

## Development and Evaluation of a Pan-European Multimodel Seasonal Hydrological Forecasting System

NIKO WANDERS,<sup>a</sup> STEPHAN THOBER,<sup>b</sup> ROHINI KUMAR,<sup>b</sup> MING PAN,<sup>c</sup> JUSTIN SHEFFIELD,<sup>d</sup>  
LUIS SAMANIEGO,<sup>b</sup> AND ERIC F. WOOD<sup>c</sup>

<sup>a</sup> Department of Physical Geography, Utrecht University, Utrecht, Netherlands

<sup>b</sup> Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany

<sup>c</sup> Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

<sup>d</sup> Geography and Environment, University of Southampton, Southampton, United Kingdom

(Manuscript received 1 March 2018, in final form 1 October 2018)


### ABSTRACT

Hydrological forecasts with a high temporal and spatial resolution are required to provide the level of information needed by end users. So far high-resolution multimodel seasonal hydrological forecasts have been unavailable due to 1) lack of availability of high-resolution meteorological seasonal forecasts, requiring temporal and spatial downscaling; 2) a mismatch between the provided seasonal forecast information and the user needs; and 3) lack of consistency between the hydrological model outputs to generate multimodel seasonal hydrological forecasts. As part of the End-to-End Demonstrator for Improved Decision Making in the Water Sector in Europe (EDgE) project commissioned by the Copernicus Climate Change Service (ECMWF), this study provides a unique dataset of seasonal hydrological forecasts derived from four general circulation models [CanCM4, GFDL Forecast-Oriented Low Ocean Resolution version of CM2.5 (GFDL-FLOR), ECMWF Season Forecast System 4 (ECMWF-S4), and Météo-France LFPW] in combination with four hydrological models [mesoscale hydrologic model (mHM), Noah-MP, PCRaster Global Water Balance (PCR-GLOBWB), and VIC]. The forecasts are provided at daily resolution, 6-month lead time, and 5-km spatial resolution over the historical period from 1993 to 2012. Consistency in hydrological model parameterization ensures an increased consistency in the hydrological forecasts. Results show that skillful discharge forecasts can be made throughout Europe up to 3 months in advance, with predictability up to 6 months for northern Europe resulting from the improved predictability of the spring snowmelt. The new system provides an unprecedented ensemble of seasonal hydrological forecasts with significant skill over Europe to support water management. This study highlights the potential advantages of multimodel based forecasting system in providing skillful hydrological forecasts.

### 1. Introduction

Extreme drought and flood events have a large societal impact and occur in all regions of the world and thus are important phenomena to accurately monitor and forecast (Kundzewicz and Kaczmarek 2000; Wanders et al. 2014). Early-warning decision support systems have been designed to provide forecasts of these impactful hydrological extreme events. Operational continental-scale forecasting systems have been developed for Europe [European Flood Awareness System (EFAS); Thielen et al. 2009],

Africa (African Flood and Drought Monitor (AFDM); Sheffield et al. 2014), North America (North American Land Data Assimilation System; Xia et al. 2012), and the entire globe (Global Flood Awareness System; Alfieri et al. 2013). These operational systems provide medium-term flood and drought outlooks with lead time up to 14 days. For this medium forecast range, hydrological model (HM) simulations can be supported by skillful high-resolution weather forecast to produce skillful hydrologic predictions. The high resolution (both space and time) of these short-range predictions can support forecasts that provide information at resolutions that are relevant to end users (Wanders and Wood 2016). Currently, no continental hydrological forecasting system exists that provides high-spatial-resolution forecasts beyond the medium-term 14-day forecast window. The potential for

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Niko Wanders, n.wanders@uu.nl

DOI: 10.1175/JHM-D-18-0040.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

continental-scale seasonal hydrological forecast is currently hampered by increased uncertainty in the meteorological forcing, increased data volumes and a mismatch between the hydrological simulations produced and the needs of forecast end users.

Recent years have seen an increase in skillful meteorological seasonal forecast from national and international weather agencies (e.g., Kirtman et al. 2014), producing forecasts that have spatial and temporal resolutions that are crucial for operational hydrological forecasting systems. High spatial resolution forecasts reduce the need for downscaling and are capable of capturing the impact of finescale topography. These are also required to provide information at local scales relevant for water management (Samaniego et al. 2017b). The increases in temporal resolution resolved the issue of temporal downscaling needed in earlier seasonal hydrological forecasting (Yuan et al. 2015; Thober et al. 2015) that inherently introduced additional uncertainty in the forecasts. The availability of high temporal meteorological forecasts results in a better matching of the spatial and temporal scales of hydrological modeling [typically, (sub)daily,  $\leq 50$  km] with those of the meteorological forecasts. Recent developments show hydrological modeling at continental scales with a spatial scale on the order of 30m–10 km (Chaney et al. 2016; Sutanudjaja et al. 2018). This is closer to the scales at which stakeholders make decisions and where they would like to receive decision support from high-resolution seasonal hydrological forecasts (Bierkens et al. 2015). The recent development in both meteorology and hydrology toward higher resolutions has created an opportunity to fulfill end-user needs and move toward high-resolution seasonal forecasts.

The two main components of a seasonal hydrological forecasting system are the atmospheric model and the hydrologic model, with the former providing the input for the latter. Both of these components carry a substantial uncertainty in the representation of the atmosphere and hydrosphere. While uncertainty in the atmosphere is widely acknowledged by issuing multiple realizations of the same model at a given forecast date (Saha et al. 2010; Kirtman et al. 2014), only one hydrologic model is usually used in operational hydrologic forecasting systems [e.g., EFAS and AFDM; Arnal et al. 2018; Sheffield et al. 2014]. This hydrological model often has a deterministic initial condition that does not reflect the uncertainty in the initial conditions. Improved simulations of forecast uncertainty in a hydrological forecast system can be obtained by 1) using perturbations in the initial conditions, 2) using different model structures, 3) using parametric uncertainty, or 4) making use of different meteorological forcing datasets

(Samaniego et al. 2017c). The correct representation of the uncertainty in hydrologic processes is an ongoing research topic of debate (Clark et al. 2015) and very relevant for seasonal hydrological predictions. Therefore, a robust hydrologic forecast system should use multiple representations of the terrestrial hydrologic cycle having reasonable predictive skill.

