHM-D-13-072.1

Jain S, Lall U (2001). Floods in a changing climate: Does the past represent the future?. Water Resour Res 37(12):3193–3205. doi:10.1029/2001WR000495

Kirtman BP, Min D (2009) Multimodel ensemble ENSO prediction with CCSM and CFS. Mon Weather Rev 137(9):2908–2930. doi:10.1175/2009MWR2672.1

Kirtman BP et al (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteorol Soc 95(April):585–601. doi:10.1175/BAMS-D-12-00050.1

Larson SM, Kirtman BP (2014) The Pacific meridional mode as an ENSO precursor and predictor in the North American multimodel ensemble. J Clim 27(18):7018–7032. doi:10.1175/JCLI-D-14-00055.1

Lawrence DM, Oleson KW, Flanner MG, Fletcher CG, Lawrence PJ, Levis S, Swenson SC, Bonan GB (2012) The CCSM4 land simulation, 1850–2005: assessment of surface climate and new capabilities. J Clim 25(7):2240–2260

Liang X, Lettenmaier DP, Wood EF, Burges SJ (1994) A simple hydrologically based model of land surface water and energy fluxes for GSMs. J Geophys Res 99(D7):14415–14428. doi:10.1029/94JD00483

Liang X, Wood EF, Lettenmaier DP (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications. Global Planet Change 13:195–206. doi:10.1016/0921-8181(95)00046-1

Lopez H, Kirtman BP (2014) WWBs, ENSO predictability, the spring barrier and extreme events. J Geophys Res Atmos 119:10114–10138. doi:10.1002/2014JD021908

Ma F, Ye A, Deng X, Zhou Z, Liu X, Duan Q, Gong W (2015a) Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China. Int J Climatol. doi:10.1002/joc.4333

Ma F, Yuan X, Ye A (2015b) Seasonal drought predictability and forecast skill over China. J Geophys Res 120(16):8264–8275. doi:10.1002/2015JD023185

Mason SJ, Baddour O (2008) Statistical modelling. In: Troccoli A, Harrison M, Andersen DLT, Mason SJ (eds) Seasonal Climate: Forecasting and Managing Risk. Springer, Dordrecht, 163–201

Merryfield WJ, Lee W-S, Boer GJ, Kharin VV, Scinocca JF, Flato GM, Ajayamohan RS, Fyfe JC, Tang Y, Polavarapu S (2013) The Canadian seasonal to interannual prediction system. Part I: models and initialization. Mon Weather Rev 141(8):2910–2945. doi:10.1175/MWR-D-12-00216.1

Misra V, Li H (2014) The seasonal climate predictability of the Atlantic Warm Pool and its teleconnections. Geophys Res Lett 41:661–666. doi:10.1002/2013GL058740

Mo KC, Lettenmaier DP (2014) Hydrologic prediction over Conterminous U.S. using the National Multi Model ensemble. J Hydrometeorol. doi:10.1175/JHM-D-13-0197.1

Mo KC, Lyon B (2015) Global Meteorological drought prediction using the North American multi-model ensemble. J Hydrometeorol. doi:10.1175/JHM-D-14-0192.1

Mo KC, Shukla S, Lettenmaier DP, Chen LC (2012) Do climate forecast system (CFSv2) forecasts improve seasonal soil moisture prediction? Geophys Res Lett 39(23):1–6. doi:10.1029/2012GL053598

Molod A, Takacs L, Suarez M, Bacmeister J, Song I-S, Eichmann A (2012) The GEOS-5 atmospheric general circulation model: mean climate and development from MERRA to Fortuna. In: Technical report series on global model data assimilation, vol 28. NASA Goddard Space Flight Cent., Greenbelt, p 175

NASS, USDA (2015) Census of agriculture. US Department of Agriculture, National Agricultural Statistics Service, Washington, DC

Rayner S, Lach D, Ingram H (2005) Weather forecasts are for wimps: why water resource managers do not use climate forecasts. Clim Change 69:197–227. doi:10.1007/s10584-005-3148-z

Rigby RA, Stasinopoulos DM (2005). Generalized additive models for location, scale and shape. J R Stat Soc Ser C (Appl Stat) 54:507–554. doi:10.1111/j.1467-9876.2005.00510.x

Rogger M, Agnoletti M, Alaoui A et al (2017) Land-use change impacts on floods at the catchment scale—challenges and opportunities for future research. Water Resour Res. doi:10.1002/2017WR020723

Roundy JK, Yuan X, Schaake J, Wood EF (2015) A framework for diagnosing seasonal prediction through canonical event analysis. Mon Weather Rev 143(6):2404–2418. doi:10.1175/MWR-D-14-00190.1

Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou Y-T, Chuang H-Y, Iredell M, Ek M, Meng J, Yang R, Peña Mendez M, van den Dool H, Zhang Q, Wang W, Chen M, Becker E (2014) The NCEP climate forecast system version 2. J Clim 27(6):2185–2208. doi:10.1175/JCLI-D-12-00823.1

Sankarasubramanian A, Lall U (2003) Flood quantiles in a changing climate: seasonal forecasts and causal relations. Water Resour Res 39(5):1134

Scheuerer M, Büermann L (2014) Spatially adaptive post-processing of ensemble forecasts for temperature. J R Stat Soc C 63(3):405–422

