AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER

10.1029/2018GL077945

Enhancing the Predictability of Seasonal Streamflow With a Statistical-Dynamical Approach

Louise J. Slater[1] **and Gabriele Villarini**[2]

[1]Department of Geography, Loughborough University, Loughborough, UK, [2]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA

Key Points:

- A statistical-dynamical framework is developed to predict seasonal streamflow 0.5–9.5 months ahead using NMME climate model forecasts
- Streamflow predictability is enhanced in all seasons by including forecasts of climate/antecedent climatic conditions as model predictors
- The streamflow forecasts are improved in certain catchments/seasons by including predictors reflecting land cover and population density

Supporting Information:

- Supporting Information S1

Correspondence to:
L. J. Slater,
l.slater@lboro.ac.uk

Citation:
Slater, L. J., & Villarini, G. (2018). Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach. *Geophysical Research Letters*, *45*, 6504–6513. https://doi.org/10.1029/2018GL077945

Received 19 MAR 2018
Accepted 10 JUN 2018
Accepted article online 19 JUN 2018
Published online 6 JUL 2018

Abstract Seasonal streamflow forecasts facilitate water allocation, reservoir operation, flood risk management, and crop forecasting. They are generally computed by forcing hydrological models with outputs from general circulation models (GCMs) or using large-scale climate indices as predictors in statistical models. In contrast, hybrid statistical-dynamical forecasts (combining statistical methods with dynamical climate predictions) are still uncommon, and their skill is largely unknown. Here we conduct systematic forecasting of seasonal streamflow using eight GCMs from the North American Multi-Model Ensemble, 0.5–9.5 months ahead, at 290 stream gauges in the U.S. Midwest. Probabilistic forecasts are developed for low to high streamflow using predictors that reflect climatic and anthropogenic influences. Results indicate that GCM forecasts of climate and antecedent climatic conditions enhance seasonal streamflow predictability; while land cover and population density predictors decrease biases or enhance skill in certain catchments. This paper paves the way for novel forecasting approaches using dynamical GCM predictions within statistical frameworks.

Plain Language Summary Streamflow forecasts several months ahead of a season are important for water management and the prevention of risks related to floods and hydrological droughts. However, existing methods for producing seasonal streamflow forecasts are often complex and computationally intensive. Here we provide a systematic evaluation of a *statistical-dynamical* approach to streamflow forecasting in several hundred river catchments across the U.S. Midwest. We assess whether global climate model forecasts can be used as predictors in statistical models to produce skillful forecasts of river flow, up to 10 months ahead. Results indicate that forecasts of rainfall and temperature, antecedent climatic conditions, as well as information on population density and land cover, can be used effectively to forecast streamflow at seasonal time scales. By including information on the future antecedent climatic conditions, streamflow forecasts can be enhanced months ahead. Information on human influences, in contrast, helps reduce the biases in the streamflow forecasts. These results pave the way for statistical-dynamical forecasting in catchments around the world and suggest that process-driven combinations of different predictors can be used to produce skillful streamflow forecasts in different seasons, for both high flows (i.e., floods) and low flows (i.e., representative of hydrological droughts).

1. Introduction

Streamflow forecasts established at subseasonal to seasonal time scales (i.e., weeks to months ahead) can be beneficial in many sectors, including water resources allocation, reservoir operation and management, flood risk planning, navigation, and agricultural crop forecasting. In Europe, for example, the estimated monetary benefit of early flood warnings provided by the continental-scale European Flood Awareness System is 400 Euro for every 1 Euro invested (Pappenberger, Cloke et al., 2015). There are principally two types of approaches: dynamical approaches (where weather/climate information is used to drive land surface hydrological models) and statistical approaches (using large-scale climate information as predictors in statistical models).

In dynamical or physically based forecasting, hydrological models are typically forced with outputs from general circulation models (GCMs; e.g., Yuan, Wood, & Ma, 2015) or ensembles of numerical weather predictions (e.g., Cloke & Pappenberger, 2009). Subseasonal to seasonal streamflow forecasts are now produced at continental to global scales (e.g., Alfieri et al., 2013; Arnal et al., 2017; Emerton et al., 2016; Thielen et al., 2009; Yuan, Roundy, et al., 2015) and implemented operationally (e.g., Global Flood Awareness System; Emerton et al., 2018). Alternatively, land surface hydrological models can also be initialized with initial hydrologic conditions and driven with historical meteorological data, as in the case of Ensemble Streamflow Prediction (Day, 1985; Harrigan et al., 2018; Van Dijk et al., 2013; Werner et al., 2004; Wood & Schaake, 2008; Yuan et al., 2016), when no dynamical seasonal forecasts are available.

In statistical or time series forecasting, the hydrological predictands are generally regressed on observational precipitation/temperature records, large-scale climate indices, climatic teleconnection patterns, and/or information on initial hydrologic conditions (e.g., Mendoza et al., 2017; Robertson et al., 2013). For example, streamflow has been successfully forecast with large-scale ocean-atmospheric predictors such as sea surface temperature, surface air temperature, geopotential height, meridional wind, or a range of climate modes of variability like the El Niño–Southern Oscillation (e.g., Regonda et al., 2006; Souza Filho & Lall, 2003). However, these stationary statistical relationships are not always reliable, due to changing patterns of flow seasonality, shifts in the relationship between streamflow and the climatic predictors (e.g., Cayan et al., 2001; Mote et al., 2005), and temporal instabilities in some of the predictors (e.g., Ionita et al., 2008; Pagano et al., 2004).