There is an increasing demand for multimodel ensemble (MME) systems that have multiple meteorological and hydrological models, to achieve improved accuracy and consistency in the forecasting operations. In seasonal meteorological forecasting, we have seen an increase in the ensemble sizes of MME forecasts, whereas hydrological forecast systems are still provided with single hydrological model setups. MME hydrological setups are faced with the difficulty of differences in parameterization, land surface properties, and surface representation. Most systems are even based on nonseamless parameterizations (Mizukami et al. 2017). For example, channel networks do not have to be identical, leading to minor displacements of major rivers, resulting in significant differences in the locally simulated river discharge. The presence and absence of reservoirs and differences in vegetation cover will add to these model differences and can locally create significant inconsistencies in the predicted hydrological variables. This indicates the importance of having a consistent multimodel hydrological forecasting system.

In this project, we aim to resolve these issues by creating the first high-resolution continental hydrological multimodel ensemble. This MME system will incorporate multiple seasonal meteorological forecasts derived from state-of-the-art dynamical forecast models which are used as forcing to multiple HMs. All HMs have a seamless parameterization (Samaniego et al. 2017b) and a common routing model [mRM, the routing model from the mesoscale hydrologic model (mHM)]. This novel system aims at fulfilling the abovementioned user needs and provides seasonal forecasts at a resolution relevant to users, reduces data volumes by using user-defined indicators, and provides an estimate on the reliability and uncertainty based on the MME system.

The system is tested against the commonly used ensemble streamflow prediction (ESP) baseline forecasts, which use climatological meteorological information in combination with the hydrological initial conditions to produce seasonal forecasts (Wood et al. 2005; Thober et al. 2015). In this study, we will also use a “reverse ESP” to quantify the impact of the initial conditions on the seasonal predictability (Wood and Lettenmaier 2008; Shukla et al. 2013). These two additional hindcast experiments will provide valuable information on the expected skill that can be derived from the historical knowledge of the meteorological and hydrological system.

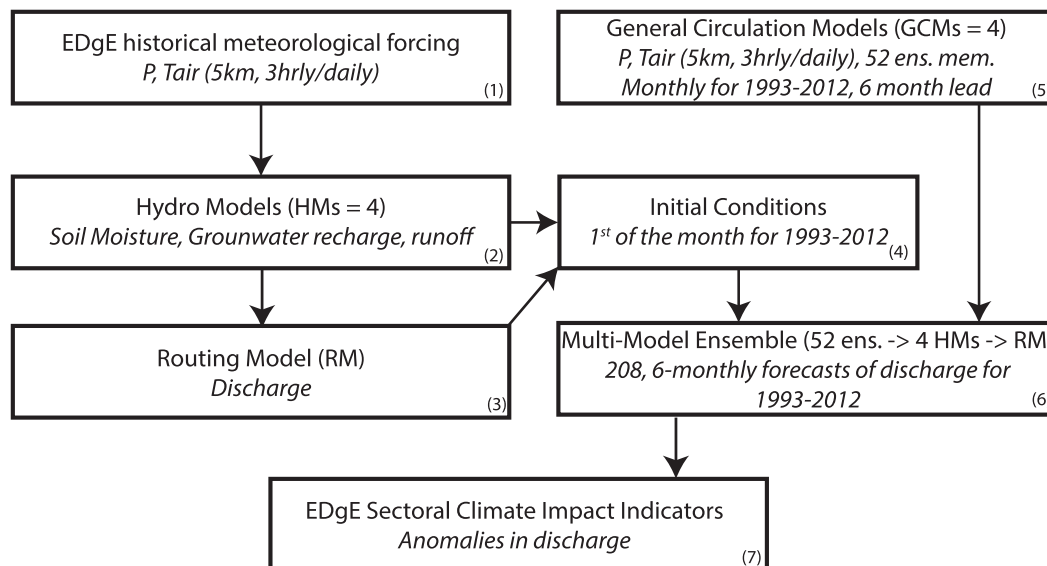


FIG. 1. Model setup of the modeling chain.

In section 2a, we describe the End-to-End Demonstrator for Improved Decision Making in the Water Sector in Europe (EDgE) project of which this MME is a component. Section 2b provides the details on the models, modeling chain, and data, and section 2c provides information about the evaluation metrics. The quality and uncertainty in the forecasting are described in detail in section 3. Finally, we discuss the implications of the obtained results and potential application of this MME in section 4.

## 2. Material and methods

### a. The EDgE project

The novel seasonal forecasting system developed in this study is part of the EDgE project, which is a proof of concept funded by the Copernicus Climate Change Service program and operationalized by the ECMWF (Samaniego et al. 2018, manuscript submitted to *Bull. Amer. Meteor. Soc.*). The EDgE project aims to support the need of European private and public sector stakeholders for improved climate information. Within the project, both climate projections and seasonal hindcasts are provided for the end users and made publically available in a web delivery system (available at <http://edge.climate.copernicus.eu/Apps/#climate-change> and <http://edge.climate.copernicus.eu/Apps/#seasonal>). In this manuscript, we focus on the seasonal forecast information, and more specifically on the hydrological modeling component of the EDgE project. The reader can find all the forecast data on <http://edge.climate.copernicus.eu/Apps/#seasonal>,

where they can look at individual forecasts over Europe.

### b. MME forecasting system

In this study we have used a total of four dynamical meteorological seasonal forecast models from different general circulation models (GCMs) and four large-scale HMs for a period of 19 years (1993–2012) over the pan-European domain to form our multimodel ensemble seasonal hydrological forecasting system. We have created the EDgE historical meteorological forcing that is derived from the E-OBS dataset (Haylock et al. 2008) and is used as a baseline in the MME (Fig. 1). The four HMs, PCRaster Global Water Balance (PCR-GLOBWB), VIC, mHM, and Noah-MP (Table 1), use this historical precipitation and temperature forcing to provide daily simulations of soil moisture content, snow water equivalent, groundwater recharge, and total runoff at a spatial resolution of 5 km for the period 1950–2015. The simulations of runoff feed into the mRM river routing model (Samaniego et al. 2010), which uses a mass conservative, time-adaptive Muskingum–Cunge approach (Todini 2007; Thober et al. 2018, manuscript submitted to *J. Adv. Model. Earth Syst.*) to generate daily discharge estimates across Europe. The initial conditions obtained from the historical simulation are used in the forecasting framework to initialize the hydrological models.

