

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR026987

Key Points:

- Forecast quality of seasonal streamflow volume varies geographically and seasonally, while streamflow predictability can be regionalized
- Streamflow predictability is strongly dependent on the basin's hydrological regime, climatology, and precipitation forecast biases
- Predictability is higher in river systems of long streamflow memory than in systems immediately responding to the precipitation signal

Supporting Information:

- Supporting Information S1

Correspondence to:

I. G. Pechlivanidis,
ilias.pechlivanidis@smhi.se

Citation:

Pechlivanidis, I. G., Crochemore, L., Rosberg, J., & Bosshard, T. (2020). What are the key drivers controlling the quality of seasonal streamflow forecasts? *Water Resources Research*, 56, e2019WR026987. <https://doi.org/10.1029/2019WR026987>

Received 21 DEC 2019

Accepted 29 APR 2020

Accepted article online 1 MAY 2020

What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts?

I. G. Pechlivanidis¹ , L. Crochemore¹ , J. Rosberg¹, and T. Bosshard¹

¹Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

Abstract Recent technological advances in representation of processes in numerical climate models have led to skillful predictions, which can consequently increase the confidence of hydrological predictions and usability of hydroclimatic services. Given that many water-related stakeholders are affected by seasonal hydrological variations, there is a need to manage such variations to their advantage through better understanding of the drivers that influence hydrological predictability. Here we analyze the seasonal forecasts of streamflow volumes across about 35,400 basins in Europe, which lie along a strong gradient in terms of climatology, scale, and hydrological regime. We then link the seasonal volumetric errors to various physiographic-hydroclimatic descriptors and meteorological biases in order to identify the key drivers controlling predictability. Streamflow volumes over Europe are well predicted, yet with some geographic and seasonal variability; however, the predictability deteriorates with increasing lead time particularly in the winter months. Nevertheless, we show that the forecast quality is well correlated to a set of descriptors, which vary depending on the initialization month. The forecast quality of seasonal streamflow volumes is strongly dependent on the basin's hydrological regime, with limited predictability in relatively flashy basins. On the contrary, snow and/or baseflow dominated regions with long recessions show high streamflow predictability. Finally, climatology and precipitation forecast biases are also related to streamflow predictability, highlighting the importance of developing robust bias adjustment methods. Overall, this investigation shows that the seasonal streamflow predictability can be clustered, and hence regionalized, based on a priori knowledge of local hydroclimatic conditions.

Plain Language Summary Hydrological information for the months ahead is of great value to existing decision-making practices, particularly to those affected by the vagaries of the climate and who would benefit from better understanding and managing climate-related risks. Currently, there is limited knowledge of the factors controlling the quality of the seasonal streamflow forecasts. We analyze such forecasts over Europe and link their predictability to basin descriptors and meteorological biases. This allows the identification of the key drivers along a strong hydroclimatic gradient. The seasonal streamflow predictability varies geographically and seasonally with acceptable values for the first lead months. Predictability deteriorates with increasing lead time particularly in the winter months. The hydrological regime is strongly linked to the forecast quality, with quickly reacting basins showing low values. Basin climatology and precipitation forecast biases are also related to the predictability of streamflow.

1. Introduction

Seasonal forecasts hold the potential for being of great value for a wide range of stakeholders who are affected by the vagaries of the climate and who would benefit from understanding and better managing climate-related risks (Bruno Soares et al., 2017; Contreras et al., 2020; Doblas-Reyes et al., 2013). In Europe, there has been relatively little uptake and use of seasonal forecasts by users for decision-making, compared to other parts of the world, that is, Africa, the United States, and Australia (Bennett et al., 2017; Hansen et al., 2011; Mendoza et al., 2017), probably due to the relatively limited skill of seasonal meteorological forecasts in Europe (Arnal et al., 2018; Giuliani et al., 2020; Greuell et al., 2018; Harrigan et al., 2018; Wanders et al., 2019). However, recent advances in our understanding and forecasting of climate have resulted in skillful and useful meteorological predictions, which can consequently increase the confidence of hydrological predictions and improve awareness and preparedness from a user perspective (Bruno Soares & Dessai, 2016; Buontempo et al., 2018; Hewitt et al., 2017).

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The accuracy of seasonal hydrological forecasts is subject to multiple sources of error/uncertainty, which are present in the various components of the production chain, that is, meteorological forecasts, bias adjustment, hydrological model(s) and setup, and model initialization (Crochemore et al., 2016; Demirel et al., 2013; Thiboult et al., 2016). Consequently, to improve hydrological forecasts, each component has to be evaluated to assess its relative contribution to the overall forecasting accuracy (Arnal et al., 2017; Wood & Lettenmaier, 2008; Yossef et al., 2017). Even though the large heterogeneity in the hydroclimatic patterns and physiographic descriptors leads to a strong spatiotemporal variability of the hydrological predictability, the understanding of the key drivers (i.e., climatological conditions, human impacts, hydrological regimes, topography, etc.) influencing hydrological predictability is still limited.

The majority of seasonal hydrological impact modeling efforts has commonly been conducted in only one or a limited number of basins limiting the need for an increased understanding of large systems, which are, for example, heavily influenced by human activities (Apel et al., 2018; Foster et al., 2018; Meißner et al., 2017; Yuan et al., 2016). Large-scale (i.e., continental) multibasin modeling can complement the “deep” knowledge from basin-based modeling, enhance process understanding, increase robustness of generalizations, and facilitate classification of basin behavior and prediction (Gudmundsson et al., 2012; Kumar et al., 2013; Pechlivanidis & Arheimer, 2015). Specifically, for seasonal hydrological forecasting, multibasin modeling can support better understanding of prediction uncertainty and go beyond sensitivities related to initial hydrological conditions and meteorological forecasts that regional investigations can only target (Lavers et al., 2020; Wood & Lettenmaier, 2008). This type of modeling has the potential to cross regional and international boundaries, while analysis over a number of basins allows the consideration of different geophysical and climatic zones and hydrological regimes (Gupta et al., 2014; Krysanova et al., 2017); hence, it can provide a deeper understanding of the underlying sensitivities in forecast quality. Such modeling can also advance hydrological science, since it founds a numerical background for comparative hydrology. The use of a large sample of stations, particularly when analyses are conducted at the continental scale, can also allow for exploration of emerging patterns and facilitate the testing of sensitivities for basins with a wide range of environmental conditions (Blöschl et al., 2013; Pechlivanidis et al., 2017, 2018; Rakovec et al., 2016; Samaniego et al., 2017).

In natural river systems, streamflow fluctuations are driven both by discharges from the basin's water storages (i.e., groundwater, snowpack, soil moisture, and channel network) and by meteorological forcings. Efforts have consequently been made to apportion the role of initial hydrological conditions and meteorological forecasts in seasonal streamflow prediction, resulting in a number of uncertainty attribution frameworks. Among others, the Ensemble Streamflow Prediction and reverse Ensemble Streamflow Prediction framework, which was proposed by Wood and Lettenmaier (2008), has received high attention. This framework was later extended to allow blending of the two sources of seasonal streamflow forecast skill and assess skill elasticity (Arnal et al., 2017; Wood et al., 2016). In various investigations, these frameworks could identify the primary contributors to seasonal hydrological skill and uncertainties (e.g., Shukla & Lettenmaier, 2011; Staudinger & Seibert, 2014; Yossef et al., 2013; Yuan et al., 2016); however, there has been limited provision of links of streamflow predictability to the physical drivers hidden behind initial conditions and meteorological forcings.

Understanding processes in large river systems is challenging, given that physical properties (e.g., vegetation and soil type) generally exhibit high spatial variability, which consequently result in significant differences in system behavior and predictability (Kuentz et al., 2017). As expected, this spatial heterogeneity introduces further uncertainty on the categorization of important drivers that influence the predictive hydrological quality. In addition, large river basins are often strongly influenced by human activities (e.g., irrigation, hydropower production, and groundwater use) for which information can be difficult to attain and therefore rarely described in hydrological model processes; hence, introducing additional uncertainty regarding process understanding and description (Andersson et al., 2015; Nazemi & Wheeler, 2015).

Here, we make a step forward by gaining insights in spatial patterns of hydrological predictability at the large scale and link this to the descriptors of the basin systems. We pose the following scientific questions: (1) *What are the limits of predictability for hydrological forecasting systems?* And (2) *what are the drivers affecting the accuracy of the seasonal hydrological forecasts?* To address these questions, we (a) assess the hydrological forecasting performance across Europe's hydroclimatic gradient for all initialization months and different

lead times (up to 7 months), (b) detect relationships between hydrological forecasting performance and physiographic-hydrological-climatic descriptors to understand the key controls of poor/good accuracy along the gradient, and (c) rank these drivers based on their potential to categorize/describe the forecasting accuracy for different initialization months. The paper is structured as follows. Section 2 presents the hydrological model setup, data used, and methodology. Section 3 presents the results, followed by a discussion in section 4. Finally, section 5 states the conclusions.

2. Model, Data, and Methods

2.1. Hydrological Model Description

HYPE (HYdrological Predictions for the Environment) is a continuous semidistributed process-based model, which simulates components of the water cycle (i.e., snow accumulation and melting, evapotranspiration, soil moisture, streamflow generation, groundwater recharge, and routing through rivers and lakes) at a daily time step (Lindström et al., 2010). HYPE simulates the water flow paths in soil, which is divided into three layers with a fluctuating groundwater table. Parameters are linked to physiographical (soil type and depths and vegetation) characteristics in the landscape. Lakes receive the local runoff and the streamflow from upstream subbasins, while the outflow from a lake is determined by a rating curve. For reservoirs, a simple regulation scheme can be used, in which the outflow is constant or follows a seasonal function for water levels above a threshold. A rating curve for the spillways can be used when the reservoir is full. Irrigation is simulated based on crop water demands calculated either with the FAO-56 crop coefficient method or relative to a reference flooding level for submerged crops (e.g., rice). Here, the crop type is assumed to be constant in time, and, therefore, cases where farmers may change crops between the seasons or years are not accounted for. The demands are withdrawn from rivers, lakes, reservoirs, and/or groundwater within and/or external to the subbasin where the demands originated. The demands are constrained by the water availability in these sources. After subtraction of conveyance losses, the withdrawn water is applied as additional infiltration to the irrigated soils from which the demands originated.

The HYPE model setup for the pan-European region (8.8 million km²), referred to as E-HYPE v.3.0 (Hundecha et al., 2016), was employed allowing analysis of model outputs in 35,408 subbasins (215 km² in average spatial resolution). Open global data sources were used to extract information on terrain, soil, land use, lakes, reservoirs, and irrigation (Donnelly et al., 2016). Meteorological variables of mean daily precipitation and temperature are derived from the HydroGFD product v2.0 (Hydrological Global Forcing Data version 2.0), which is an observation-corrected reanalysis data set provided daily at a 0.5° gridded resolution (Berg et al., 2018). HydroGFD is the reference data set and was used to drive the hydrological model for the period 1991–2015 (hereafter referred to as “reference simulation”; the first 2 years were used to spin up the model states). E-HYPE was calibrated and evaluated against multiple variables (i.e., streamflow, evapotranspiration, snow, water quality) extracted by in situ observations and earth observations. For instance, the performance of E-HYPE in validation in terms of streamflow reaches a median Nash-Sutcliffe Efficiency of 0.53 over Europe. Details about the model performance and its relation to physiographic-climatic characteristics can be found in Hundecha et al. (2016). The modeled monthly means (1991–2015) for precipitation, temperature, and streamflow can be found in Figures S1–S3 in the supporting information respectively.