In contrast with both of these long-standing approaches, the coupling of climate predictions with statistical models, hereafter referred to as statistical-dynamical forecasting, is new (here we use the term *dynamical* to refer to dynamical climate predictions, rather than to dynamic hydrological models). Hybrid statistical-dynamical models have different strengths: they are computationally efficient, can integrate and *learn* from a broad selection of nonstationary input data (e.g., dynamical GCM forecasts, Earth Observation time series, and teleconnections), and can benefit from recent statistical progress (e.g., multimodel blending and post-processing). The implementation of multimodel systems such as the Copernicus Climate Change Service or the North American Multi-Model Ensemble (NMME) has facilitated the development of such approaches. In the NMME project, participating North American modeling centers (Table S1 in the supporting information) now provide publicly accessible global climate hindcasts and forecasts at 1° latitude by 1° longitude resolution, with lead times ranging from 0.5 to 11.5 months (Kirtman et al., 2014). Yet the few papers that have investigated the use of the NMME for streamflow forecasting have mostly used hydrological models such as the variable infiltration capacity (VIC) model (e.g., Mo & Lettenmaier, 2014; Yuan, Roundy, et al., 2015; Yuan, Wood, & Ma, 2015) or other land surface models (e.g., Thober et al., 2015).

Thus, one of the current challenges in seasonal streamflow forecasting is to assess the predictability of streamflow using a statistical-dynamical methodology in different catchments around the globe, in comparison with the existing operational forecasting approaches. Until now, the predictability arising from the inclusion of temperature or land cover in statistical-dynamical models has only been evaluated in a handful of carefully selected, snowmelt-dominated (Lehner et al., 2017) or agricultural (Slater, Villarini, Bradley, Vecchi, 2018) river catchments.

Here we develop a statistical-dynamical framework to forecast seasonal streamflow between 0.5 and 9.5 months ahead, using NMME outputs from 94 GCM members, in 290 river catchments with widely varying physical, climatic, and land cover characteristics. We investigate (1) the predictability of streamflow across the Midwestern USA, using just dynamical NMME precipitation forecasts; (2) to what extent this predictability can be enhanced by including other climatic and *anthropogenic* predictors; and (3) how the predictability varies across different catchments, seasons, and streamflow quantiles.

2. Data and Methods

2.1. Data Sets of Observed Streamflow, Precipitation, Temperature, Agriculture, and Population Density

We develop our forecasting framework in the Midwest because the drivers of streamflow variability are relatively well understood in the region (e.g., Slater & Villarini, 2017). The climate exhibits average annual temperatures ranging from less than 3 °C in northern Minnesota to over 15 °C in southeastern Missouri and average annual precipitation from ~500 mm to ~1200 mm in the same areas (Andresen et al., 2012). The Midwest has also witnessed a notable increase in precipitation (Alter et al., 2018) and flooding (Mallakpour & Villarini, 2015) in recent decades.

Daily streamflow data are obtained from the U.S. Geological Survey's (USGS) National Water Information System for all stream gauges in the 12 U.S. Midwestern states with a continuous record of at least 50 years before 2015 (Figure 1). The size of the 290 catchments varies from 20 km^2 to 153,327 km^2. We retain

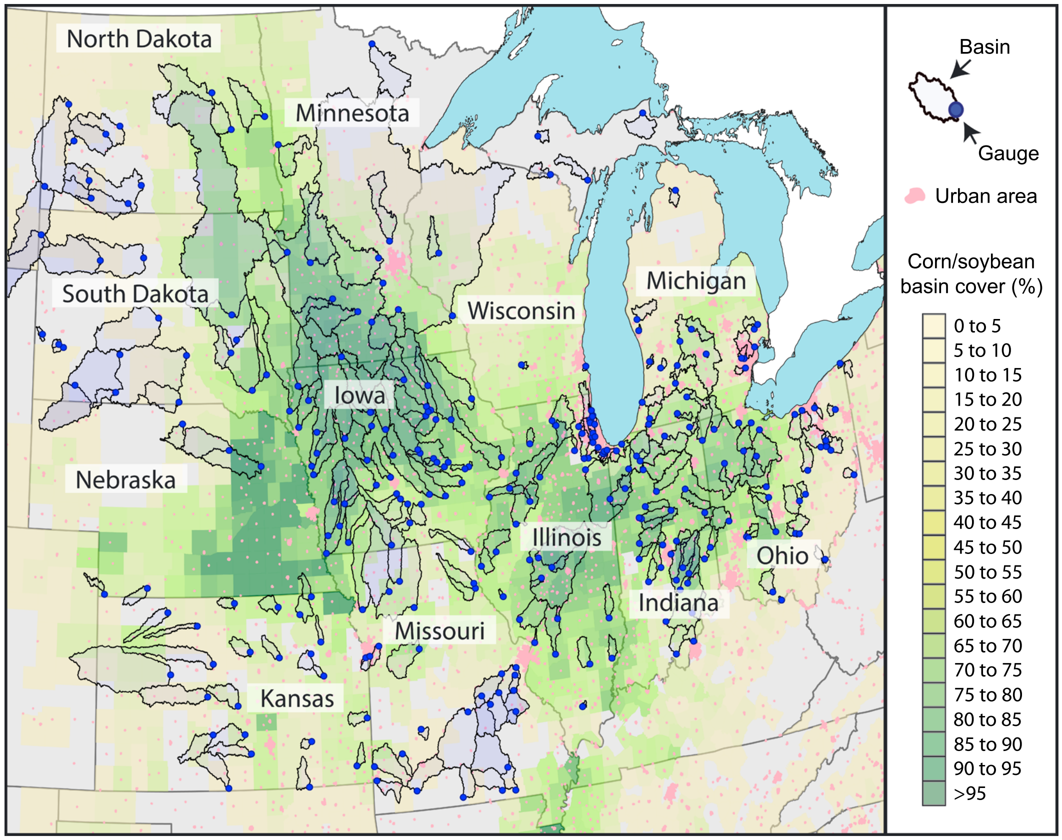

Figure 1. Location of the 290 catchments across the Midwest. Stream gauges are indicated as blue circles, and catchments are indicated as black outlines. The agricultural extent in 2014 within each county is indicated in shades of green, and urban areas are indicated in pink.

catchments where the peak flows are not notably affected by regulation or diversion (absence of flags *5* or *6* in the USGS peak flows database). We also retain catchments that have witnessed urban development, to test the influence of population density as a predictor. The catchments are the same as those used in Slater and Villarini (2017) to assess the ability of our model across sites with variable physical, climatic, and streamflow characteristics. Observed streamflow (Q) quantiles are computed from daily records at every site and for every season, ranging from Q_0 (minimum flow) to Q_1 (maximum flow). For example, if our predictand is summer Q_1, then we compute the maximum of the daily streamflow distribution for the 3 months ranging from June until August and repeat this for every year from 1983 to 2015.