The four GCMs, CanCM4, GFDL Forecast-Oriented Low Ocean Resolution version of CM2.5 (GFDL-FLOR), ECMWF Season Forecast System 4 (ECMWF-S4), and Météo-France LFPW (Table 2), all produce a new seasonal meteorological forecast at the beginning of each

TABLE 1. Properties of the four large-scale hydrological models used in this study.

Model	Temporal resolution	Institute	Reference
PCR-GLOBWB	Daily	Utrecht University, The Netherlands	Sutanudjaja et al. (2018)
VIC	Daily	University of Washington, United States	Liang et al. (1996)
mHM	Daily	Helmholtz Centre for Environmental Research–UFZ, Germany	Samaniego et al. (2010, 2017a)
Noah-MP	3-hourly	National Center for Atmospheric Research, United States	Niu et al. (2011)
mRM	Daily	Helmholtz Centre for Environmental Research–UFZ, Germany	Samaniego et al. (2010)

month for the upcoming 6 months with a daily temporal resolution. CanCM4 and GFDL-FLOR are selected because of their high forecast skill (Wanders and Wood 2016). CanCM4 exhibits a high skill at the short leads ( $\leq 3$  months), while GFDL-FLOR was found to be good at the longer lead times. Both ECMWF-S4 and LFPW were selected since these are regarded as skillful over Europe (Molteni et al. 2011). The meteorological forecasts are downscaled with kriging, using terrain elevation as external drift, to a spatial resolution of 5 km. The E-OBS historical meteorological station data were used to derive variograms for precipitation and temperature. All forecast models were corrected with the same E-OBS variogram, but individual forecast members are corrected separately to keep the original ensemble spread. This produces a seasonal forecast dataset at 5-km resolution with mixed statistical properties. The aggregated statistical properties of the 5-km dataset at the original resolution of the GCMs are identical to those of the original GCM dataset. This ensures that the ensemble spread and spatiotemporal dynamics of the original GCM dataset are preserved, which are the basis for a meaningful analysis of forecasting skill. At the 5-km resolution, the spatial variability is similar to that of the E-OBS dataset. Additionally, the DEM is used as an external drift, which leads to higher precipitation and colder temperatures with higher altitude. Thus, mountains are better represented in the 5-km dataset than in the original coarse-scale GCM data. For Noah-MP and VIC 3-hourly forcing data are required, which are obtained by

applying the Mountain Microclimate Simulation Model (MTCLIM) algorithm (Bohn et al. 2013). This methodology produces a full ensemble of 52 meteorological forecasts at the start of each month for the period 1993–2012, which is used to drive the hydrological models. The four HMs and initial conditions are paired with the 52 meteorological ensemble members to form an ensemble of 208 hydrological forecasts of soil moisture, groundwater recharge, runoff, and discharge conditions for the coming 6-month period. Finally, the EDgE Sectoral Climate Impact Indicators (SCIIs) are then used to study anomalies in the discharge compared to the historical simulations. Within EDgE 34 SCIIs are available, but for this experiment, we only use the discharge quintiles. The quintile ranges are  $Q_0$ – $Q_{20}$ ,  $Q_{20}$ – $Q_{40}$ ,  $Q_{40}$ – $Q_{60}$ ,  $Q_{60}$ – $Q_{80}$ , and  $Q_{80}$ – $Q_{100}$ , and their cutoffs are derived from the reference E-OBS–based simulation that is also used to create the initial conditions. Moreover, these quintile limits are estimated for each month and hydrological model separately for the period 1993–2012. For each forecast quintile, a forecast probability is provided by each individual combination of GCM–HM, and these probabilities are computed for each forecast lead separately.

The ESP provides the baseline seasonal forecast skill that can be obtained from providing known initial hydrological conditions in combination with the average climatological meteorological forcing (e.g., Wood et al. 2005; Thober et al. 2015). A 15-member ESP is created by resampling the historical meteorological observations

TABLE 2. Properties of the four dynamical seasonal forecast models used in this study.

Model	Ensemble size	Spatial resolution	Institute	Reference
CanCM4	10	1.0°	Environment and Climate Change Canada, Canada	Merryfield et al. (2013)
GFDL-FLOR	12	1.0°	Geophysical Fluid Dynamics Laboratory, United States	Vecchi et al. (2014)
ECMWF-S4	15	0.54°	European Centre for Medium-Range Weather Forecasts, United Kingdom	Molteni et al. (2011)
LFPW	15	0.54°	Météo-France, France	Météo-France (2015)
ESP	15	0.05°	—	e.g., Wood et al. (2005); Thober et al. (2015)

from the years 1993–2012, where for each forecast time step 15 random years are selected (excluding the target year). The period 1993–2012 is selected to ensure that trends in temperature and precipitation do not influence the ESP forecasts. In addition, the selection of these years ensures that the climatology is preserved.

We have added a reverse ESP experiment (rev-ESP) to provide the baseline seasonal forecast skill that can be obtained from providing known meteorological forcing in combination with the average hydrological initial conditions (e.g., Wood and Lettenmaier 2008; Shukla et al. 2013). The rev-ESP is created by resampling the historical hydrological initial conditions from the years 1993–2012, where for each forecast time step 15 random years are selected (excluding the target year). The selection of these years ensures that there is now a trend in the hydrological conditions that could impact the results. We have alternated the initial conditions from using only climatological hydrological conditions to selecting the known conditions for the individual hydrological components (e.g., soil moisture, groundwater levels). This experiment was only performed for the PCR-GLOBWB model, because it is highly computationally intensive and not all model structures allow for these perturbations. From the rev-ESP we compute the portion of explained variance compared to a simulation where all initial conditions are known. This explained variance is taken for all seasons and geographical locations together. This will clearly impact the findings, as components like snow will be important for some regions but irrelevant for others. However, to get a more general estimate of the important components, we have focused on the pan-European number.

We use consistent soil, land surface, and land cover data to ensure the largest possible consistency across models. For example, vegetation rooting depths, the thickness of the unsaturated zone, and subsurface properties are consistent across the four hydrological models. Because forcings and all geophysical properties are identical across the models, the impact of the model structural differences and parameterizations on streamflow forecast can be effectively investigated in this study. It is worth noting that a single routing model is selected to minimize the influence of channel network configuration.