2.2. Seasonal Meteorological Forecasts and Bias Adjustment

Seasonal predictions of daily mean precipitation and temperature were taken from the fifth generation seasonal forecasting system of the European Centre for Medium-Range Weather Forecasts, named SEAS5 (Johnson et al., 2019). The SEAS5 reforecasts (also known as hindcasts) used here consist of 25 ensemble members available at a grid spacing of approximately 36 km. SEAS5 reforecasts are available for the period 1993–2015 and consist of 7-month forecasts initialized at the beginning of each month.

The reforecast data were bias adjusted using a modified version of the Distribution Based Scaling (DBS) method (Yang et al., 2010) to account for drifting. Using the data for the whole analysis period, the bias adjustment parameters are conditioned on the lead month and the forecast issue date. DBS has originally been developed to adjust biases in climate projections and is a quantile-mapping method adapted here for seasonal forecasting. Bias adjustment was conducted on all monthly initialized forecasts using the

HydroGFD data set as reference. After bias adjustment, the cumulative distribution of daily precipitation and temperature forecasts follows closely the one of the HydroGFD data.

2.3. Seasonal Hydrological Forecasts—Evaluation

E-HYPE runs with the bias-adjusted SEAS5 reforecasts as forcing input, taking initial hydrological model states (snow, water levels in reservoirs/lakes/wetlands, soil moisture, and streamflow) from the “reference simulation.” In this paper, we extract E-HYPE streamflow (m^3/s) in 35,408 subbasins to assess the model’s predictive accuracy on seasonal timescales. Seasonal reforecasts are evaluated with respect to their performance against the “reference simulation” (also known as “pseudo-observations” or perfect forecast) per initialization month and lead month. It is important to note that the comparison of modeled reforecasts to modeled pseudo-observations eventually leads to a theoretical, and not to the actual, accuracy. We do not base the evaluation on real observations of streamflow, as their availability does not cover the entire hydroclimatic gradient of the European river systems; streamflow time series are only available and quality assured at 1,366 stations (see Kuentz et al., 2017). In addition, an evaluation in the modeled world is independent from the hydrological model’s imperfection. The evaluation of the reforecasts is performed on monthly mean streamflow (i.e., daily streamflow averaged over a month) for the 1993–2015 period. Here, the first month of the forecast (model just initialized) is referred to as Lead Month 0 (e.g., January 1993 streamflow for forecasts issued on 1 January 1993). The second month of the forecast is referred to as Lead Month 1 (e.g., February 1993 streamflow for forecasts issued on 1 January 1993), etc.

In this paper, we assess the forecasts in terms of their accuracy (—), that is, the performance of the monthly streamflow forecasts in comparison to modeled time series (named “pseudo-observations” hereafter), which are both expressed in m^3/s . For this purpose, the Continuous Rank Probability Score (*CRPS*; Hersbach, 2000) was calculated for each subbasin, target month and lead time. *CRPS* is defined as the integral of the squared distance between the cumulative distribution of the forecast members and a step function for the “pseudo-observations.” The score is the average of this integral computed at each time step of the evaluation period. We next standardized the *CRPS* score (*CRPS'*) by dividing it by the monthly mean of the “pseudo-observations” (*MQ*):

$$CRPS' = 1 - CRPS / MQ.$$

While *CRPS* values (in m^3/s) range between 0 and ∞ with 0 indicating a perfect forecast, *CRPS'* values (—) range between $-\infty$ and 1 with 1 being the optimum. *CRPS'* allows a clear categorization of its values into (very) good/poor performances, and it is used to present results hereafter.

2.4. Seasonal Hydrological Forecasts—Identifying Key Controlling Drivers

To better understand the controls of the forecasting accuracy, we explore the spatial streamflow patterns across the entire continent by analyzing the *CRPS'* score in all 35,408 basins modeled by the E-HYPE model on every initialization month and lead month. We apply the classification and regression trees (CART) method (Breiman et al., 1984) to identify regions of similar forecasting accuracy and their key controlling drivers using the Statistics and Machine Learning Toolbox™ for MATLAB® (The MathWorks, 2019). CART is a recursive-partitioning algorithm that classifies the space defined by the input descriptors (i.e., physiographic, hydrological, and climatic) based on the output variable (e.g., *CRPS'* score for Lead Month 2 and target month March). The “tree” consists of a series of nodes, where each node is a logical expression based on a similarity metric in the input space (here the physiographic-hydroclimatic descriptors). The method also provides information on the probabilities of different output groups at each “leaf” node (see a CART example in supporting information Figure S4a). We divided *CRPS'* into five groups—bad ($CRPS' \leq 0.2$), poor ($0.2 < CRPS' \leq 0.4$), medium ($0.4 < CRPS' \leq 0.6$), good ($0.6 < CRPS' \leq 0.8$), and very good ($CRPS' > 0.8$). A terminal “leaf” exists at the end of each branch of the “tree,” where the probability of belonging to any of the five output groups can be inspected. Here, we summarize the basin descriptors into climatic, topographic, human impacts, biases in forcing input, and hydrological similarity (Table 1). The degree of regulation at each dam (DoR) was calculated by dividing the dam capacity with the mean annual inflows to the dam. The dam capacity is provided by the Global Reservoir and Dam (Grand) database (Lehner et al., 2011). DoR indicates the dam’s capacity to store the runoff generated over a year, that is, if DoR = 1, the dam can hold all runoff generated within 1 year, and if DoR = 0.5, the dam can hold half of

Table 1
Basin Descriptors Used in the CART Analysis

Type	Name [reference]	Abbreviation	Unit
Climatology/forcing biases (9 descriptors)	Precipitation	Prec	mm/month
	Temperature	Temp	°C
	<i>Snow depth</i>	<i>Snow</i>	<i>cm/month</i>
	<i>Actual evapotranspiration</i>	<i>AET</i>	<i>mm/month</i>
	<i>Potential evapotranspiration</i>	<i>PET</i>	<i>mm/month</i>
	<i>Dryness index</i>	<i>PET/Prec</i>	—
	Evaporative index	AET/Prec	—
	Bias in precipitation	BiasPrec	%
	Bias in temperature	BiasTemp	°C
Topography (4)	Area	Area	km ²
	Elevation	Elev	m
	Relief ratio	Relief	—
	<i>Slope</i>	<i>Slope</i>	%
Human impact (1)	Degree of regulation	DoR	%
Hydrological signatures (15)	Mean annual specific runoff (Viglione et al., 2013)	Qm	mm/year
	Normalized high streamflow (Viglione et al., 2013)	q05	—
	Normalized low streamflow (Viglione et al., 2013)	q95	—
	Normalized relatively low streamflow (Viglione et al., 2013)	q70	—
	Slope of streamflow duration curve (Viglione et al., 2013)	mFDC	%/%
	Range of Pardé coefficient (Viglione et al., 2013)	DPar	—
	Coefficient of variation (Donnelly et al., 2016)	CV	—
	Flashiness (Donnelly et al., 2016)	Flash	—
	Normalized peak distribution (Euser et al., 2013)	PD	—
	Rising limb density (Euser et al., 2013)	RLD	—
	Declining limb density (Euser et al., 2013)	DLD	—
	Baseflow index (Kuentz et al., 2017)	BFI	—
	Runoff coefficient (Kuentz et al., 2017)	RC	—
	Streamflow elasticity (Sawicz et al., 2011)	EQP	—
	High pulse count (Yadav et al., 2007)	HPC	—

Note. Descriptors in italic were omitted in the regression analysis due to strong correlation with other descriptors.

the runoff generated within 1 year. The descriptors were also analyzed for interdependence, and highly interdependent (correlation coefficient greater or smaller than +0.7 or −0.7 respectively; see Figure S5) descriptors were omitted to avoid overfitting in the CART analysis (see Table 1).

We next calculate the descriptors' importance by summing changes in the probability of splitting on every descriptor and dividing the sum by the number of branch nodes (Loh, 2011; The MathWorks, 2019). The descriptor's importance is further normalized to range from 0 to 1 since the number of generated branch nodes depends on the input space defined by the descriptors, and hence allowing an intercomparison between months and lead months. In order to avoid the high dimensionality in the CART analysis, hydrological signatures (i.e., a set of 15 statistics describing hydrological behavior and listed in Table 1) were firstly clustered in groups receiving an identification number (named FlowID), which is further introduced in the CART. The hydrological signatures were used to identify the dominant processes and to determine the temporal characteristics (extremes and information at daily, seasonal, and annual timescales) of the streamflow response (Sawicz et al., 2011; Westerberg & McMillan, 2015). A k-means clustering approach within the 15-dimension space (consisting of the 15 calculated hydrological signatures in Table 1) is applied to categorize the subbasins based on their combined similarity in streamflow signatures. Through the mapping of the spatial pattern one can gain insights into the similarities of basin functioning and further identify the dominant streamflow generating processes for specific regions.

3. Results

3.1. Bias Adjustment of SEAS5 Forecasts

SEAS5 precipitation and temperature forecasts were bias adjusted for all initialization months and lead months using HydroGFD as reference. These biases, that is, the average difference between the forecast

ensemble and the HydroGFD, are not similar in terms of magnitude and spatial variability, while as expected following bias adjustment, they are significantly reduced both for precipitation and temperature. Figure 1 shows large positive and negative biases in seasonal precipitation forecasts with sharp gradients between regions, which coincide with regions of complex topography and coastal areas (e.g., Spain, the United Kingdom, France, and southeastern Europe). In general, SEAS5 tends to overpredict precipitation in the winter (rain and snowfall season) in northern Europe. SEAS5 also tends to overpredict precipitation in northern Europe from the middle of spring till late summer. In most parts of Europe, temperature is underpredicted by on average 1 and 2°C in all months and lead months; only in February is temperature overpredicted by about 2°C in northeastern Europe (see Figure S6). Note also that Figures S1 and S2 present the spatial variability of monthly mean precipitation and temperature, respectively, allowing an estimation of the proportional biases before and after bias adjustment.

Despite the effectiveness of the DBS method, some biases unavoidably remain in the meteorological forecasts (particularly in precipitation). Such remaining biases are the result of an assumed theoretical distribution of daily data, which in general does not perfectly fit but rather tries to minimize the number of parameters in the correction model. Such remaining biases are further propagated in the forecasting production chain, and hence potentially affecting the hydrological forecast quality. For instance, most remaining precipitation (positive) biases are located in southern Europe and the Mediterranean, and generally in highly elevated regions. However, after the DBS bias adjustment only negligible bias remains for temperature and the spatial pattern becomes essentially identical to the HydroGFD. These results together with supporting information Figures S1–S3 provide supporting information of the monthly means (1991–2015) for precipitation, temperature, and streamflow, respectively.