Observed monthly precipitation (millimeter) and temperature (degrees Celsius) data at approximately 4-km resolution are obtained from the PRISM climate group (Daly et al., 2002) and used to compute monthly catchment-averaged values. Agricultural land cover data are obtained from the U.S. Department of Agriculture's National Agricultural Statistics Services *quickstats* database. Time series of harvested corn and soybean acreage (as a proportion of total land cover within each catchment) are computed by aggregating county-level data (see, e.g., Villarini & Strong, 2014) and vary from near 0% to 92%. Population density statistics are from the U.S. Census Bureau's Population Estimates Program. Time series are computed to reflect the average population density in each catchment, which varies from 0.6 pers/km^2 to 1339 pers/km^2 (see Slater & Villarini, 2017). Predictions of agricultural land cover and population density are computed on an annual basis using the last available year's estimate (persistence forecast; Figure S1), which proved to be more skillful than forecasts based on other simple methods.

2.2. Seasonal Precipitation and Temperature NMME Forecasts

Monthly NMME forecasts of precipitation and temperature are available every month and for lead times of between 0.5 and 11.5 months. We use eight NMME GCMs with a total of 94 model members (CanCM3,

Table 1
Formulation of the Six Statistical Models

Model acronym	Model formulation
P	$\begin{cases} \log(\mu_1) = \alpha_1 + \beta_1 \cdot x_p \\ \log(\sigma_1) = \kappa_1 \end{cases}$
P + *T*	$\begin{cases} \log(\mu_2) = \alpha_2 + \beta_2 \cdot x_p + \gamma_2 \cdot x_t \\ \log(\sigma_2) = \kappa_2 \end{cases}$
P + *AntP*	$\begin{cases} \log(\mu_3) = \alpha_3 + \beta_3 \cdot x_p + \gamma_3 \cdot x_{\mathrm{ap}} \\ \log(\sigma_3) = \kappa_3 \end{cases}$
P + *AntT*	$\begin{cases} \log(\mu_4) = \alpha_4 + \beta_4 \cdot x_p + \gamma_4 \cdot x_{\mathrm{at}} \\ \log(\sigma_4) = \kappa_4 \end{cases}$
P + *Ag*	$\begin{cases} \log(\mu_5) = \alpha_5 + \beta_5 \cdot x_p + \gamma_5 \cdot x_{\mathrm{ag}} \\ \log(\sigma_5) = \kappa_5 \end{cases}$
P + *Pop*	$\begin{cases} \log(\mu_6) = \alpha_6 + \beta_6 \cdot x_p + \gamma_6 \cdot x_{\mathrm{pop}} \\ \log(\sigma_6) = \kappa_6 \end{cases}$

Note. The σ parameter does not depend on predictors.

CanCM4, CCSM3, CCSM4, CFSv2, GFDL2.1, FLORb01, and GEOS5; see Table S1 for details). Catchment-averaged time series of precipitation and temperature outputs are computed for every site, month, and lead time, for each of the 94 NMME members (Kirtman et al., 2014). We compute the catchment ensemble forecast as the mean of the available members (for every site, month, and lead time) because the multimodel mean forecast tends to outperform any single model (Becker et al., 2014; Krakauer, 2017; Slater, Villarini, & Bradley, 2018). The seasonal precipitation and temperature forecasts are computed by aggregating monthly forecasts for every initialization time (e.g., the spring forecast issued in March is the sum of the 0.5-lead forecast for March, the 1.5-lead forecast for April, and the 2.5-lead forecast for May; Figure S1).

Antecedent precipitation and temperature are computed as the total precipitation and mean temperature in the 3-month period prior to the streamflow forecast (e.g., the winter precipitation is the antecedent precipitation for a spring streamflow forecast). Whether the antecedent precipitation is computed using observations or forecast precipitation depends on the initialization month. For example, for a spring forecast issued in *March*, the observations of December, January, and February precipitation would already be available, so the antecedent precipitation is effectively the observed winter rainfall. For a spring forecast issued in *February*, however (1.5 months ahead), we would not yet have the observed precipitation for February and thus would need to use the 0.5-lead forecast for February.

2.3. Forecasting Approach

Because of the limited length of available GCM data we compute the seasonal streamflow forecasts using a leave-one-out cross-validation approach, where each year is dropped and the model is trained using the rest of the forecasts (e.g., Grantz et al., 2005). For example, to forecast spring streamflow in 2015, we would use the spring GCM forecasts for 1983–2014 to train the model. Our forecasting approach is probabilistic, to better convey forecast uncertainties (e.g., Ramos et al., 2013). We use a gamma distribution with two parameters, μ and σ, that depend linearly on the predictors via a logarithmic link function; the σ parameter is kept constant (i.e., does not depend on predictors) because previous work indicates that it does not enhance model fits (Villarini & Strong, 2014). We implement the Generalized Additive Models for Location, Scale and Shape (GAMLSS) using the gamlss package in *R* (Rigby & Stasinopoulos, 2005). Our *baseline* model uses only precipitation (x_{p}) as predictor (model *P* in Table 1). However, previous work has shown that the inclusion of additional predictors, such as antecedent climatic conditions and land cover, enhances the model fit to seasonal streamflow data (Slater & Villarini, 2017). Therefore, we develop five additional models that all include precipitation and one other predictor: temperature (x_{t}), antecedent precipitation (x_{ap}), antecedent temperature (x_{at}), agricultural land cover (x_{ag}), or population density (x_{pop}). These five models serve to assess whether the inclusion of additional climatic and land cover predictors can enhance the seasonal predictability of streamflow. For brevity, we will refer to these models as *P*, *P* + *T*, *P* + *AntP*, *P* + *AntT*, *P* + *Ag*, and *P* + *Pop* (see Table 1 for model formulations). We consider each predictor separately to evaluate whether it increases the forecast skill for different initialization months.