### c. Evaluation metrics

#### 1) SKILL METRICS

The seasonal forecasts are evaluated against a 1993–2012 reference simulation that is also used to generate the initial hydrological conditions. An adapted Brier score (BS; Brier 1950) is computed to quantify the skill, reliability, resolution, and uncertainty of the model forecast and is given by

$$QS = \frac{1}{NTQ} \sum_{n=1}^N \sum_{t=1}^T \sum_{q=1}^Q [(R_t \leq p_q) - (P_{n,t,q} \geq p_q)]^2, \quad (1)$$

where  $p_q$  is the statistical likelihood that a forecast will fall in a certain quintile  $q$  which equals 0.2 in this formulation,  $P_{t,n,q}$  is the probability that the forecasted flow falls in quintile  $q$  for ensemble member  $n$  at time  $t$ ,  $R_t$  is the reference-run-driven quintile,  $Q$  is the total number of quintiles ( $Q = 5$ ),  $N$  is the total number of ensemble members, and  $T$  is the total length of the forecast period. The quintile score (QS) ranges between 0 and 1, where 0 indicates perfect forecasts. This implies that the forecasts for a given month always lay within the same quintile as the reference simulation, whereas 1 indicates that none of the forecasts lay in the respective quintile. Since a quintile distribution is applied here, the theoretical no-skill value for the QS is 0.8 because there is a 20% chance of a randomly correct forecast. The QS estimated here is model specific because of its dependency to the reference run simulations.

#### 2) UNCERTAINTY METRICS

The uncertainty contributions of the models are determined using the differences in the QS between the individual model chain components. We distinguish between uncertainty in the GCMs and the HMs, where the uncertainty is defined as the deviation from the average QS for a given model  $m$  and is given by

$$\overline{QS}_m = \frac{1}{N} \sum_{n=1}^N QS_{m,n}, \quad (2)$$

where  $N$  is the number of GCMs or HMs that can be combined with GCM or HM  $m$ . The  $\overline{QS}_m$  acts as the benchmark used to determine the deviation from the normal. The uncertainty score is then defined by

$$U_m = \frac{1}{N} \sum_{n=1}^N (QS_{m,n} - \overline{QS}_m), \quad (3)$$

where  $U_m$  provides an estimate of the uncertainty of the forecasts for model  $m$ . As an example, where  $m$  is one of the hydrological models and  $N$  is the four GCMs,  $U_m$  will then give the uncertainty from the dynamical models for the hydrological model  $m$ . The uncertainties in the QS only describe the variations in the QS within the model ensemble; it will not inform the user on the uncertainty that can result between different forecast years. Due to the low number of years and the potentially strong impact of the seasonality, it would be difficult to get a consistent estimate of the uncertainty between years. Seasonal analysis can be done using the aggregated QS for each season from the individual months, but on the other hand, this method does allow



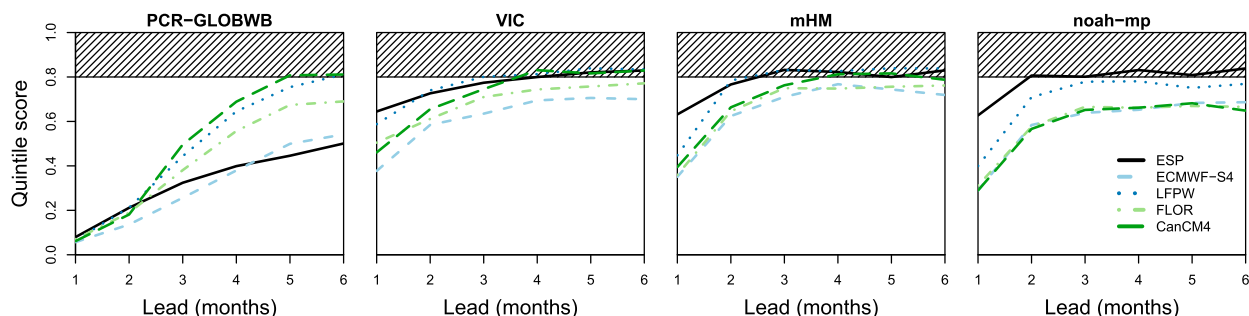


FIG. 2. Pan-European average forecast skill as depicted by the QS for four hydrological models. The baseline is indicated by the solid black line, and the area of no skill is indicated in hashed black. Using the QS formulation and the quintile forecast produced in EDgE, QSs above 0.8 indicate no skill, and a value of 0.0 indicates perfect forecasts.

us to quantify the average uncertainty contribution of the GCMs and HMs by selecting all combinations of GCMs and HMs.

### 3. Results

#### a. Ensemble streamflow prediction baseline

Figure 2 shows the pan-European average QS for all the GCM–HM model combinations separately. It shows that there is a significant spread in the ESP skill for the HMs. This indicates that some models are more sensitive to the initial conditions than others since the climatological meteorological forcings are identical for all models. This sensitivity to the initial conditions is mainly caused by the different process parameterizations used in the HMs. All models show, as expected, that skill is decreasing with lead time, but the extent of this decrease is model dependent. Noah-MP stabilizes already beyond lead times of 2 months, whereas PCR-GLOBWB shows an almost linear decrease with increasing lead times. Due to the unique setup of the EDgE MME setup, all HMs have identical soil and land cover properties information for the derivation of their parameters. Therefore, all differences in baseline skill can be directly attributed to the interaction between the hydroclimatic variables and the model structures. The reference simulations show that the PCR-GLOBWB model has the slowest hydrological response to changes in the meteorology and highlight that it has the largest groundwater discharge among the four considered hydrologic models. The groundwater response of the model is driven by the percolation from the unsaturated zone, resulting in a relatively strong influence of the initial conditions on the predictive skill of the model. The mHM, VIC, and Noah-MP models, on the contrary, exhibit less sensitivity to the initial conditions, due to the relative importance of the surface runoff fluxes that are driven by

precipitation events. For example, Noah-MP is used with its free drainage subsurface runoff option, which gives the best representation to the observed streamflow among the available options provided in the source code (not shown). At the same time, this option has no designated groundwater storage, which leads to a very rapid reaction of the model to changes in the meteorological forcings (e.g., precipitation events). This explains the strong differences in the forecast skill of Noah-MP in the first 3 months of the ESP and shows that model structure can have a dominant impact on the forecast skill when using ESP forecast validated to reference simulations.