3.2. Forecast Evaluation of Seasonal Streamflow Volumes

Results of forecast quality across the European domain show that forecasts can adequately predict the streamflow achieving high $CRPS'$ values, with the highest predictability roughly from April to August and the lowest one generally in autumn and winter (Figure 2). This could be related to the relatively small streamflow volumes occurring in the warm months (May–August) in comparison to autumn and winter. In addition, temperature is more influential in summer and is easier to predict than precipitation. Overall, in Lead Month 0 forecast quality can be described as very good ($CRPS' > 0.8$; however, with the exception of forecasts during autumn), while in the high lead months the forecast quality can only be described as good ($0.6 < CRPS' \leq 0.8$).

Figure 3 shows the spatial variability of streamflow volumetric errors (described by $CRPS'$) for the winter and summer seasons and for Lead Months 0, 2, and 4. Overall, hydrological forecast quality varies both geographically and seasonally with acceptable performance over the entire domain in Lead Month 0. High forecast quality is generally shown in central and northern Europe, particularly in winter in the short lead months, and in central and western Europe in summer. The hydrological response in the cold regions of northern Europe is controlled by snow accumulation/melting processes, and hence, temperature forecasts are expected to be an important driver controlling accuracy. The performance of hydrological forecasts is higher in the summer months (May–July), in comparison to winter, during which snow melting controls the basin response. Consequently, streamflow forecasts are influenced by accurately estimating the snowpack in the previous months. In addition, as expected, the forecasting performance deteriorates with increased lead month, particularly in southern and eastern Europe. The hydrological response in these regions is generally driven by rainfall since temperature is generally above freezing temperatures and snow only occurs in highly elevated basins.

3.3. Understanding Processes and Predictability Along a Hydroclimatic Gradient

3.3.1. Hydrological Similarity and Dominant Processes

To identify river systems of similar hydrological behavior, 15 streamflow signatures from the E-HYPE hydrological model setup were categorized, resulting into 11 different clusters of different size and varying distributions in the signatures (Figure 4; see also Figures S7 and S8). The following properties characterize the clusters, which are presented (in terms of key streamflow signatures, geographical domains, and dominant processes) in Table 2 and further supported by supporting information Figure S9.

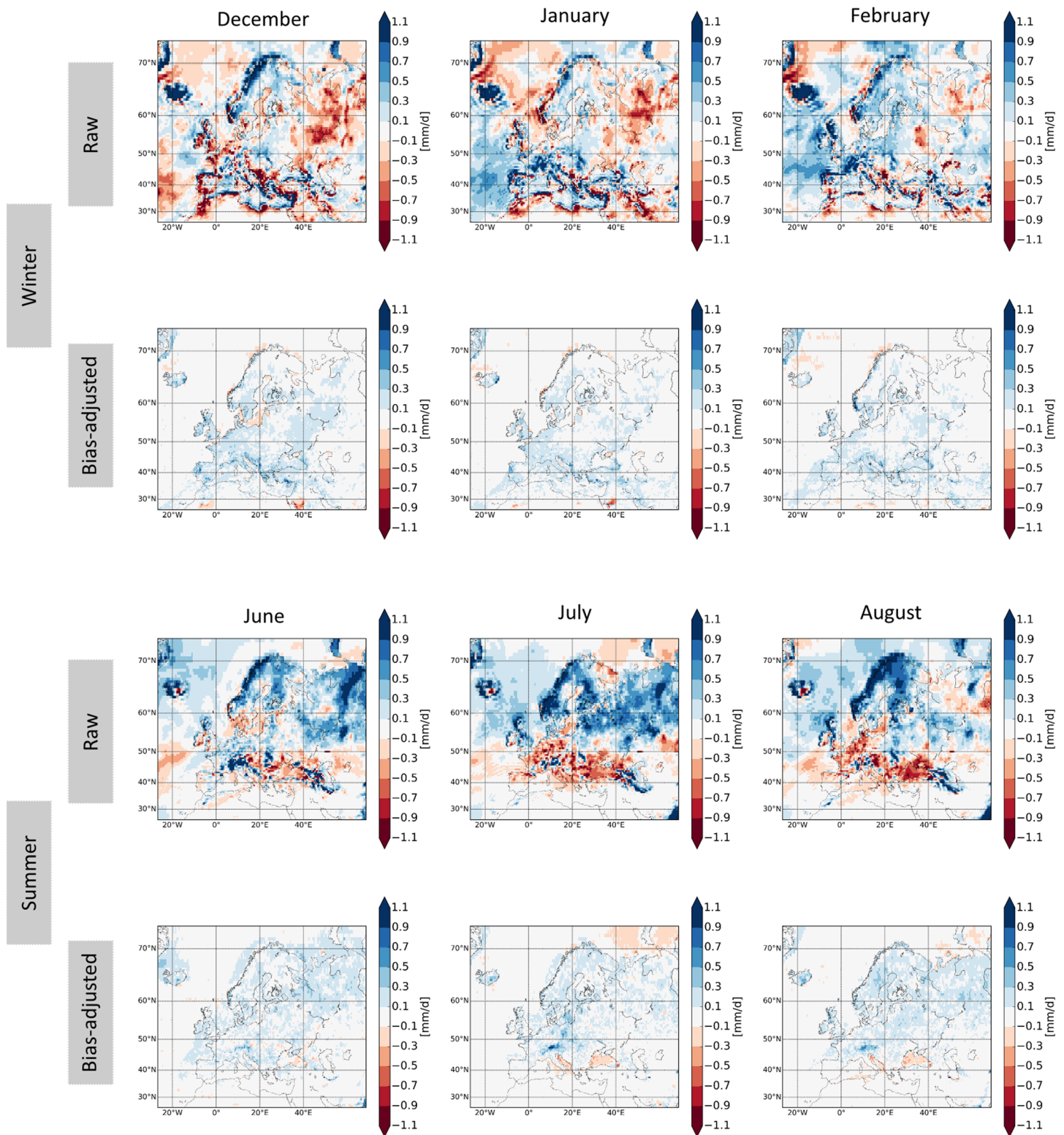


Figure 1. Biases in raw and bias-adjusted precipitation forecasts for the winter and summer months and Lead Month 0.

Basins in Clusters 1 (Poland and Denmark), 8 (located in central Europe), and 9 (Scandinavia and Russia) experience large memory since they are mainly baseflow dominated. These basins have long recessions with small annual variability and hence very little response to precipitation. In particular, large river channels

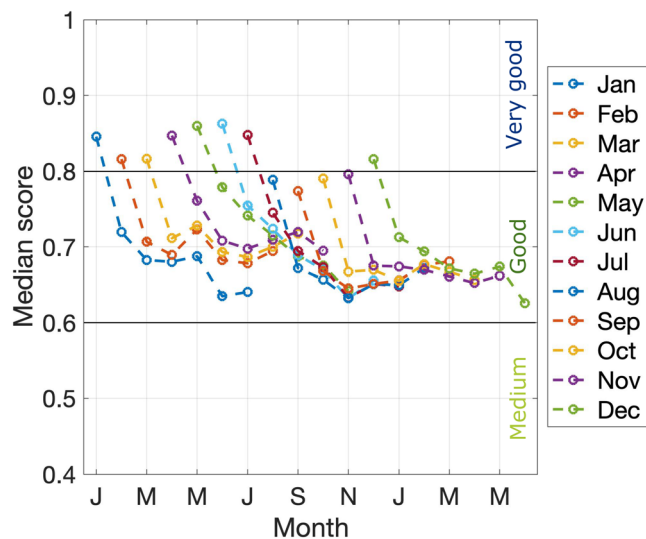


Figure 2. Median forecasting quality (in terms of the CRPS' score) as a function of lead time and initialization month over the entire European domain. Each colored curve corresponds to the hindcasts initialized in a single month.

and water bodies dampen streamflow in basins of Cluster 1. Basins in Cluster 2 are characterized by long recessions in their hydrological response, yet the frequent precipitation events result in frequent peak streamflows. These basins mainly lie in France, Germany, Sweden, south Finland, and Russia. Streamflow regime in north-central and northwestern parts of Russia is driven by snow processes (i.e., snow-melt during spring), while the presence of lakes and wetlands results into dampening of streamflow (Cluster 3). High seasonality caused by snow-melting characterize basins in Clusters 4 and 5, which mostly lie in highly elevated regions and/or cool continental climate. Regulation for hydropower production during winter is also affecting the streamflow regime in Cluster 4.

Basins in Cluster 6 are generally spread around Europe and are very responsive to precipitation yet with long recessions. Basins in Cluster 7 lie in warm and temperate Mediterranean climate yet they are placed at high altitudes while their streamflow regime is characterized by high variability. Streamflow response is highly sensitive to precipitation (and hence interannual variability is driven by precipitation climatology) while evapotranspiration is low. Typical streamflow responses of Mediterranean river systems characterize the basins in Cluster 10.

These systems are located at low elevations and experience low flows and relatively low runoff coefficients due to high evapotranspiration. Finally, basins in Cluster 11 are located in eastern Ukraine and southeastern Russia and are characterized by low runoff coefficient and relatively high annual variability influenced by irrigation.

3.3.2. Linking Forecast Quality to Basin Descriptors

Results from the CART analysis lead to the identification of relationships (for each forecast month and lead month) between predictive accuracy and physiographic-hydrologic-climatic descriptors and consequently to the identification of the key controls, which affect the streamflow forecast quality. An example of the CART tree results for target month December, and Lead Month 2 is provided in supporting information Figure S4b. We next group all CART results and present the ranking of the 10 descriptors for all months and for two lead months (Figure 5). Overall, the descriptors' importance varies as a function of forecast month and lead month, yet some seem to be well identified as key drivers. Results show that the dominant descriptors resulting in poor/good streamflow forecast quality are the basin hydrological behavior (described by FlowID), temperature (Temp), precipitation (Prec), and evaporative index (AET/Prec). It is generally expected that remaining biases in temperature can have a significant impact on the form of precipitation (rainfall or snowfall) and the processes (i.e., changing from (to) snow accumulation to (from) melting) during the cold months. Here the remaining biases in temperature forecasts are negligible to observe an effect on the streamflow forecast quality. However, the remaining biases in precipitation forecasts (BiasPrec) are larger than for temperature and consequently results show a relation to streamflow predictability. Here, we note that a higher importance in BiasPrec could be identified if remaining biases were higher due to the application of an "inappropriate" bias adjustment method. Note also that uncertainty/errors in streamflow volumes at upstream locations are generally propagated further in the downstream basins of the modeling chain. Such alternation in the processes occur in north Europe during April where the mean average temperature is close to 0°C and hence small deviations (or biases) in the meteorological forecasts will affect the basin response. Snow-related processes commonly occurring in highly elevated regions result into a hydrological regime, which is better forecasted in comparison to rain-fed basins. This is the reason that the descriptors of temperature (strongly correlated to snow), evaporative index, and elevation are emerging between March and May.