2.4. Skill Evaluation

We evaluate the forecast skill of our six models using both deterministic and probabilistic evaluation metrics. The deterministic model forecast skill is computed using only the 50th percentile of the probabilistic streamflow forecast, for every season and initialization time. We use four metrics that compose the deterministic mean square error skill score (MSESS): Pearson's correlation coefficient (*R*), the potential skill (PS), and two components that reflect the model biases, the standardized mean error (SME) and the slope reliability (SREL). Please see Murphy and Winkler (1992) for the formulation of these standard model evaluation metrics.

For probabilistic evaluation, we also implement the continuous ranked probability (CRPS) skill score (CRPSS) using the easyVerification package (Weigel & Mason, 2011) in *R*. The CRPS compares the forecast distribution with the observed distribution and is widely recommended for probabilistic forecast evaluation (Pappenberger, Ramos et al., 2015). The CRPSS provides a comprehensive evaluation of flow predictability across the entire probability distribution, penalizing forecasts with large biases or low sharpness; it is increasingly used for benchmarking ensemble forecast skill relative to a baseline forecast (e.g., Arnal et al., 2017; Harrigan et al., 2018). Here we compute the CRPSS of the forecast distribution relative to the observed streamflow value, for every site, season, and streamflow quantile.

3. Results

To what extent can forecasts of climate and anthropogenic influences be used to enhance the seasonal predictability of streamflow using a statistical-dynamical forecasting approach? A systematic comparison of model performance across all seasons and initialization times provides insight into the predictability arising from precipitation and temperature (section 2.1), antecedent climate (section 2.2), and anthropogenic influences (section 2.3).

3.1. Basic Predictability Arising From NMME Precipitation and Temperature Forecasts

We begin by considering how the NMME climate forecast skill translates into the statistical streamflow forecasts across 290 catchments with widely varying physical and climate characteristics (section 2.1). On average, at the shortest initialization time and when all forecast/observation pairs are pooled across all sites, the NMME precipitation forecasts are better in winter/fall (precipitation R = 0.79/0.58, MSESS = 0.21/0.11, respectively) than in spring/summer (R = 0.57/0.27, MSESS = −0.13/−0.54). The NMME temperature forecasts tend to be better in the fall/spring (R = 0.87/0.84; MSESS = 0.67/0.66, respectively) than in the summer/winter (R = 0.77/0.81; MSESS = −0.90/0.61, respectively). One possible explanation for the poor skill (and tendency to overpredict) of NMME summer forecasts (Figure S1) may be their inability to capture recent decreases in summer surface air temperature in the central United States due to agricultural intensification (alongside increased surface humidity and rainfall; Alter et al., 2018). Spatially, however, the skill of NMME precipitation forecasts is extremely variable when considering individual catchments: the regions with high precipitation forecast skill tend to match those where a statistically significant relationship exists between heavy precipitation and certain climate modes of variability (Figure S2), such as the Pacific Decadal Oscillation in eastern Midwest in the winter or the North Atlantic Oscillation in the northwestern Midwest in the summer (Mallakpour & Villarini, 2016). Spatially, the temperature forecasts are considerably more skillful than precipitation across most seasons ($R > 0.5$) except in the southern part of the domain in winter/summer (Figure S3).

When only NMME precipitation is used as a predictor (model P), the skill of the streamflow forecasts varies considerably across sites, for low to high flows (Figures 2, 3, and S4). When all sites are pooled together, the predictability of $Q_{0.5}$ in terms of R and PS is consistently the highest in the spring/winter, and spatially, in the central Midwest around Iowa. Streamflow forecast skill does not decrease monotonically with longer initialization times, because the skill of NMME precipitation and temperature predictions is inconsistent for different lead times (see also Slater, Villarini, Bradley, Vecchi, 2018). When sites are pooled together we find that conditional biases (SREL) are the dominant source of bias (Figure 2), reflecting the inaccuracy of forecast variability (standard deviation). When considering sites separately, the unconditional biases (SME) tend to be the dominant source of bias (not shown here), likely due to overprediction/underprediction in the NMME precipitation forecasts (see Slater, Villarini, & Bradley, 2018).

The inclusion of the NMME temperature forecast as a predictor (model $P + T$) increases predictability mainly in spring and fall for most initialization times, and to some extent in the winter (Figure 2). In spring, this suggests that the temperature is an important predictor during the warming season in snowmelt-dominated catchments (see also Lehner et al. 2017). In the fall, the enhanced skill occurs possibly because temperature is a good proxy for evapotranspiration. In summer, however, temperature does not improve model skill, which may reflect inaccuracies in summer temperature forecasts as discussed above.

3.2. Enhancing Predictability by Including Antecedent Climate

The inclusion of antecedent precipitation as a predictor (model $P + AntP$) increases forecast skill considerably for the shortest initialization times (i.e., 0.5 to 2.5 months ahead of every season) across most seasons and streamflow quantiles (Figures 2, 3, and S4–S9). This enhancement is to be expected because x_{ap} includes observed precipitation and therefore provides a proxy for soil moisture conditions (section 2.2). When