Apart from the model structure, we also observe a strong impact of the local hydroclimatic regimes on the ESP seasonal forecast skill (Fig. 3). We observe high skills for Scandinavia, especially for the winter season when discharge predictions are primarily driven by the presence or absence of snow. Snowmelt conditions clearly dominate the forecast skill in most of Europe for the period March–May. Quintile scores for the Rhine and Danube River basins show that ESP has good skill in predicting the onset of the high flows related to the runoff produced by snowmelt.

The differences between HMs are most predominant for the short lead times, but spatial patterns are comparable. The HMs show the highest forecast skills in Scandinavia and lower forecast skills for central Europe (Fig. A1). The highest ESP forecast skill is observed for PCR-GLOBWB, where in addition to the high skill in Scandinavia, high skill is observed for the Baltic states. The other HMs have the tendency to rapidly drop in ESP forecast skill for the long lead times, with the lowest skills being observed in the Rhine and Danube catchments.

#### b. Dynamical streamflow predictions

Four GCMs are used to compute the seasonal prediction skill for the pan-European domain. Compared to the ESP baseline, only the meteorological input data

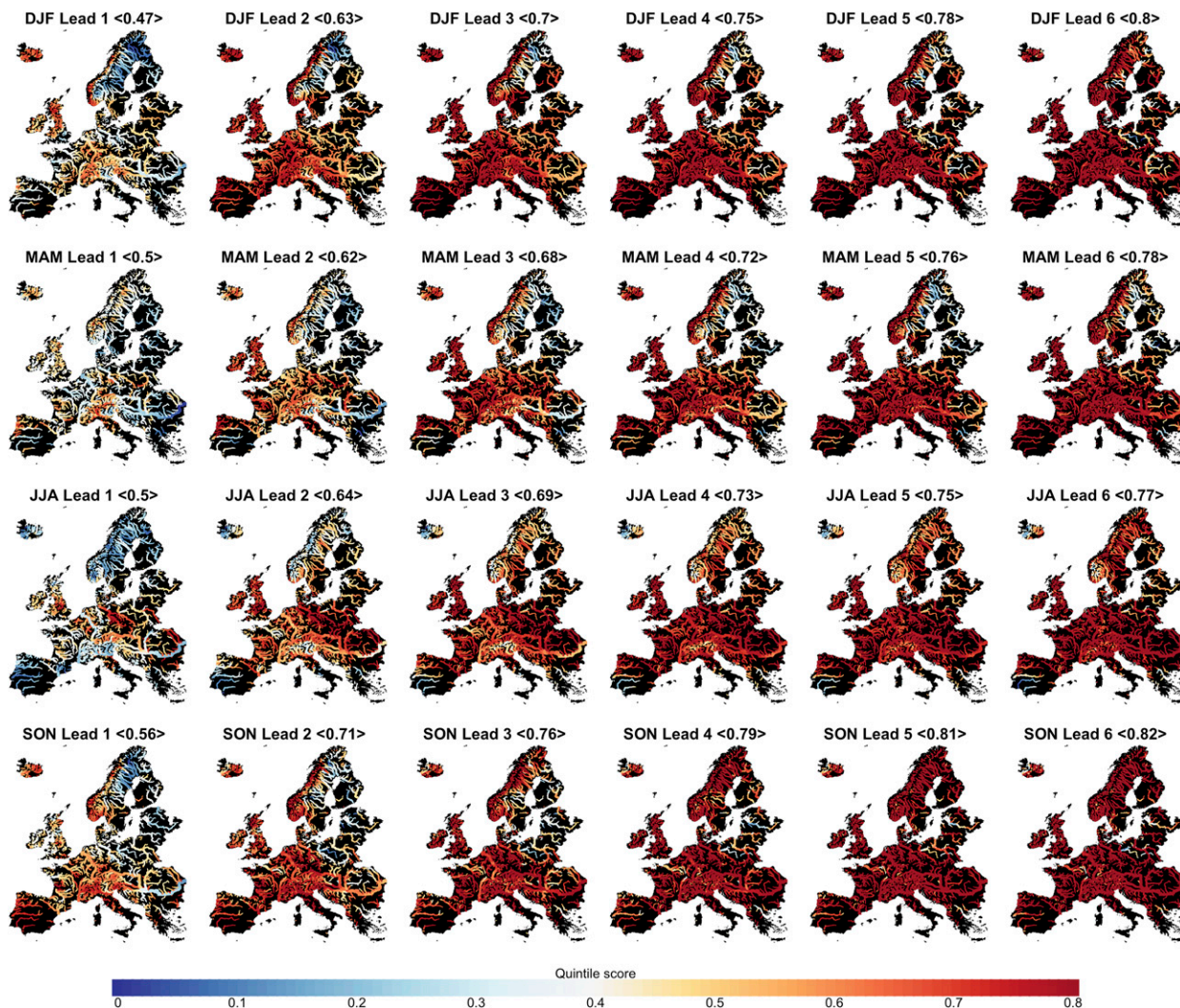


FIG. 3. Pan-European multihydrological model ESP forecast skill as depicted by the seasonal average QS for four target seasons (DJF = winter, MAM = spring, JJA = summer, SON = autumn). The header indicates the lead and pan-European average QS. Using the QS formulation and the quintile forecast produced in EDgE, QSs of 0.8 indicate no skill, and a value of 0.0 indicates perfect forecasts.

were changed to use state-of-the-art dynamical seasonal forecast models from major meteorological services in Europe and North America. Compared to the ESP baseline, we observe major improvements in forecast skill for large parts of Europe, with the exception of eastern and northern Europe (Fig. 4). The northern regions especially show high skill in the ESP baseline forecast and therefore have a low potential for improvement. We observe that the patterns of forecast improvement change with increasing lead times, due to the relative skill in the forecast of precipitation and temperature anomalies at these time scales (Wanders and Wood 2016). With longer lead times we mostly see improvement in the Balkan regions and parts of the

Danube, but we observe no clear improvement in the pan-European forecast skill compared to the ESP.

It is clear from this experiment that the EDgE MME has difficulties in estimating snow accumulation and melt forecasts comparable to the ESP baseline. The small increases and sometimes decreases in forecast skill for the alpine regions indicate a lack of snowpack predictability in high mountains. This is likely caused by the need for skillful precipitation and temperature anomalies. When either one of these forecasts is inaccurate, it will have a significant impact on either the gain or loss in snowpack. The forecast skill in these regions is mostly independent of the HM selection (Fig. A1).