The basin hydrological similarity (FlowID) seems to be a key descriptor with basins of similar streamflow properties achieving similar predictive performance. There are processes in the river systems (i.e., routing in lakes and lateral groundwater flow) that have higher memory in comparison to the phenomena occurring in the atmosphere, and hence, it is expected that hydrological variables (i.e., streamflow, soil moisture) can have higher predictability than meteorological variables (i.e., precipitation). However, the link between

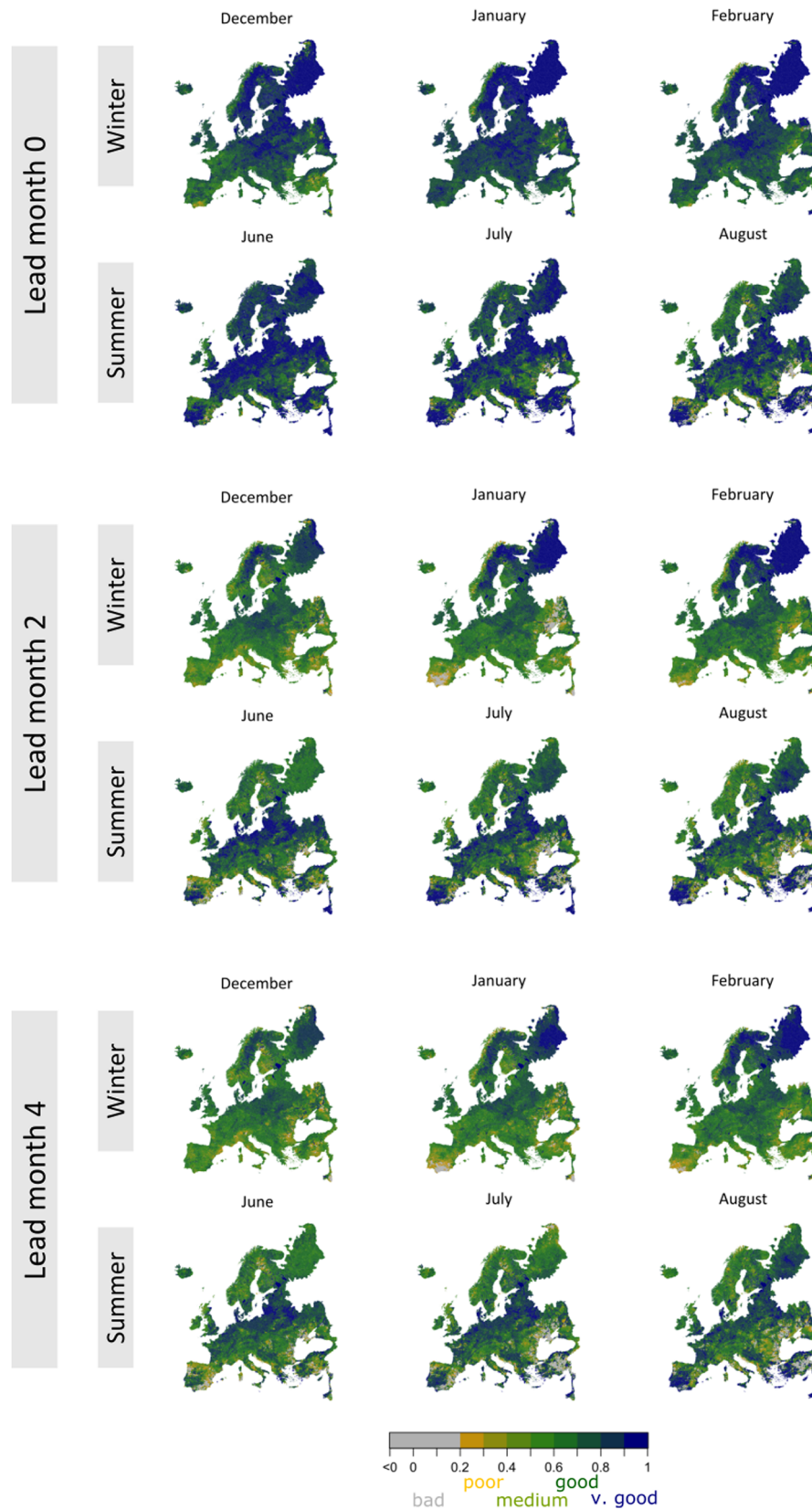


Figure 3. Spatial variability of the forecast quality (in terms of the CRPS' score) for different target months and lead months.

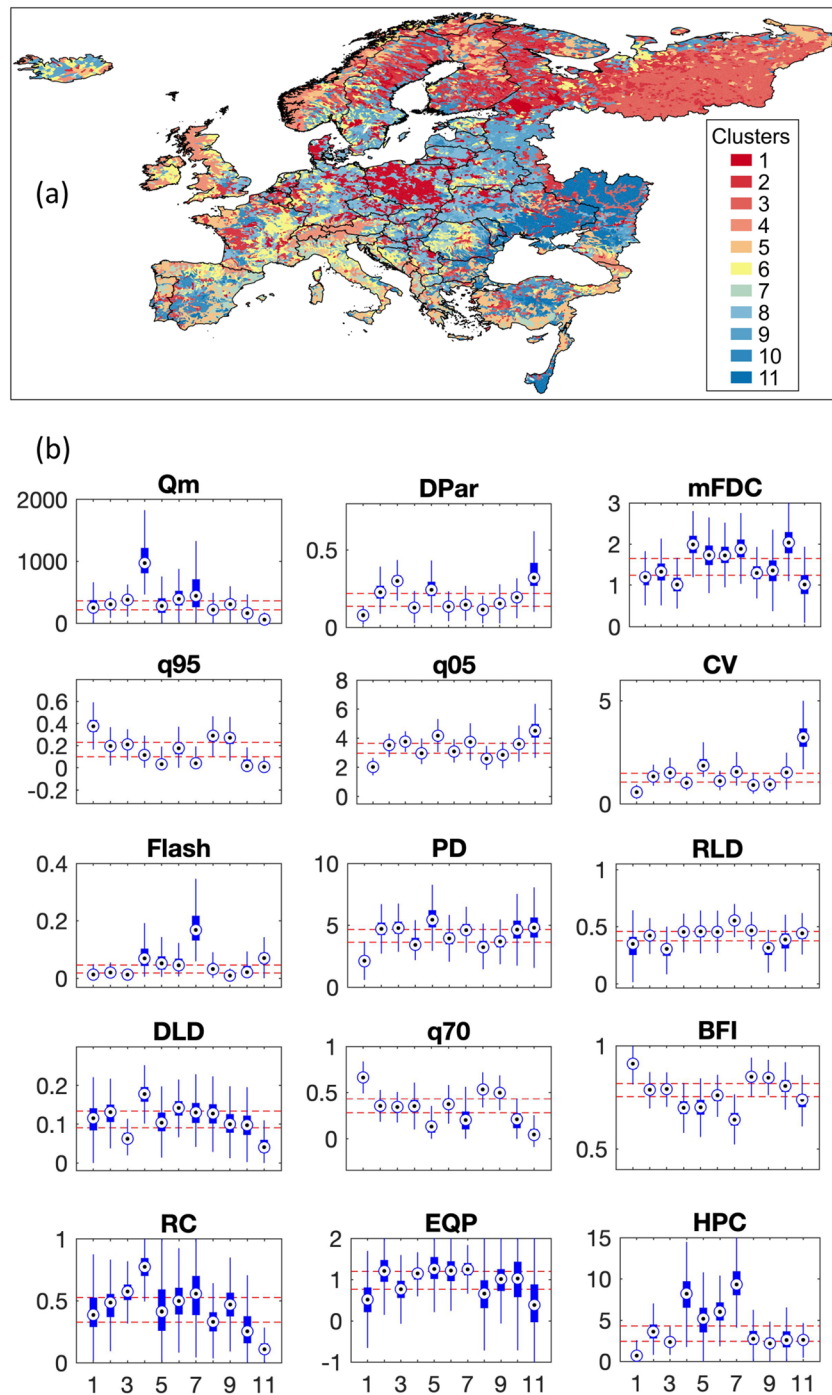


Figure 4. (a) Spatial distribution of hydrologically similar (clusters) basins over Europe and (b) distribution of streamflow signatures in each cluster group. The red lines represent the terciles (33rd and 66th) for each signature.

meteorological and hydrological predictability is not linear, since the precipitation-streamflow process is also nonlinear, with different systems responding differently to the meteorological signal.

3.4. Relating Forecast Quality to Hydrological Processes

We next investigate the pattern of the *CRPS'* score at sites of different hydrological regimes (clusters defined by FlowID) aiming for a deep understanding of the main processes controlling seasonal forecasts of

Table 2
Summary of the Findings on Basin Classification (Key Streamflow Signatures, Geographical Domains, and Dominant Processes)

Cluster group (no. of basins)	Streamflow signatures (FS) FS		Location	Dominant hydrological process
	low FS	high		
1 (2,694)	DPar, q05, CV, Flash, PD, RLD, EQP, HPC	q95, q70, BFI	Poland, Denmark	Streamflow temporal characteristics are controlled by baseflow and hydrographs are not very responsive to precipitation due to dampening by large river channels and water bodies. Some influence of snowmelt-driven streamflows.
2 (4,827)	—	DPar, DLD, EQP	Basins lie at the temperate oceanic and temperate continental climates (spread over Europe but mainly in France, Germany, south Finland, Sweden and Russia)	Streamflow responses, which are precipitation-driven and result into frequent peak streamflows, yet with long recession.
3 (2,870)	mFDC, Flash, RLD, DLD, HPC	DPar, q05, PD, RC	Mainly north-central and northwestern Russia	High interannual variability, particularly variable between the low and high streamflow segments. Snow-dominated streamflow regime with significant snowmelt during spring. Dampening of streamflow due to presence of lakes and wetlands and low actual evapotranspiration.
4 (1,983)	DPar, CV, PD, BFI	Qm, mFDC, Flash, DLD, RC, HPC	Highly elevated regions in Norway, Alps, Iceland, Spain, western Ireland and western UK	Regions receive high precipitation and are very responsive. Occasionally flow is regulated for hydropower production during winter (snow and ice melt), but still with some tendency of spring streamflow.
5 (3,798)	q95, q70, BFI	DPar, mFDC, q05, CV, Flash, PD, EQP, HPC	Most lie at the cool continental (subarctic) climate (mountainous areas of Norway, Sweden, Finland and Russia) and temperate Mediterranean climate (Portugal, Spain, Italy, Greece and Turkey).	Basins of a highly variable streamflow regime. Response is sometimes driven by snow melting, but also precipitation driven.
6 (3,906)	DPar	Qm, mFDC, Flash, DLD, EQP, HPC	Basins spread around Europe but mostly located in southern and central Europe at the temperate continental climate and also south Scandinavia	Highly variable streamflow regime quickly responding to precipitation, yet with long recessions.
7 (1,853)	q95, q70, BFI	Qm, mFDC, Flash, RLD, RC, EQP, HPC	Most lie at the warm and temperate Mediterranean climate (Spain, Italy, Albania, and Greece)	Elevated basins with high variability in the streamflow regime. Precipitation results into flashy streamflow responses, also due to low actual evapotranspiration, which results to high runoff coefficients.
8 (4,450)	DPar, q05, CV, PD, RC, EQP	q95, RLD, q70, BFI	Mostly in central Europe at the temperate continental climate	Hydrographs are baseflow dominated with sharp rising limbs. This corresponds to a small annual variability.
9 (3,658)	q05, CV, Flash, RLD, HPC	q95, q70, BFI	Mainly in Scandinavia and Russia, but also spread around Europe	Highly baseflow dominated streamflow with very little response to precipitation.
10 (2,55-8)	Qm, q95, Flash, q70, RC	mFDC, CV	Mostly in the Mediterranean countries (Portugal, Spain, Greece and Turkey)	Basins (elevation about 300 m a.s.l) with very low streamflows during the entire year, and relatively low runoff coefficients.
11 (2,81-1)	Qm, mFDC, q95, DLD, q70, BFI, RC, EQP	DPar, q05, CV, Flash, PD	Eastern Ukraine and southeastern Russia	Basins with low runoff coefficient, yet they experience relatively high annual variability, that is, fast response to precipitation and fast hydrograph recession. Streamflow could also be influenced by human impacts (i.e., irrigation).