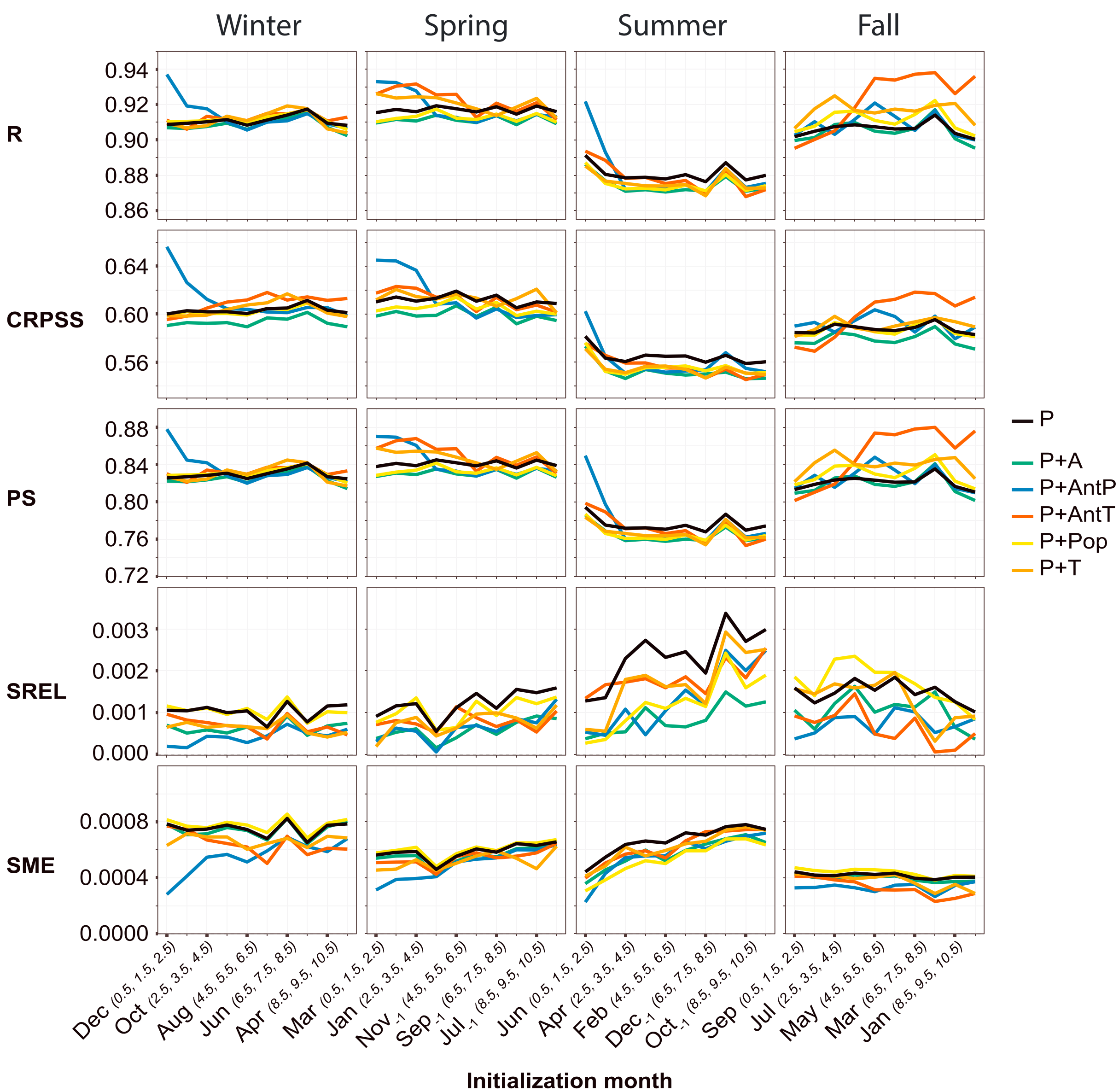

Figure 2. Verification for $Q_{0.5}$ leave-one-out cross-validated forecasts, computed by pooling forecasts and observation pairs for all 290 sites over the period 1983–2015. The verification is shown for all six models (colored lines; Table 1), for every season (columns), initialization month (x axis), and five evaluation metrics (rows; described in section 2.4). On the x axis, forecasts initialized in the previous year are indicated by the subscript -1 and lead times used to compute the precipitation/temperature forecasts are indicated in parentheses. Computing the verification metrics this way augments model skill but helps discriminate between different model formulations at the regional level when forecasts are computed over such a short period. When verification is completed on a site-by-site basis, we find a range of high to low skill (see Figure 3 and the supporting information for all sites, seasons, and lead times). CRPSS = continuous ranked probability skill score; PS = potential skill; R = Pearson's correlation coefficient; SREL = slope reliability; SME = standardized mean error.

pooling across 290 catchments, the increased skill arising from the inclusion of antecedent precipitation is negligible in the fall but ranges from 0.02 to 0.03 for the other seasons (with much higher enhancements for specific sites). However, at initialization times exceeding 2.5 months ahead of any given season, x_{ap} is strictly composed of *forecast* x_{ap} values, and so including antecedent precipitation only enhances the forecasts in the fall, between 3.5 and 5.5 months ahead; elsewhere, the forecast quality is deteriorated (although the biases are reduced compared to model *P*).

At individual sites, we find considerable improvements in streamflow forecast skill using model *P + AntP*. The correlation coefficient increases by >0.5 at 34% of sites at the shortest initialization time and at 23% of sites 3 months ahead (Figure S8). At the shortest initialization time, the greatest increase in model skill occurs in