Schilling KE, Jha MK, Zhang Y-K, Gassman PW, Wolter CF (2008). Impact of land use and land cover change on the water balance of a large agricultural watershed: historical effects and future directions. Water Resour Res 44:1–12. doi:10.1029/2007WR006644

Schottler SP, Ulrich J, Belmont P et al. (2014) Twentieth century agricultural drainage creates more erosive rivers. Hydrol Process 28:1951–1961. doi:10.1002/hyp.9738

Shrestha RR, Schnorbus MA, Cannon AJ (2015). A dynamical climate model–driven hydrologic prediction system for the Fraser River, Canada. J Hydrometeorol 16(3):1273–1292. doi:10.1175/JHM-D-14-0167.1

Sikder MS, Chen X, Hossain F, Roberts JB, Robertson F, Shum CK, Turk FJ (2015) Are General Circulation Models Ready for Operational Streamflow Forecasting for Water Management in Ganges and Brahmaputra River basins? J Hydrometeorol 17(1):195–210. doi:10.1175/JHM-D-14-0099.1

Slater LJ, Villarini G, Bradley AA (2017) Evaluation of the skill of North-American multi-model ensemble (NMME) global climate

models in predicting average and extreme precipitation and temperature over the continental USA. Clim Dyn. doi:10.1007/s00382-016-3286-1

Smith A, Matthews J (2015) Quantifying uncertainty and variable sensitivity within the U.S. Billion-dollar weather and climate disaster cost estimates. Nat Hazards 77(3):1829–1851. doi:10.1007/s11069-015-1678-x

Stasinopoulos DM, Rigby RA, Akantziliotou C (2006) Instructions on how to use the GAMLSS package in R. Accompanying documentation in the current GAMLSS help files (see also http://www.gamlss.org/. Accessed 30 Jan 2016)

Stasinopoulos M, Rigby B, with contributions from Akantziliotou C, Heller G, Ospina R, Motpan N, McElduff F, Voudouris V, Djennad M, Enea M, Ghalanos A (2016) gamlss.dist: distributions to be used for GAMLSS modelling. R package version 5.0-0. https://CRAN.R-project.org/package=gamlss.dist

Thober S, Kumar R, Sheffield J, Mai J, Schäfer D, Samaniego L (2015) Seasonal soil moisture drought prediction over Europe using the North American multi-model ensemble (NMME). J Hydrometeorol 16(6):2329–2344. doi:10.1175/JHM-D-15-0053.1

Towler E, Rajagopalan B, Gilleland E, Summers RS, Yates D, Katz RW (2010) Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. Water Resour Res 46:W11504. doi:10.1029/2009WR008876

Twedt T, Schaake JJ, Peck E (1977) National weather service extended streamflow prediction. In: Proceedings 45th annual western snow conference, Albuquerque, NM, Western Snow Conference, 52–57

Vecchi GA, Delworth T, Gudgel R, Kapnick S, Rosati A, Wittenberg A, Zeng F, Anderson W, Balaji V, Dixon K, Jia L, Kim H-S, Krishnamurthy L, Msadek R, Stern WF, Underwood SD, Villarini G, Yang X, Zhang S (2014) On the seasonal forecasting of regional tropical cyclone activity. J Clim 27(21):7994–8016. doi:10.1175/JCLI-D-14-00158.1

Vernieres G, Rienecker MM, Kovach R, Keppenne CL (2012) The GEOS-iODAS: description and evaluation. In: GEOS5 technical report NASA/TM-2012-104606, vol 30. 61 pp. Available online at http://gmao.gsfc.nasa.gov/pubs/docs/Vernieres589.pdf

Villarini G, Serinaldi F (2012) Development of statistical models for at-site probabilistic seasonal rainfall forecast. Int J Climatol 32(14):2197–2212 doi:10.1002/joc.3393

Villarini G, Strong A (2014), Roles of climate and agricultural practices in streamflow changes in an agricultural watershed in Iowa. Agric Ecosyst Environ 188:204–211. doi:10.1016/j.agee.2014.02.036

Wanders N, Wood EF (2016) Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. Environ Res Lett 11:94007. doi:10.1088/1748-9326/11/9/094007

Wang H (2014) Evaluation of monthly precipitation forecasting skill of the National Multi-model Ensemble in the summer season. Hydrol Processes 28:4472–4486. doi:10.1002/hyp.9957

Wood AW, Hopson T, Newman A, Brekke L, Arnold J, Clark M (2016) Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. J Hydrometeorol 17(2):651–668. doi:10.1175/JHM-D-14-0213.1

Yuan X, Wood EF, Roundy JK, Pan M (2013) CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States. J Clim 26(13):4828–4847. doi:10.1175/JCLI-D-12-00683.1

Yuan X, Roundy JK, Wood EF, Sheffield J (2015) Seasonal forecasting of global hydrologic extremes: system development and evaluation over GEWEX basins. Bull Am Meteorol Soc 96(11):1895–1912. doi:10.1175/BAMS-D-14-00003.1

Zhang YK, Schilling KE (2006) Increasing streamflow and baseflow in Mississippi River since the 1940s: effect of land use change. J Hydrol 324:412–422. doi:10.1016/j.jhydrol.2005.09.033

Zhang S, Harrison MJ, Rosati A, Wittenberg A (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. Mon Weather Rev 135(10):3541–3564. doi:10.1175/MWR3466.1