Note. A streamflow signature for the cluster group of interest is identified as high (low) when its median value exceeds (does not exceed) the upper (lower) tercile set from all cluster groups as shown in Figure 4.

streamflow volumes. Figure 6 shows the pattern of forecast quality as a function of forecast month and lead month for eight case studies belonging to different hydrological clusters. The *CRPS'* value remains high for all lead months and months for Cases 1 and 2 (the Angermanälven and Torneälven Rivers) located in Scandinavia. In these basins, snow is accumulated over the winter months and melted over the spring (usually May–July), with this climatological pattern being well defined and adequately predicted in most months. In April, the forecast quality drops in Scandinavia, where temperature is close to 0°C and hence small deviations can erroneously result into snowfall or rainfall. It is important to note that the

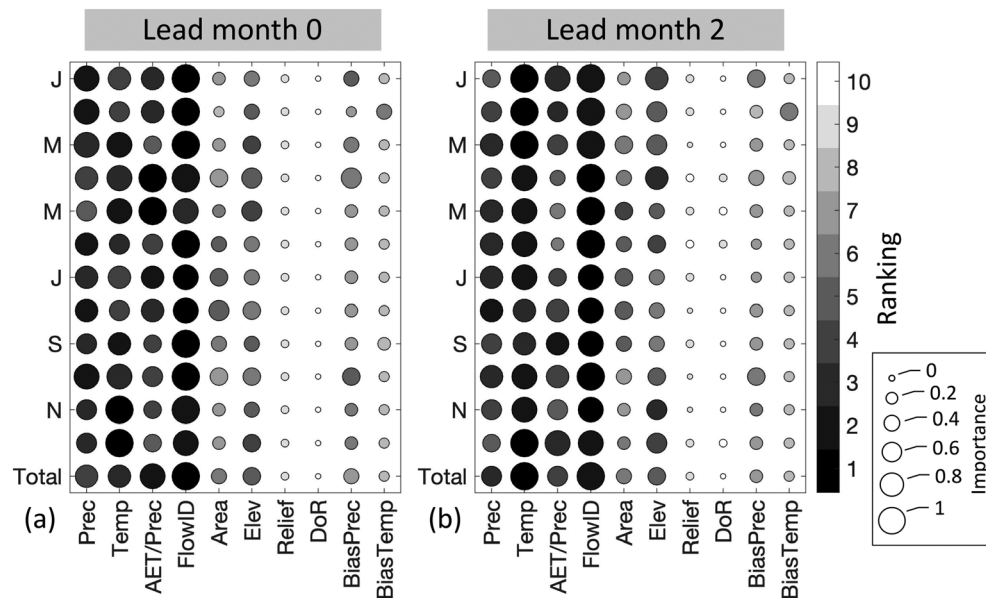


Figure 5. Importance ranking of 10 descriptors (with 1 being the most important) that influence the hydrological forecast quality for all months (and “Total” as the average of each descriptor’s monthly importance) and in lead months: (a) 0 and (b) 2. Shades of gray indicate the ranking of each descriptor, while circle diameters indicate the normalized descriptor’s importance.

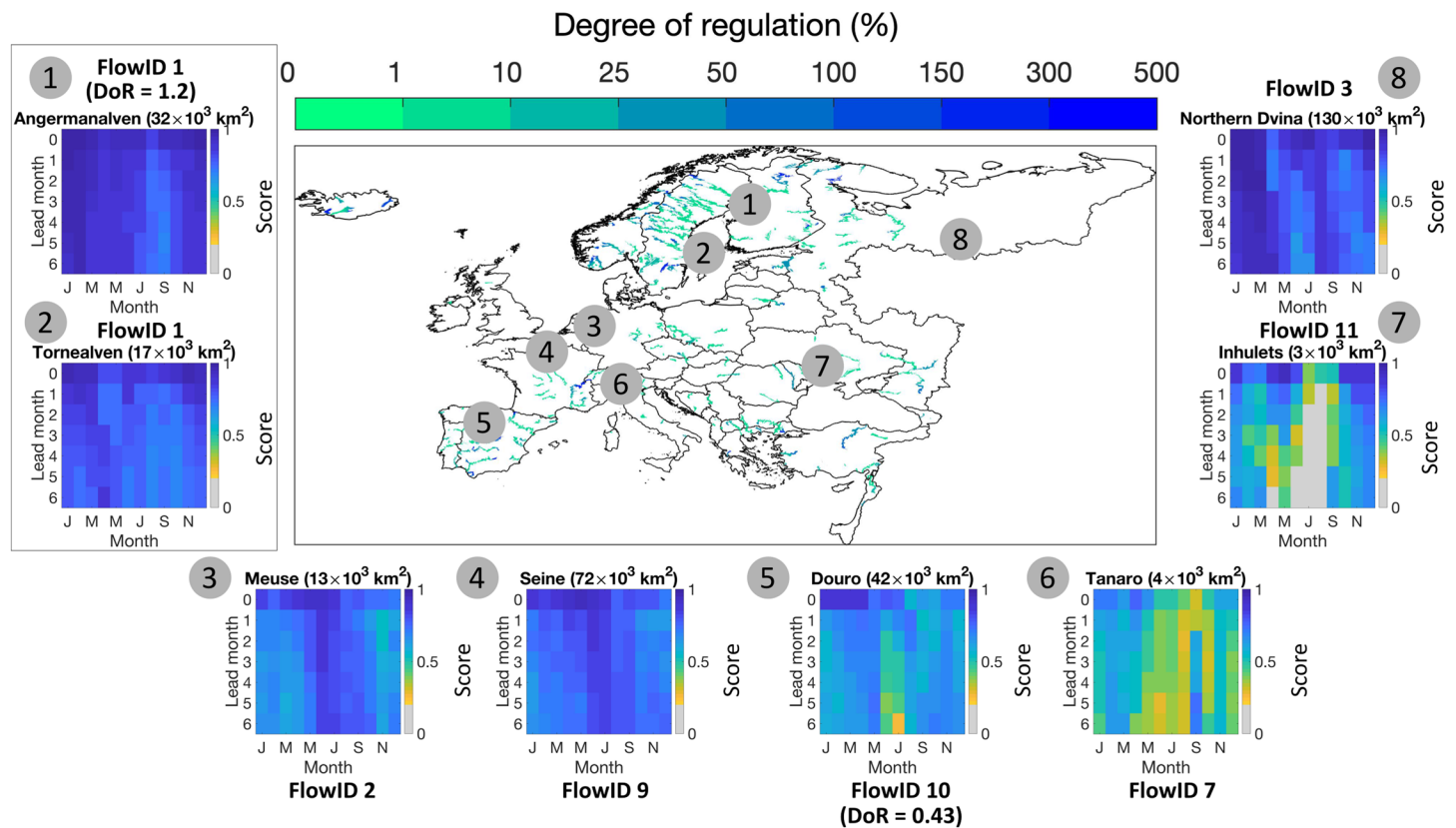


Figure 6. Quality of hydrological forecasts (in terms of $CRPS'$ score) for all months and lead months in eight basins of different hydrological regimes. Central figure shows the degree of streamflow regulation based on the E-HYPE model setup. DoR is shown for Basins 1 and 5, which have a degree of regulation greater than 0%.

Angermanälven basin is regulated for hydropower energy production, and hence, the signal of the seasonal meteorological forecasts can be masked by the regulation scheme (having a constant or seasonal response when water levels are above a threshold) and further affect the naturalized hydrological response. This condition is applied to both the perfect model simulation and SEAS5 model forecasts that consequently results to good predictions. The degree of regulation for Angermanälven is relatively small (DoR of 1.2); however, the forecasting quality is generally higher than the unregulated Torneälven, which has similar climatology and physiography, particularly at the high lead months.

Similarities are further shown at the Sites 3 and 4 (the Meuse and Seine Rivers, respectively). These basins of high river memory due to, for instance, baseflow domination and/or long recessions in the streamflow, show high forecast quality in almost all months and lead months. In addition, in the Mediterranean, the patterns can vary depending on local conditions and human interventions (reservoir regulation and irrigation). Case 5 (the Douro river) is generally characterized by low streamflow and runoff coefficients due to high evapotranspiration rates. Reservoir regulation is affecting the basin's response, which, although DoR is relatively small (0.43), in semiarid regions of low runoff coefficients can influence streamflow forecast quality. The forecast quality varies depending on the month and lead month for the relatively flashy Tanaro River (Case 6). The river's response is driven by precipitation and experiences high variability in the hydrological regime. The *CRPS'* score varies strongly for the fast-responding river system of low runoff coefficient and generally highly variable regime (Case 7). In all seasons, the forecasting quality is generally acceptable in Lead Month 0, yet it (often rapidly) decreases with increasing lead time. Finally, the forecast quality is high in the snow-dominated Northern Dvina river basin (Case 8), whose streamflow is also affected by the lakes and wetlands controlling and dampening the seasonal peaks.