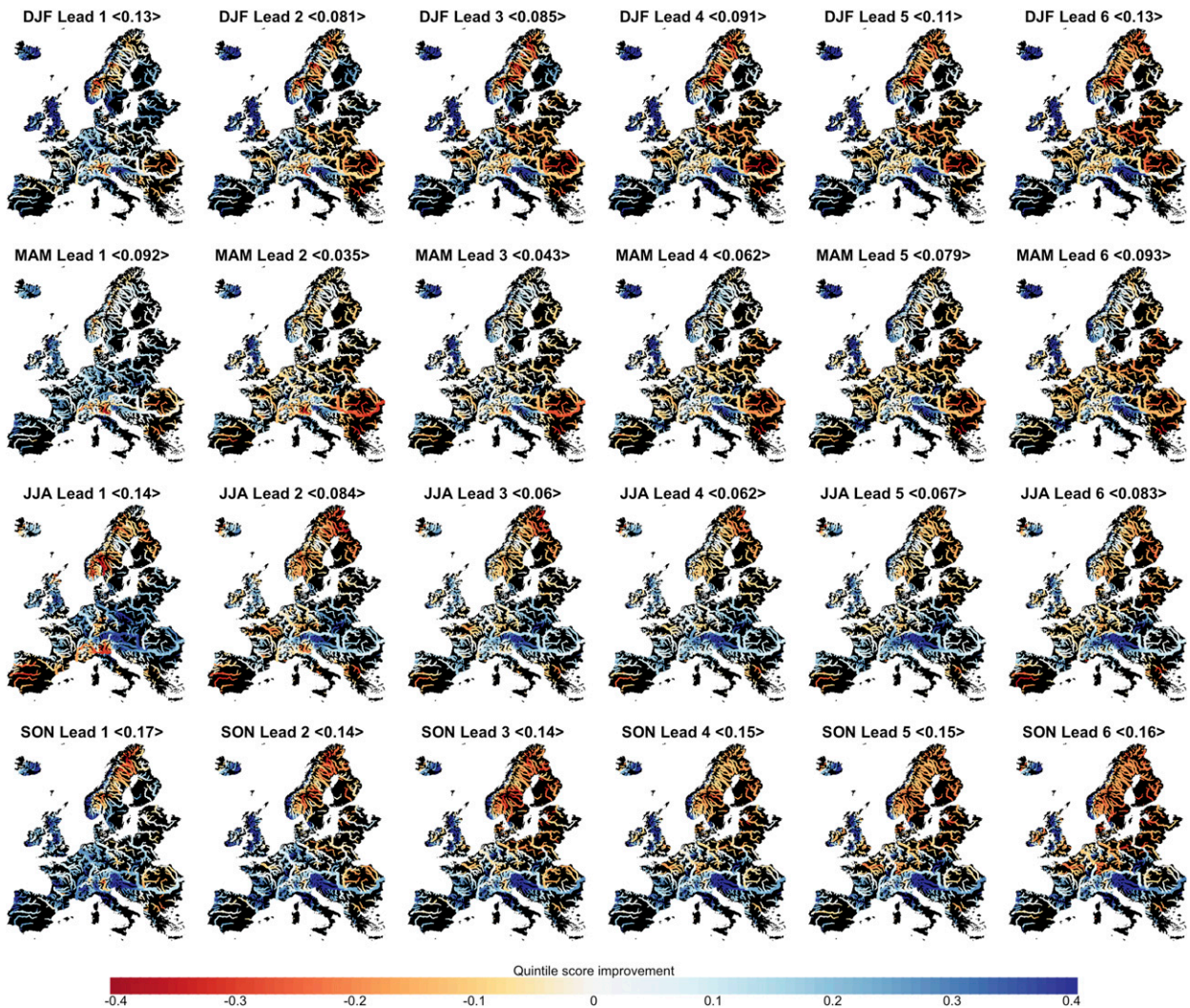


FIG. 4. Pan-European multihydrological model and multidynamical forecast model skill as depicted by the seasonal average QS for four target seasons (DJF = winter, MAM = spring, JJA = summer, SON = autumn). The header indicates the lead and pan-European improvement in the average QS compared to the baseline ESP forecast (Fig. 3). Positive values indicate an increase in forecast skill, whereas negative values indicate a deterioration in the forecast skill compared to the ESP baseline.

We observe some remarkable forecast skill for the United Kingdom and Ireland, where we see an improvement in the forecast skill for most leads and seasons (Fig. 4). In general, regions that have a hydrology dominated by a high number of low-intensity precipitation events normally exhibit low seasonal forecast predictability. In our ensemble, most of the skill for these regions is obtained from the GCM forecasts of ECMWF-S4 and CanCM4 (not shown).

In general, we conclude that the dynamical models will only provide limited skill for some selected regions in the pan-European domain in the current forecasting experiment (Fig. A1). This is mostly caused by the high information content and resulting skill of the initial

conditions in combination with climatological meteorological forcing (Fig. A2). This makes the ESP, on the one hand, a powerful tool for seasonal forecast and, on the other hand, a very challenging benchmark to improve upon.

### c. Uncertainty contributions

The variability of the ensemble QS is used to determine the different sources of uncertainty. The uncertainty contributions of GCMs and HMs are computed separately to allow for uncertainty attribution. In general, the HM uncertainty dominates the total uncertainty in the forecasts (Fig. 5). This is also in line with the observed QS differences in the