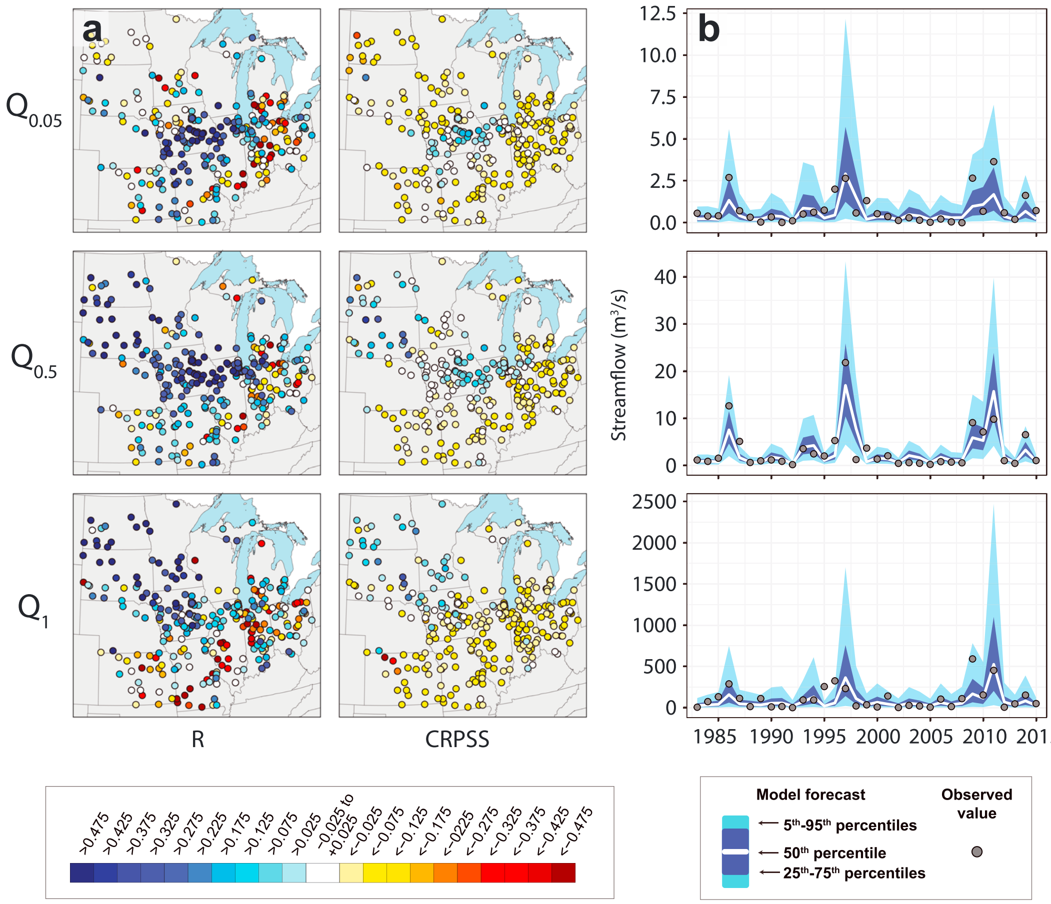

Figure 3. Spring streamflow forecast skill for model *P* + *AntP* at individual stream gauges, initialized at the start of the spring, with rows indicating streamflow quantiles ($Q_{0.05}$, $Q_{0.5}$, Q_1). (a) Maps indicate forecast skill at all 290 stream gauges; columns indicate correlation and CRPSS, with the same color scheme, ranging from blue (positive skill) to white (no skill) to red (negative skill). For other seasons and flow quantiles, see Figures S4–S7. (b) The time series illustrates the probabilistic model forecast at site 06359500 (Moreau River near Faith, SD). Time series for all other sites/seasons are shown in Figure S10. CRPSS = continuous ranked probability skill score.

winter (0.67) and the smallest in summer (0.12). In fall, the skill of model *P* + *AntP* remains largely negative except in the central Midwest around the state of Iowa. In the spring/summer, the median flows tend to be more predictable (R for $Q_{0.5}$ = 0.18/0.20, respectively) than the extremes (R for $Q_{0.05}$ = 0.06/0.18; Q_1 = 0.04/0.09). However, in fall/winter, the low flows tend to be more predictable (R = 0.25/0.35; see Figures 3 and S5). Note also that the correlation values provided here are computed using the Pearson method, and therefore, poor skill may reflect the influence of outliers.

The time series of fits and forecasts for model *P* + *AntP* are also verified visually for every site, season, and flow quantile (Figure S10). For most sites, the probabilistic forecasts reproduce the variability in the period 1983–2015 relatively well. There may nonetheless be some biases arising from real-world processes, as the USGS water year reports for some catchments indicate flow abstraction for irrigation or streamflow augmentation by effluent or return water from sewage treatment plants. Thus, climate and anthropogenic influences may also decrease model skill in the absence of site-specific calibration.

Antecedent temperature enhances the seasonal flow forecast skill (model *P* + *AntT*) at longer initialization times compared to other models (Figure 2). In terms of *R* and PS, model *P* + *AntT* outperforms model *P* for most lead times in the spring. In terms of CRPSS, model *P* + *AntT* shows a better performance in winter and fall for long initialization times. The improved skill at longer lead times is especially true when considering individual sites (not shown). Biases are also decreased across all four seasons by the inclusion of x_{at}.

3.3. Secondary Predictability Arising From Anthropogenic Influences

Alongside climate predictors, we also consider ancillary influences that could broadly be described as *anthropogenic*, that is, agriculture and population density. Previous work has shown that the use of corn and soybean catchment extent as a predictor improves seasonal streamflow model fits in the summer and fall, alongside antecedent precipitation (Slater & Villarini, 2017). Here we find that x_{ag} enhances the skill of streamflow forecasts considerably at certain sites (e.g., in fall and winter in Iowa; not shown). Yet when pooling the data across sites, we only find negligible improvements in R/PS for some lead times (Figure 2). Model biases (SME and SREL) are reduced by the inclusion of x_{ag} mainly in the summer, with some improvement in conditional biases (SREL) across all seasons.

Last, population density also increases streamflow forecast skill notably at certain sites, but not everywhere. When observation/forecast pairs are pooled from all sites, the improvement is most visible in the fall. The CRPSS does not increase notably, as the model biases are also enhanced (Figure 2). In spring and summer, the conditional biases (SREL) are reduced, and the unconditional biases (SME) are reduced only in summer. In earlier work, we found that population density enhanced seasonal flow model fits in urban catchments specifically (Slater & Villarini, 2017). Here the average increase in correlation with predictor x_{pop} in the 20 catchments with more than 500 persons per square kilometer is +0.43 in the fall (versus +0.13 in catchments with $<$500 pers/km^2). Therefore, the inclusion of population density as a predictor does seem to enhance forecast skill, perhaps by better reflecting water consumption patterns or, over longer time scales, the changes in runoff properties due to the increasing extent of impervious areas.

4. Conclusions

This work reveals that a simple statistical streamflow forecasting approach, using only the dynamical precipitation forecast as input, can be enhanced with a range of climatic and land cover predictors in different seasons and initialization months. Results vary notably depending on the chosen forecasting and validation approaches, but they suggest that considerable improvements in forecast skill may be obtained by choosing the predictors that are most appropriate for specific catchments, streamflow quantiles, seasons, or initialization months (e.g., using both antecedent precipitation and antecedent temperature in winter).