We next assess whether insights from the individual case studies can be regionalized to other locations within Europe. We therefore explore the similarity of the forecast quality (in terms of the *CRPS'* score) within river systems of similar hydrological behavior (Figure 7). As expected, the distribution of the *CRPS'* values for the different clusters varies between the months. However, it is interesting to note that the clusters with good (or bad) forecast quality in relation to the other clusters will always be the same independently of the target month. This is due to the intraannual variability of the streamflow response which consistently varies between the basins from the different clusters. Overall, there is a clear identification of the basins with high (or low) *CRPS'* values. Basins in clusters 1 and 3 have the highest forecast quality of streamflow volumes. Basins in those clusters are characterized by processes of high river memory, for instance, high ranges of baseflow (Cluster 1) and presence of lakes and wetlands, which delay and dampen the streamflow signal (see Table 2) and are thus driven by previous hydrological conditions rather than meteorological forcing. Similar results are observed in Clusters 2, 8, and 9 that are defined by high baseflow index and long recessions. *CRPS'* reaches the lowest range of values in the highly and immediately responsive basins that define Cluster 7. These basins are characterized by short river memory with flashy response to the precipitation signal, high seasonal variability, and small baseflow contribution. Flashy basins that belong to Cluster 11 also experience low *CRPS'* values, particularly in the months when precipitation occurs (autumn and winter). Moreover, forecast quality in the basins of Clusters 4 and 5 is not adequate. These highly elevated basins also have very small baseflow contribution and although they are not as flashy as Cluster 7, their streamflow distribution is characterized by snow melting during spring. Similar insights are overall concluded for higher lead months; however, since the forecast quality decreases as a function of lead month for generally all basins, the distinction of forecast quality between clusters is not preserved as strongly as for Lead Month 0 (see Figure S10 for Lead Month 2).

4. Discussion

4.1. Regionalization Over a Hydroclimatic Gradient

Here for the first time to our knowledge, an investigation demonstrates that the quality of seasonal streamflow forecasts can be clustered, and hence regionalized, based on a priori knowledge of the local hydroclimatic conditions. The insights are of high value to operational continental and global climate services and to users/stakeholders that are dependent on seasonal water fluctuations. In particular, the identified key

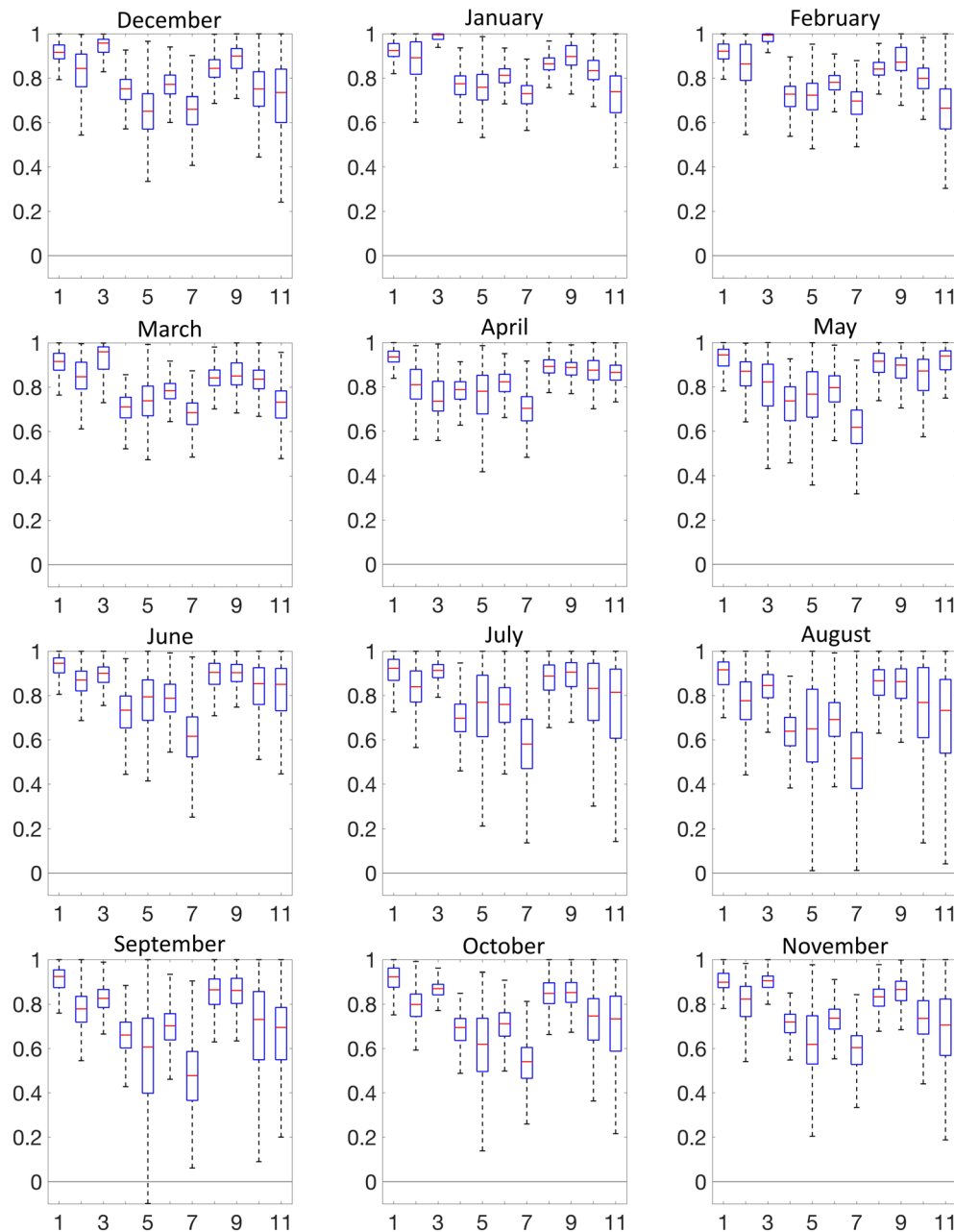


Figure 7. Distribution of streamflow forecast quality (in terms of the $CRPS'$ score) for Lead Month 0 in each cluster group and all target months.

drivers can be used as diagnostics that allow an a priori estimation of the performance of a forecasting service. The insights also set a new scientific scope in seasonal hydrological driven by the genuine interest in identifying additional drivers that can better diagnose the regionalization performance of the forecast quality. Our results show that in general the seasonal streamflow forecast quality is very good at river systems of generally long memory, that is, river systems that are snow dominated and/or experience long recessions, or even systems with lakes and wetlands that dampen streamflow. However, the forecast quality is not adequate in cold and semiarid climates with the river systems immediately responding to the precipitation signal (short river memory). Note that these results are shown in river systems where the biases in the meteorological forecasts are generally small. In regions where these biases are large (i.e., due to the poor performance of the bias adjustment method), the hydrological forecasts are expected to be of poor quality.

This investigation is conditioned to Europe's hydroclimatic gradient; however, we note that the European conditions undoubtedly describe only a portion of the global hydroclimatic gradient. For instance, there are river systems whose streamflow response is dependent on ice processes (i.e., the Himalayan Plateau), and other systems whose response depends on large upstream floodplains (i.e., Niger River Delta; Aich et al., 2016; Andersson et al., 2017; Pechlivanidis et al., 2016). This indicates that there can be hydrological clusters that are not represented in this investigation, and hence might have an unexplored relationship to the drivers used here. Although our investigation focuses on Europe, the methodology followed to identify the key drivers is not limited to any continental scale, and a geographic extension of the investigation is subject to future research using existing global hydrological setups (Arheimer et al., 2019; Emerton et al., 2018).

4.2. Uncertainty in Seasonal Streamflow Forecasting

The quality of seasonal streamflow forecasts relies on a forecasting chain that includes at least seasonal meteorological forcing, initialization of hydrological model states and a hydrologic model setup (Mazrooei et al., 2015; Pechlivanidis et al., 2014). To improve the forecast quality and further the decision-making, this chain can be advanced by introducing additional components that allow assimilation of data to set the initial model states (e.g., in situ/Earth observations of soil moisture and snow water equivalent; Draper & Reichle, 2015; Griessinger et al., 2016; Liu et al., 2012; Musuza et al., 2020), postprocessing of seasonal meteorological forecasts (e.g., bias adjustment and model output statistics; Dobrynin et al., 2018; Manzananas et al., 2019; Zhao et al., 2017), and postprocessing of hydrologic forecasts (e.g., conditioning to local data; Lucatero et al., 2018; Madadgar et al., 2014; Wood & Schaake, 2008). Currently, forecast service development is ad hoc with improvements made to single parts of the forecasting chain when and where available, and with only very limited guidance on the relative importance of each component to the forecasting chain performance (Arheimer et al., 2011; Sinha et al., 2014; Thibault et al., 2016).

To date, only few investigations identified the dominant sources of predictability in seasonal hydrological forecasting (i.e., the initial hydrological conditions and meteorological forcing) at the continental and global scale; however, these only consider different forcing data, model setups, and benchmarking (Greuell et al., 2019; Li et al., 2009; Shukla & Lettenmaier, 2011; Yossef et al., 2013, 2017; Zhang et al., 2017). The lack of large-sample studies across a variety of modeling settings at multiple spatiotemporal scales and under changing environmental conditions has limited the understanding of how predictability evolves in space and time. Sensitivity analysis methods have been used at place-based investigations for seasonal streamflow forecasting to (1) identify critical lead times after which the streamflow predictability mainly depends on the meteorological forecasts (and less on initial conditions; Wood & Lettenmaier, 2008) and (2) quantify the increase in hydrological predictability as a results of increasing the predictability in one of the dominant sources (Arnal et al., 2017; Wood et al., 2016). Application of these methods at the continental/global scale can better exploit our understanding of the sources of predictability in seasonal predictions supplying the users/stakeholders with evidence to guide forecast developments.

Results here show that the streamflow forecast quality depends on the biases in the precipitation forecasts, with large biases being capable of masking the potential of a streamflow forecasting service. Although the existing bias adjustment methods can significantly remove biases in temperature, considerable biases could remain in precipitation. However, note that the requirements in meteorological variables depend on the hydrological models; for instance, wind speed, humidity, and solar radiation are being used in similar hydrological model setups (Arheimer et al., 2019). Promising areas to improve the quality and utility of seasonal meteorological forecasts include a mixture of longer hindcast data sets and improved bias adjustment methods, capable of for instance taking into account the joint variability of multiple variables (Clark et al., 2017; Li et al., 2014). Moreover, model output statistics, which is a type of statistical postprocessing, can produce more reliable seasonal forecasts than in typically available forecasts from climate prediction systems (Wood & Schaake, 2008). In particular, model output statistics-type methods have shown considerable advantages over simple bias adjustment methods, that is, setting climatology-like forecasts in the absence of seasonal forecast skill at long lead times (Zhao et al., 2017).