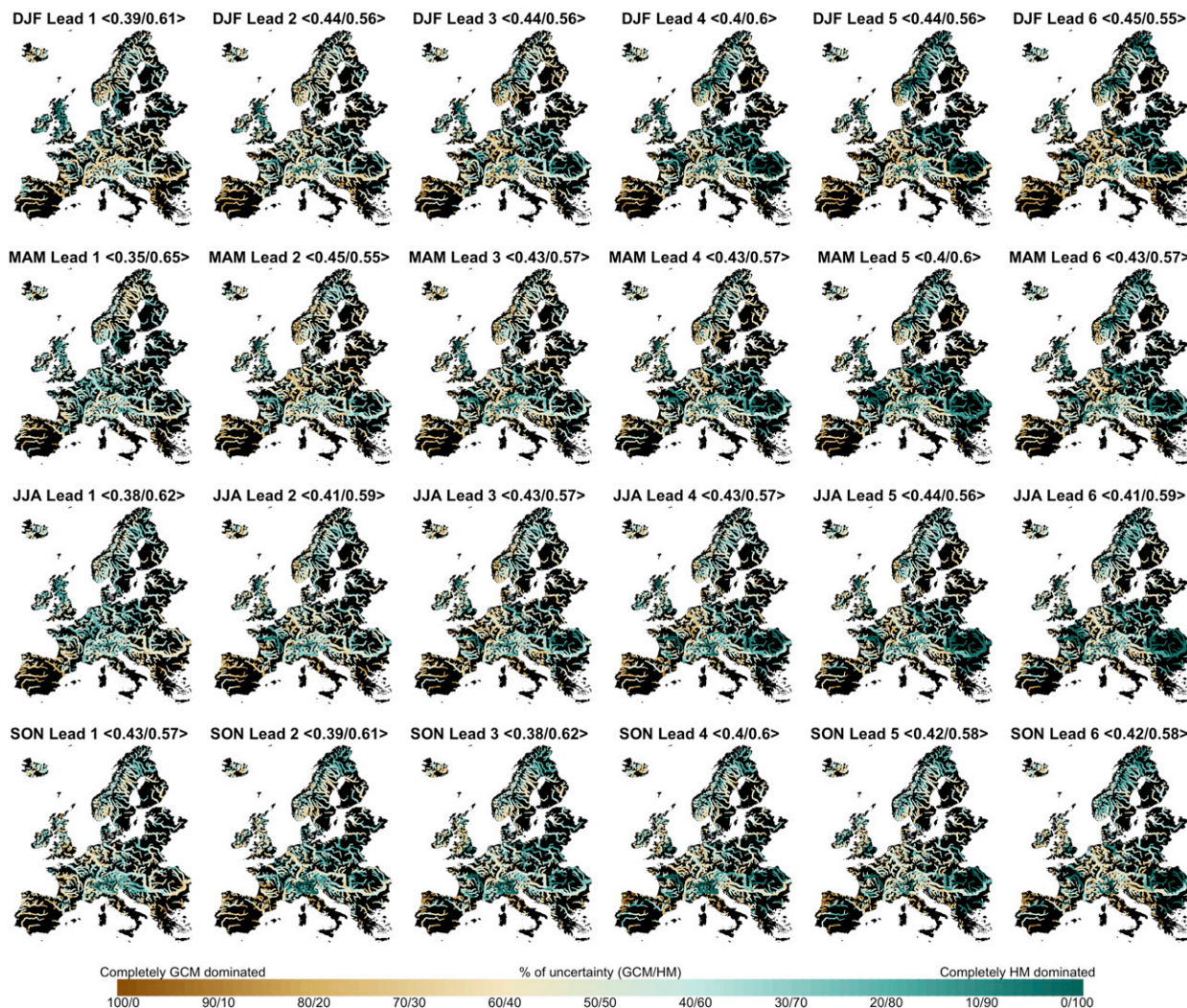


FIG. 5. Pan-European uncertainty contributions from HMs and GCM forecasts for four seasons. The header indicates the lead and pan-European ratio of GCM/HM uncertainty.

pan-European averages (Fig. 2), where the difference between the HMs is larger than for the GCMs. It shows that it is HM dependent whether a certain GCM will show a systematic higher performance than observed for the ESP (e.g., CanCM4 for PCR-GLOBWB and Noah-MP).

Eastern and northern Europe clearly show a domination of the HM uncertainty, whereas the Iberian Peninsula shows a stronger contribution of GCM uncertainty. This can be related to the performance of the dynamical forecasts of the MME compared to the ESP baseline. Areas that show an improvement in forecast skill from the use of GCMs are often linked to areas where GCM uncertainty will be the dominant source of the total uncertainty (Figs. 4, 5). Regions exhibiting a weak performance compared to the ESP baseline are

dominated by HM uncertainty. In these areas, it is likely that the HM will dominate the uncertainty since their ability to reproduce critical hydrological process, in combination with a large hydrological memory, results in a high forecast skill for the ESP.

*d. Impact of the initial conditions*

To study the seemingly strong impact of the initial conditions in PCR-GLOBWB, the rev-ESP was performed for this HM over Europe (Fig. 6). The initial conditions dominate up to 50% of the predictability up to a period of 3 months. The major part of the explained variance is explained by the initial groundwater conditions, whereas riverine storage (storage in lakes and rivers) only explains 30% of the variance in the first month. Soil moisture and snow conditions

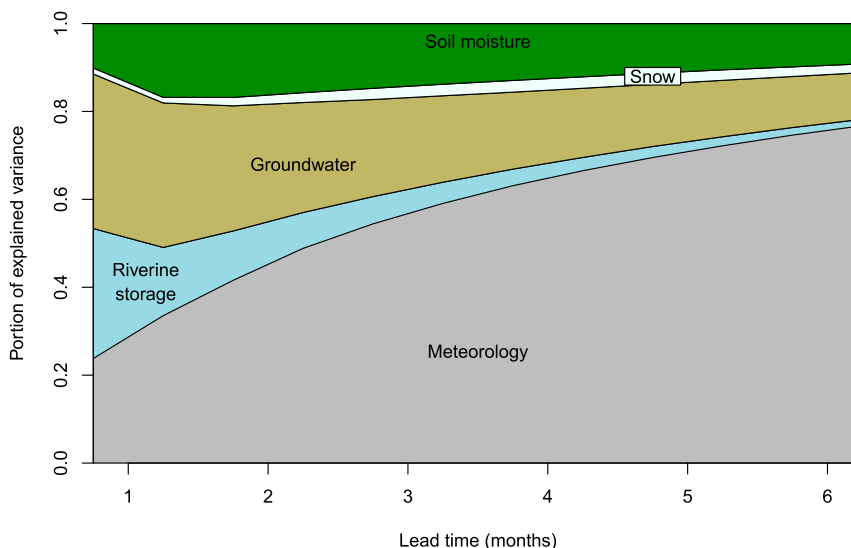


FIG. 6. Explained variance in hydrological forecasts for PCR-GLOBWB derived from a rev-ESP experiment. Values are average over the pan-European domain. The rev-ESP experiment was performed for the period 1993–2010.

account for the remaining 10%–20% of the variance, where snow dominates the signal in the northern parts of Europe.