In future work, statistical-dynamical seasonal flow forecasting systems can be further enhanced by improving the quality of model inputs, outputs, and the model formulation. Model inputs can be improved by using the entire GCM forecast ensemble instead of simply the mean value and applying forecast ensembling approaches such as Bayesian updating (Bradley et al., 2015), optimal weights (Wanders & Wood, 2016), Bayesian joint probability (Schepen et al., 2018), or forecast monetary value (Cloke et al., 2017). The types of model inputs can also be expanded to include information about the initial land surface conditions or large-scale climate precursors such as the North Atlantic Oscillation or El Niño–Southern Oscillation (Emerton et al., 2017; Yuan, Wood, & Ma, 2015). Model formulation may be enhanced by using objective selection of model predictors (e.g., Robertson & Wang, 2012), interaction terms (e.g., between precipitation and antecedent climatic conditions), or conditioning approaches that subsample/weight the historical streamflow time series (Crochemore et al., 2017). Last, adjustments can be made to account for human interferences in the water cycle (e.g., reservoirs and flow abstraction; see Pagano et al., 2014) or catchment differences, via model calibration or statistical postprocessing methods such as Bayesian model averaging (Zsótér et al., 2016) to reduce forecast biases and spread.

This work represents a first step in the systematic testing of NMME-based statistical-dynamical streamflow forecasts at seasonal time scales. Further research will seek to evaluate a broad range of statistical-dynamical approaches at global scales, with predictors representing both climatic and anthropogenic influences, and with the aim of making these forecasts available to users and decision makers.

Acknowledgments
We thank two anonymous reviewers for helpful feedback, Christel Prudhomme for constructive discussion, as well as the different modeling centers, and the USGS for making their data publicly available. The NMME seasonal climate forecasts can be downloaded at http://www.cpc.ncep.noaa.gov/products/NMME/, the streamflow data at https://waterdata.usgs.gov/nwis/, and agricultural statistics at https://quickstats.nass.usda.gov/. This study was supported in part by the Broad Agency Announcement (BAA) Program and the Engineer Research and Development Center (ERDC)-Cold Regions Research and Engineering Laboratory (CRREL) under contract W913E5-16-C-0002, and by the National Science Foundation under CAREER grant AGS-1349827. The authors declare that there are no conflicts of interest.

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Alter, R. E., Douglas, H. C., Winter, J. M., & Eltahir, E. A. B. (2018). Twentieth century regional climate change during the summer in the Central United States attributed to agricultural intensification. *Geophysical Research Letters*, *45*, 1586–1594. https://doi.org/10.1002/2017GL075604

Andresen, J., Hilberg, S., & Kunkel, K. (2012). Historical climate and climate trends in the Midwestern USA. *U.S. National Climate Assessment Midwest Technical Input Report*, 1–18. Retrieved from http://glisa.Msu.Edu/docs/NCA/MTIT_Historical.Pdf

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., et al. (2017). Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences Discussions*, (October), 1–27. https://doi.org/10.5194/hess-2017-610

Becker, E., van den Dool, H., & Zhang, Q. (2014). Predictability and forecast skill in NMME. *Journal of Climate*, *27*(15), 5891–5906. https://doi.org/10.1175/JCLI-D-13-00597.1

Bradley, A. A., Habib, M., & Schwartz, S. S. (2015). Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resources Research*, *51*, 7382–7400. https://doi.org/10.1002/2014WR016811

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M., & Peterson, D. H. (2001). Changes in the onset of spring in the Western United States. *Bulletin of the American Meteorological Society*, *82*, 399–415. https://doi.org/10.1175/1520-0477(2001)082<2265:CAACOC>2.3.CO;2

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3–4), 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Cloke, H. L., Pappenberger, F., Smith, P. J., & Wetterhall, F. (2017). How do I know if I've improved my continental scale flood early warning system? *Environmental Research Letters*, *12*(4), 1–10. https://doi.org/10.1088/1748-9326/aa625a

Crochemore, L., Ramos, M.-H., Pappenberger, F., & Perrin, C. (2017). Seasonal streamflow forecasting by conditioning climatology with precipitation indices. *Hydrology and Earth System Sciences*, *21*(3), 1573–1591. https://doi.org/10.5194/hess-21-1573-2017

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate Research*, *22*, 99–113. https://doi.org/10.3354/cr022099

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, *111*(2), 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., et al. (2018). Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS v2.2 seasonal v1.0. *Geoscientific Model Development Discussion*. https://doi.org/10.5194/gmd-2018-118

Emerton, R. E., Cloke, H. L., Stephens, E. M., Zsoter, E., Woolnough, S. J., & Pappenberger, F. (2017). Complex picture for likelihood of ENSO-driven flood hazard. *Nature Communications*, *5*, 71–99. https://doi.org/10.1038/NCOMMS14796

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., et al. (2016). Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water*, *3*(3), 391–418. https://doi.org/10.1002/wat2.1137

Grantz, K., Rajagopalan, B., Clark, M., & Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, *41*, W10410. https://doi.org/10.1029/2004WR003467

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., & Tanguy, M. (2018). Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, *22*, 2023–2039. https://doi.org/10.5194/hess-22-2023-2018

Ionita, M., Lohmann, G., & Rimbu, N. (2008). Prediction of spring Elbe discharge based on stable teleconnections with winter global temperature and precipitation. *Journal of Climate*, *21*(23), 6215–6226, https://doi.org/10.1175/2008JCLI2248.1.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The north American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(April), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Krakauer, N. Y. (2017). Temperature trends and prediction skill in NMME seasonal forecasts. *Climate Dynamics*, *0*(0), 1–13. https://doi.org/10.1007/s00382-017-3657-2

Lehner, F., Wood, A. W., Llewellyn, D., Blatchford, D. B., Goodbody, A. G., & Pappenberger, F. (2017). Mitigating the impacts of climate non-stationarity on seasonal streamflow predictability in the U.S. southwest. *Geophysical Research Letters*, *44*, 12,208–12,217. https://doi.org/10.1002/2017GL076043

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250–254. https://doi.org/10.1038/nclimate2516