Finally, the hydrological model (setup, data, structure, and parameters) is another source of uncertainty. Here the analysis was conducted using pseudo-observations as reference, which are not always comparable to real observations, but provide complete information in the spatial and temporal domain. Nevertheless, the assessment against pseudo-observations reduced model errors from the analysis to the minimum, and

Acknowledgments

This study was partially funded by the EU Horizon 2020 project IMPREX (Improving predictions and management of hydrological extremes) under Grant Agreement 641811. Funding was also received from the EU Horizon 2020 project S2S4E (Subseasonal to seasonal forecasting for the energy sector) under Grant Agreement 776787. This study was also partially funded by the EU Horizon 2020 project PrimeWater (Delivering advanced predictive tools from medium to seasonal range for water dependent industries exploiting the cross-cutting potential of EO and hydroecological modeling) under the Grant Agreement 870497. This study was also partially funded by the EU Horizon 2020 project CLARA (Climate forecast enabled knowledge services) under the Grant Agreement 730482. Finally, this study was partially funded by the project named “Long-term forecasts of wind and hydropower supply in one fluctuating climate—Importance for production planning and investments in energy storage and power transmission” funded by the Swedish Energimyndigheten under Grant Agreement 46412-1. The investigation was performed at the SMHI Hydrology Research unit, where much work benefits from joint efforts in developing models and concepts by the whole team. The seasonal meteorological forecasts SEAS5 from the European Centre for Medium-Range Weather Forecasts are freely accessible from the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/>). The HYPE model code is available from the HYPEweb portal (<http://hypeweb.smhi.se/model-water/>). Real-time seasonal forecasts obtained through E-HYPE are openly available on the HYPEweb portal (<http://hypeweb.smhi.se/explore-water/forecasts/seasonal-forecasts-europe/>). The authors would finally like to express their sincere gratitude to Marc Bierkens and two anonymous reviewers for their constructive comments. I. G. P. contributed with the basic idea, the overall study design, result analysis, figures, and writing the manuscript; L. C. contributed with the hydrological model runs, the interpretation of results, and commenting on the manuscript; J. R. contributed with the downloading and reformatting of the meteorological historical data and forecasts; T. B. contributed with the bias adjustment of the seasonal meteorological forecasts and commenting on the manuscript.

hence, results can be attributed to the hydrological processes rather than to the model performance (Bierkens and Van Beek, 2009; Crochemore et al., 2020; Van Dijk et al., 2013). In the case of real observations, small sensitivities (e.g., bias adjustment) would have been hard to detect. However, results would have a direct meaning to users and stakeholders (Bruno Soares & Dessai, 2016; Buontempo et al., 2018); hence, improving the hydrological model setup would be another path to improve the seasonal streamflow forecasts.

5. Conclusions

Herein, we analyzed the seasonal forecasts of streamflow volumes over Europe from the E-HYPE hydrological model forced with bias-adjusted European Centre for Medium-Range Weather Forecasts SEAS5 meteorological forecasts. About 35,400 basins were investigated, which lie along a strong gradient in terms of climatology, scale, and hydrological regime. We further linked the quality of the seasonal streamflow forecasts to a set of physiographic-hydroclimatic descriptors and meteorological biases, which consequently allowed the identification of the key drivers controlling the seasonal streamflow predictability. This investigation sets a benchmark over which further methodologies and systems, beyond those used here, can be tested to assess potential improvements in forecast predictability and its regionalization.

The main conclusions from this study are as follows:

1. SEAS5 meteorological forecasts have biases that need to be adjusted prior to their use in an impact (hydrological) model. These biases are not similar in terms of magnitude and spatial variability; large positive and negative biases in precipitation forecasts with sharp gradients between regions, which coincide with regions of complex topography, and a general underestimation of temperature in most months and lead months. Even when a bias adjustment methodology is applied, remaining biases still exist and their magnitude depends on the variable of interest. Most remaining precipitation biases are located in southern Europe and the Mediterranean and generally in highly elevated regions, while temperature biases are negligible.
2. The European basins can be categorized into 11 clusters based on similarities in streamflow signatures revealing dominating hydrological processes. The hydrological clusters vary spatially in terms of different characteristics of the streamflow signal, that is, mean, variability, extremes, and seasonality. Overall, dominant streamflow generation processes, including baseflow, dampening, human alterations, and climate, could explain the hydrological clustering across Europe. Also, in some regions, distinct patterns of hydrological similarity could appear, for example, mountainous areas, warm Mediterranean region, and central Europe.
3. The quality of the seasonal streamflow forecasts varies both geographically and seasonally, depends on the initialization month, and deteriorates with increased lead months. The highest predictability over Europe overall is shown from April to August, and the predictability decreases in autumn and winter. High forecast quality is shown in central and northern Europe in winter in the short lead months, and in central and western Europe in summer.
4. The quality of the seasonal streamflow forecasts is linked to physiographic and hydroclimatic descriptors, while the descriptors' importance varies with initialization month and lead month. The hydrological similarity, temperature, precipitation, evaporative index, and precipitation forecast biases are strongly linked to the streamflow forecast quality. Seasonal streamflows can be well predicted in river systems of generally long memory (due to snow-related processes, dampening from lakes/wetlands, aquifer contribution, and long recessions); however, the predictability is poor in cold and semiarid climates with the river systems immediately responding to the precipitation signal (short river memory systems).

References

- Aich, V., Liersch, S., Vetter, T., Fournet, S., Andersson, J., Calmanti, S., et al. (2016). Flood projections within the Niger River Basin under future land use and climate change. *Science of the Total Environment*, 562, 666–677. <https://doi.org/10.1016/j.scitotenv.2016.04.021>
- Andersson, J. C. M., Arheimer, B., Traoré, F., Gustafsson, D., & Ali, A. (2017). Process refinements improve a hydrological model concept applied to the Niger River basin. *Hydrological Processes*, 31(25), 4540–4554. <https://doi.org/10.1002/hyp.11376>
- Andersson, J. C. M., Pechlivanidis, I. G., Gustafsson, D., Donnelly, C., & Arheimer, B. (2015). Key factors for improving large-scale hydrological model performance. *European Water*, 49, 77–88.

- Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gavrilenko, N., Gerlitz, L., et al. (2018). Statistical forecast of seasonal discharge in Central Asia for water resources management: Development of a generic linear modelling tool for operational water resource management. *Hydrology and Earth System Sciences*, 22, 2225–2254. <https://doi.org/10.5194/hess-22-2225-2018>
- Arheimer, B., Lindström, G., & Olsson, J. (2011). A systematic review of sensitivities in the Swedish flood-forecasting system. *Atmospheric Research*, 100(2–3), 275–284. <https://doi.org/10.1016/j.atmosres.2010.09.013>
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2019). Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. *Hydrology and Earth System Sciences Discussions*, 1–34. <https://doi.org/10.5194/hess-2019-111>
- Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., et al. (2018). Skillful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, 22, 2057–2072. <https://doi.org/10.5194/hess-22-2057-2018>
- Arnal, L., Wood, A. W., Stephens, E., Cloke, H. L., & Pappenberger, F. (2017). An efficient approach for estimating streamflow forecast skill elasticity. *Journal of Hydrometeorology*, 18, 1715–1729. <https://doi.org/10.1175/JHM-D-16-0259.1>
- Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M., & Michael, K. (2017). Assessment of an ensemble seasonal streamflow forecasting system for Australia. *Hydrology and Earth System Sciences*, 21(12), 6007–6030. <https://doi.org/10.5194/hess-21-6007-2017>
- Berg, P., Donnelly, C., & Gustafsson, D. (2018). Near real-time adjusted reanalysis forcing data for hydrology. *Hydrology and Earth System Sciences*, 22, 989–1000. <https://doi.org/10.5194/hess-22-989-2018>
- Bierkens, M. F. P., & Van Beek, L. P. H. (2009). Seasonal predictability of European discharge: NAO and hydrological response time. *Journal of Hydrometeorology*, 10(4), 953–968. <https://doi.org/10.1175/2009JHM1034.1>
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (2013). *Runoff prediction in ungauged basins. In Synthesis across processes, places and scales*. Cambridge, UK: Cambridge University Press.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth, Belmont, CA: CRC Press.
- Bruno Soares, M., Alexander, M., & Dessai, S. (2017). Sectoral use of climate information in Europe: A synoptic overview. *Climate Services*, 1–16. <https://doi.org/10.1016/j.cliser.2017.06.001>
- Bruno Soares, M., & Dessai, S. (2016). Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, 137(1), 89–103. <https://doi.org/10.1007/s10584-016-1671-8>
- Buontempo, C., Hanlon, H. M., Bruno Soares, M., Christel, I., Soubeyroux, J. M., Viel, C., et al. (2018). What have we learnt from EUPORIAS climate service prototypes? *Clim. Serv.*, 9, 21–32. <https://doi.org/10.1016/j.cliser.2017.06.003>
- Clark, R. T., Bett, P. E., Thornton, H. E., & Scaife, A. A. (2017). Skillful seasonal predictions for the European energy industry. *Environmental Research Letters*, 12, 024002. <https://doi.org/10.1088/1748-9326/aa94a7>
- Contreras, E., Herrero, J., Crochemore, L., Pechlivanidis, I. G., Photiadou, C., Aguilar, C., & Polo, M. J. (2020). Advances in the definition of needs and specifications for a climate service tool operation and management. *Energies*, 13(7), 1827. <https://doi.org/10.3390/en13071827>
- Crochemore, L., Ramos, M.-H., & Pappenberger, F. (2016). Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 20, 3601–3618. <https://doi.org/10.5194/hess-20-3601-2016>
- Crochemore, L., Ramos, M.-H., & Pechlivanidis, I. G. (2020). Can continental models convey useful seasonal hydrologic information at the catchment scale? *Water Resources Research*, 56, 1–21. <https://doi.org/10.1029/2019WR025700>
- Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013). Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, 49, 4035–4053. <https://doi.org/10.1002/wrcr.20294>
- Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4), 245–268. <https://doi.org/10.1002/wcc.217>
- Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Bell, L., Brune, S., Bunzel, F., et al. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, 45, 3605–3614. <https://doi.org/10.1002/2018GL077209>
- Donnelly, C., Andersson, J. C. M., & Arheimer, B. (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, 61(2), 255–273. <https://doi.org/10.1080/02626667.2015.1027710>
- Draper, C., & Reichle, R. (2015). The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual time-scales. *Hydrology and Earth System Sciences*, 19(12), 4831–4844. <https://doi.org/10.5194/hess-19-4831-2015>
- 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-Seasonal v1.0. *Geoscientific Model Development*, 11(8), 3327–3346. <https://doi.org/10.5194/gmd-11-3327-2018>
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, 17(5), 1893–1912. <https://doi.org/10.5194/hess-17-1893-2013>
- Foster, K., Uvo, C. B., & Olsson, J. (2018). The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers. *Hydrology and Earth System Sciences*, 22(5), 2953–2970. <https://doi.org/10.5194/hess-22-2953-2018>
- Giuliani, M., Crochemore, L., Pechlivanidis, I. G., & Castelletti, A. (2020). From skill to value: isolating the influence of end-user behaviour on seasonal forecast assessment. *Hydrology and Earth System Sciences Discussions*, In Review. <https://doi.org/10.5194/hess-2019-659>
- Greuell, W., Franssen, W. H. P., Biemans, H., & Hutjes, R. W. A. (2018). Seasonal streamflow forecasts for Europe—Part I: Hindcast verification with pseudo- and real observations. *Hydrology and Earth System Sciences*, 22, 3453–3472. <https://doi.org/10.5194/hess-22-3453-2018>
- Greuell, W., Franssen, W. H. P., & Hutjes, R. W. A. (2019). Seasonal streamflow forecasts for Europe—Part 2: Sources of skill. *Hydrology and Earth System Sciences*, 23(1), 371–391. <https://doi.org/10.5194/hess-23-371-2019>
- Griessinger, N., Seibert, J., Magnusson, J., & Jonas, T. (2016). Assessing the benefit of snow data assimilation for runoff modeling in Alpine catchments. *Hydrology and Earth System Sciences*, 20(9), 3895–3905. <https://doi.org/10.5194/hess-20-3895-2016>
- Gudmundsson, L., Wagener, T., Tallaksen, L. M., & Engeland, K. (2012). Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. *Water Resources Research*, 48, W11504. <https://doi.org/10.1029/2011WR010911>
- Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463–477. <https://doi.org/10.5194/hess-18-463-2014>
- Hansen, J. W., Mason, S. J., Sun, L., & Tall, A. (2011). Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. *Experimental Agriculture*, 47(2), 205–240. <https://doi.org/10.1017/S0014479710000876>
- 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(3), 2023–2039. <https://doi.org/10.5194/hess-22-2023-2018>

- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5), 559–570. [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2)
- Hewitt, C., Buontempo, C., Newton, P., Doblas-Reyes, F., Jochumsen, K., & Quadfasel, D. (2017). Climate observations, climate modeling, and climate services. *Bulletin of the American Meteorological Society*, 98(7), 1503–1506. <https://doi.org/10.1175/BAMS-D-17-0012.1>
- Hundechea, Y., Arheimer, B., Donnelly, C., & Pechlivanidis, I. G. (2016). A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, 6, 90–111. <https://doi.org/10.1016/j.ejrh.2016.04.002>
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>
- Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., et al. (2017). Intercomparison of regional-scale hydrological models in the present and future climate for 12 large river basins worldwide—A synthesis. *Environmental Research Letters*, 12(10), 105,002. <https://doi.org/10.1088/1748-9326/aa8359>
- Kuentz, A., Arheimer, B., Hundechea, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6), 2863–2879. <https://doi.org/10.5194/hess-21-2863-2017>
- Kumar, R., Livneh, B., & Samaniego, L. (2013). Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme. *Water Resources Research*, 49, 5700–5714. <https://doi.org/10.1002/wrcr.20431>
- Lavers, D. A., Ramos, M. H., Magnusson, L., Pechlivanidis, I., Klein, B., Prudhomme, C., et al. (2020). A vision for hydrological prediction. *Atmosphere (Basel)*, 11(3), 237. <https://doi.org/10.3390/atmos11030237>
- Lehner, B., Liermann, C. R., Revenga, C., Vösmarty, C., Fekete, B., Crouzet, P., et al. (2011). Global Reservoir and Dam (GRanD) database—Technical documentation version 1.1, Bonn.
- Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., & Michalak, A. M. (2014). Joint bias correction of temperature and precipitation in climate model simulations. *Journal of Geophysical Research: Atmospheres*, 119, 153–162. <https://doi.org/10.1002/2014JD022514>
- Li, H., Luo, L., Wood, E. F., & Schaake, J. (2009). The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research*, 114, D04114. <https://doi.org/10.1029/2008JD010969>
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B. (2010). Development and testing of the HYPE (hydrological predictions for the environment) water quality model for different spatial scales. *Hydrology Research*, 41(3–4), 295–319. <https://doi.org/10.2166/nh.2010.007>
- Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., et al. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10), 3863–3887. <https://doi.org/10.5194/hess-16-3863-2012>
- Loh, W.-Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1, 14–23. <https://doi.org/10.1002/widm.8>
- Lucatero, D., Madsen, H., Refsgaard, J. C., Kidmose, J., & Jensen, K. H. (2018). Seasonal streamflow forecasts in the Ahlegraarde catchment, Denmark: The effect of preprocessing and post-processing on skill and statistical consistency. *Hydrology and Earth System Sciences*, 22(7), 3601–3617. <https://doi.org/10.5194/hess-22-3601-2018>
- Madaadgar, S., Moradkhani, H., & Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, 28(1), 104–122. <https://doi.org/10.1002/hyp.9562>
- Manzanas, R., Gutierrez, J. M., Bhend, J., Hermi, S., Doblas-Reyes, F. J., Torralba, V., et al. (2019). Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Climate Dynamics*, 53(3–4), 1287–1305. <https://doi.org/10.1007/s00382-019-04640-4>
- Mazrooei, A., Sinha, T., Sankarasubramanian, A., Kumar, S., & Peters-Lidard, C. D. (2015). Decomposition of sources of errors in seasonal streamflow forecasting over the U.S. Sunbelt. *Journal of Geophysical Research: Atmospheres*, 120, 11,809–11,825. <https://doi.org/10.1002/2015JD023687>
- Meißner, D., Klein, B., & Ionita, M. (2017). Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe. *Hydrology and Earth System Sciences*, 21(12), 6401–6423. <https://doi.org/10.5194/hess-21-6401-2017>
- 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>
- Musuuz, J. L., Gustafsson, D., Pimentel, R., Crochemore, L., & Pechlivanidis, I. (2020). Impact of satellite and in situ data assimilation on hydrological predictions. *Remote Sensing*, 12(5), 811. <https://doi.org/10.3390/RS12050811>
- Nazemi, A., & Wheeler, H. S. (2015). On inclusion of water resource management in earth system models—Part 1: Problem definition and representation of water demand. *Hydrology and Earth System Sciences*, 19, 33–61. <https://doi.org/10.5194/hess-19-33-2015>
- Pechlivanidis, I. G., & Arheimer, B. (2015). Large-scale hydrological modelling by using modified PUB recommendations: The India-HYPE case. *Hydrology and Earth System Sciences*, 19, 4559–4579. <https://doi.org/10.5194/hess-19-4559-2015>
- Pechlivanidis, I. G., Arheimer, B., Donnelly, C., Hundechea, Y., Huang, S., Aich, V., et al. (2017). Analysis of hydrological extremes at different hydro-climatic regimes under present and future conditions. *Climatic Change*, 141(3), 467–481. <https://doi.org/10.1007/s10584-016-1723-0>
- Pechlivanidis, I. G., Bosshard, T., Spångmyr, H., Lindström, G., Gustafsson, D., & Arheimer, B. (2014). Uncertainty in the Swedish operational hydrological forecasting systems. In M. Beer, S. K. Au, & J. M. Hall (Eds.), *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation and Management* (pp. 253–262). UK: Liverpool.
- Pechlivanidis, I. G., Gupta, H., & Bosshard, T. (2018). An information theory approach to identifying a representative subset of hydro-climatic simulations for impact modeling studies. *Water Resources Research*, 54, 1–14. <https://doi.org/10.1029/2017WR022035>
- Pechlivanidis, I. G., Olsson, J., Bosshard, T., Sharma, D., & Sharma, K. C. (2016). Multi-basin modelling of future hydrological fluxes in the Indian subcontinent. *Water*, 8(177), 1–21. <https://doi.org/10.3390/w8050177>
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., et al. (2016). Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology*, 17, 287–307. <https://doi.org/10.1175/JHM-D-15-0054.1>
- Samaniego, L., Kumar, R., Breuer, L., Chamorro, A., Flörke, M., Pechlivanidis, I. G., et al. (2017). Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins. *Climatic Change*, 141(3), 435–449. <https://doi.org/10.1007/s10584-016-1778-y>
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, 15, 2895–2911. <https://doi.org/10.5194/hess-15-2895-2011>

- Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences*, 15(11), 3529–3538. <https://doi.org/10.5194/hess-15-3529-2011>
- Sinha, T., Sankarasubramanian, A., & Mazrooei, A. (2014). Decomposition of sources of errors in monthly to seasonal streamflow forecasts in a rainfall-runoff regime. *Journal of Hydrometeorology*, 15(6), 2470–2483. <https://doi.org/10.1175/JHM-D-13-0155.1>
- Staudinger, M., & Seibert, J. (2014). Predictability of low flow—An assessment with simulation experiments. *Journal of Hydrology*, 519, 1383–1393.
- The MathWorks, I (2019). Statistics and machine learning toolbox, Natick, Massachusetts, United State, <https://www.mathworks.com/help/stats/>
- Thibault, A., Anctil, F., & Boucher, M. A. (2016). Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrology and Earth System Sciences*, 20(5), 1809–1825. <https://doi.org/10.5194/hess-20-1809-2016>
- 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>
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, 17(6), 2263–2279. <https://doi.org/10.5194/hess-17-2263-2013>
- Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., & Wood, E. F. (2019). Development and evaluation of a Pan-European multi-model seasonal hydrological forecasting system. *Journal of Hydrometeorology*, 20, 99–115. <https://doi.org/10.1175/JHM-D-18-0040.1>
- Westerberg, I., & McMillan, H. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, 19, 3951–3968. <https://doi.org/10.5194/hess-19-3951-2015>
- Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., & Clark, M. (2016). Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. *Journal of Hydrometeorology*, 17(2), 651–668. <https://doi.org/10.1175/JHM-D-14-0213.1>
- Wood, A. W., & Lettenmaier, D. P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, 35, L14401. <https://doi.org/10.1029/2008GL034648>
- 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>
- Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30, 1756–1774. <https://doi.org/10.1016/j.advwatres.2007.01.005>
- Yang, W., Andréasson, J., Graham, P. L., Olsson, J., Rosberg, J., & Wetterhall, F. (2010). Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. *Hydrology Research*, 41(3–4), 211–229. <https://doi.org/10.2166/nh.2010.00>
- Yossef, N., Van Beek, R., Weerts, A., Winsemius, H., & Bierkens, M. F. P. (2017). Skill of a global forecasting system in seasonal ensemble streamflow prediction. *Hydrology and Earth System Sciences*, 21(8), 4103–4114. <https://doi.org/10.5194/hess-21-4103-2017>
- Yossef, N. C., Winsemius, H., Weerts, A., Van Beek, R., & Bierkens, M. F. P. (2013). Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing. *Water Resources Research*, 49, 4687–4699. <https://doi.org/10.1002/wrcr.20350>
- 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 1: Understanding the role of initial hydrological conditions. *Hydrology and Earth System Sciences*, 20(6), 2437–2451. <https://doi.org/10.5194/hess-20-2437-2016>
- Zhang, X., Tang, Q., Leng, G., Liu, X., Li, Z., & Huang, Z. (2017). On the dominant factor controlling seasonal hydrological forecast skill in China. *Water*, 9(902), 1–19. <https://doi.org/10.3390/w9110902>
- Zhao, T., Bennett, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E., & Ramos, M. H. (2017). How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *Journal of Climate*, 30(9), 3185–3196. <https://doi.org/10.1175/JCLI-D-16-0652.1>