Mallakpour, I., & Villarini, G. (2016). Investigating the relationship between the frequency of flooding over the central United States and large-scale climate. *Advances in Water Resources*, *92*, 159–171. https://doi.org/10.1016/j.advwatres.2016.04.008

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., et al. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *21*(7), 3915–3935. https://doi.org/10.5194/hess-21-3915-2017

Mo, K. C., & Lettenmaier, D. P. (2014). Hydrologic prediction over the conterminous United States using the national multi-model ensemble. *Journal of Hydrometeorology*, *15*(4), 1457–1472. https://doi.org/10.1175/JHM-D-13-0197.1

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society*, *86*(1), 39–50. https://doi.org/10.1175/BAMS-86-1-39

Murphy, A. H., & Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, *7*, 435–455. https://doi.org/10.1016/0169-2070(92)90028-8

Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of official western US seasonal water supply outlooks, 1922–2002. *Journal of Hydrometeorology*, *5*(5), 896–909. https://doi.org/10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2

Pagano, T. C., Wood, A. W., Ramos, M.-H., Cloke, H. L., Pappenberger, F., Clark, M. P., et al. (2014). Challenges of Operational River forecasting. *Journal of Hydrometeorology*, *15*(4), 1692–1707. https://doi.org/10.1175/JHM-D-13-0188.1

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Environmental Science & Policy*, *51*, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., et al. (2015). How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology*, *522*, 697–713. https://doi.org/10.1016/j.jhydrol.2015.01.024

Ramos, M. H., van Andel, S. J., & Pappenberger, F. (2013). Do probabilistic forecasts lead to better decisions? *Hydrology and Earth System Sciences*, *17*(6), 2219–2232. https://doi.org/10.5194/hess-17-2219-2013

Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E. (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River basin. *Water Resources Research*, *42*, W09404. https://doi.org/10.1029/2005WR004653

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *54*(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Robertson, D. E., Pokhrel, P., & Wang, Q. J. (2013). Improving statistical forecasts of seasonal streamflows using hydrological model output. *Hydrology and Earth System Sciences*, *17*(2), 579–593. https://doi.org/10.5194/hess-17-579-2013

Robertson, D. E., & Wang, Q. J. (2012). A Bayesian approach to predictor selection for seasonal streamflow forecasting. *Journal of Hydrometeorology*, *13*(1), 155–171. https://doi.org/10.1175/JHM-D-10-05009.1

Schepen, A., Zhao, T., Wang, Q. J., & Robertson, D. E. (2018). A Bayesian modelling method for post-processing daily sub-seasonal to seasonal rainfall forecasts from global climate models and evaluation for 12 Australian catchments. *Hydrology and Earth System Sciences*, *22*, 1615–1628. https://doi.org/10.5194/hess-22-1615-2018

Slater, L. J., & Villarini, G. (2017). Evaluating the drivers of seasonal streamflow in the U.S. Midwest. *Water*, *9*(9), 1–22. https://doi.org/10.3390/w9090695

Slater, L. J., Villarini, G., & Bradley, A. A. (2018). 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. *Climate Dynamics*. https://doi.org/10.1007/s00382-016-3286-1

Slater, L. J., Villarini, G., Bradley, A. A., & Vecchi, G. A. (2018). A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed. *Climate Dynamics*. https://doi.org/10.1007/s00382-017-3794-7

Souza Filho, F. A., & Lall, U. (2003). Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resources Research*, *39*(11), 1307. https://doi.org/10.1029/2002WR001373

Thielen, J., Bartholmes, J., Ramos, M.-H., & de Roo, A. (2009). The European flood alert system—Part 1: Concept and development. *Hydrology and Earth System Sciences*, *13*(2), 125–140. https://doi.org/10.5194/hess-13-125-2009

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). *Journal of Hydrometeorology*, *16*(6), 2329–2344, https://doi.org/10.1175/JHM-D-15-0053.1.

Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., & Beck, H. E. (2013). Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resources Research*, *49*, 2729–2746. https://doi.org/10.1002/wrcr.20251

Villarini, G., & Strong, A. (2014). Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. *Agriculture, Ecosystems & Environment*, *188*, 204–211. https://doi.org/10.1016/j.agee.2014.02.036

Wanders, N., & Wood, E. F. (2016). Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, *11*(9), 94,007. https://doi.org/10.1088/1748-9326/11/9/094007

Weigel, A. P., & Mason, S. J. (2011). The generalized discrimination score for ensemble forecasts. *Monthly Weather Review*, *139*(9), 3069–3074. https://doi.org/10.1175/MWR-D-10-05069.1

Werner, K., Brandon, D., Clark, M., & Gangopadhyay, S. (2004). Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. *Journal of Hydrometeorology*, *5*(6), 1076–1090. https://doi.org/10.1175/JHM-381.1

Wood, A. W., & Schaake, J. C. (2008). Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, *9*(1), 132–148. https://doi.org/10.1175/2007JHM862.1

Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye, A., & Peng, S. (2016). An experimental seasonal hydrological forecasting system over the Yellow River basin—Part I: Understanding the role of initial hydrological conditions. *Hydrology and Earth System Sciences Discussions*, 1–26. https://doi.org/10.5194/hess-2016-101

Yuan, X., Roundy, J. K., Wood, E. F., & Sheffield, J. (2015). Seasonal forecasting of global hydrologic extremes: System development and evaluation over GEWEX basins. *Bulletin of the American Meteorological Society*, *96*(November), 1895. https://doi.org/10.1175/BAMS-D-14-00003.1–1912.

Yuan, X., Wood, E. F., & Ma, Z. (2015). A review on climate-model-based seasonal hydrologic forecasting: Physical understanding and system development. *Wiley Interdisciplinary Reviews: Water*, *2*(5), 523–536. https://doi.org/10.1002/wat2.1088

Zsótér, E., Pappenberger, F., Smith, P., Emerton, R. E., Dutra, E., Wetterhall, F., et al. (2016). Building a multimodel flood prediction system with the TIGGE archive. *Journal of Hydrometeorology*, *17*(11), 2923–2940. https://doi.org/10.1175/JHM-D-15-0130.1