The meteorological conditions only influence a small fraction of the total model forecast variance, which is important given that this will be the part of the results that leads to a significant difference between the ESP baseline and the dynamic MME forecasts. Even with perfect forecasts from the GCMs, we can only explain up to 25% of the total variance in the first month, which also shows why the relative improvement in most regions is only minor compared to the ESP baseline forecasts. Longer lead times show a higher potential if perfect forecasts are produced by the GCMs. However, skillful MME precipitation anomaly predictability for larger regions is generally limited to 2 months and temperature predictability up to 4–6 months, resulting in a limited potential for MME forecasts compared to an ESP baseline (Wanders and Wood 2016).

Even though the meteorological forecasts show a lower skill, a tendency is observed in all HMs to have better forecasts in the lower quintiles of the discharge forecasts (not shown). Improvements against the ESP baseline are most dominant in these low flows and below normal conditions, whereas the seasonal flood forecasts show a lack of skill compared to the ESP. It is clear that the impact of the initial conditions is highly important for the seasonal forecast skill (Fig. 6), not only for the dynamical GCM forecast but also for the ESP baseline, which is often used as a benchmark in seasonal forecasting experiments.

## 4. Discussion

### a. Impact of model structure

The results show a mixed performance of the dynamical GCM forecasts compared to the ESP baseline forecasts. We clearly observe that a low initial skill of the baseline forecast is a prerequisite for skill improvements by the dynamical model (Yuan et al. 2015). Model forecasts that are highly dependent on the initial conditions (e.g., PCR-GLOBWB, groundwater) also show a lower dependency on the meteorological input. This was also confirmed by Greuell et al. (2016); however, they indicated that soil moisture is a dominant driver for predictability. This already indicates that impact of the initial conditions is highly model specific. Both studies show that this narrows the chances for improvement by the dynamical forecast since the meteorological forcing only explains a small part of the total variance in the first months, when dynamical seasonal meteorological forecasts are most skillful (Fig. 6).

The impact of the initial conditions is interlinked with the model structure and the dominant processes in the model representation (Wanders et al. 2014), especially when one wants to forecast hydrological extremes (Mo and Lettenmaier 2014). Different model families have a different development history and therefore one would typically find different dominant processes in the different model families (Bierkens 2015). The Noah-MP comes from a background of land–atmosphere interaction modeling and the modeled processes are more focused on an accurate representation of the

land–atmosphere fluxes than streamflow generation (Yang et al. 2011). For example, Noah-MP has a surface runoff component that is only active during strong precipitation events. The explained variance might also be different for flood and drought events, as a result of this flashy runoff behavior. Hydrologic models such as mHM have two interflow components with different recession constants (fast and slow) that account for the rapid response of strong precipitation events. Resulting from this difference in the water partitioning, we hypothesize that Noah-MP is more sensitive to the meteorological forcings than the other GHMs, which will, in turn, explain a larger portion of the total variance. This ultimately results in lower ESP scores and a higher potential for the GCM forecasts to outperform the baseline forecast. Similar behavior to Noah-MP is observed for mHM and VIC, which also show a low skill in the baseline forecast (Fig. 2). PCR-GLOBWB, on the other hand, exhibits a higher dependency on the initial hydrologic conditions leading to a comparatively higher ESP skill than for the other models. This clearly has implications for the development of MME forecasting systems, where we at least in the seasonal meteorological forecasts observe a tendency to simply take the ensemble mean for the forecast (e.g., Kirtman et al. 2014). One way forward would be to use different weights in the construction of the MME mean, which has proven to provide more reliable forecasts in multimodel systems (Wanders and Wood 2016). From the analysis in this study we observe that for a balanced MME, one must preferably select models that originate from the different model families and have different process representations. This will ensure that the MME will capture the full range of uncertainty and will not be overconfident in its assessment of the uncertainty in the forecast (Samaniego et al. 2018). To ensure that all processes are represented and the uncertainty is not underestimated, we have selected four hydrological models from four different model families within the EDgE project. Only the inclusion of an infinitely large number of models, preferably from many different model families, will provide the true answer to that aim. However, with the limited resources available within EDgE, a selection of four hydrological models from an identical number of model families provides one of the largest hydrological ensembles to date used for seasonal forecasting.

### *b. Model calibration*

Within this study, only some minor calibrations were performed for the individual models [detailed descriptions in Thober et al. (2018) and Marx et al. (2018)]. Extensive model calibration has a clear advantage when it comes to the local representativeness of the seasonal forecasts

(Duan et al. 2007). However, Shi et al. (2008) show that for seasonal forecasting, calibration only provides marginal gains in terms of forecasting skill compared to bias correction methods. Another downside is that calibration often results in an unrealistic model representation for the locations where no calibration was performed or in conditions that were not observed during the calibration period. It is also important to realize that most objective functions in calibration procedures strongly favor medium and high flows (e.g., Nash–Sutcliffe efficiency, RMSE, bias). This could lead to unrealistic simulations of low-flow conditions, which is an important component when a seasonal hydrological forecasting system is used for decision support. Additionally, calibration against multiple objectives focusing on high and low flows could alleviate this problem. It is, however, important to derive a unique parameter set that allows for a physically realistic, seamless continental-scale simulation (Samaniego et al. 2017b). Recommendations for multisite parameter estimations can be found in Rakovec et al. (2016) and Samaniego et al. (2017b).

In addition to the physical impacts of the calibration procedures, the end users in the EDgE project indicated that they are more interested in forecast anomalies than in absolute values. Anomalies in the seasonal forecast can be related to their own forecast system, whereas absolute values are often biased and have no physical meaning for the end user. With that in mind, future calibration of seasonal forecasting systems should focus on the calibration of the anomalies rather than the absolute values.

### *c. Model validation*

For the skill evaluation within EDgE, we used the historical reference simulations (section 2c). Given the fact that we want to perform a spatial analysis of the model's skill to reproduce discharge anomalies, we used a historic simulation for reference. Clearly, this will have an impact on the skill of the ESP and the dynamical forecasts, since they are not compared to “real” observations. The use of the historic simulations will reduce the impact of existing biases and incorrect estimates of the dynamic behavior on the overall forecast performance. This might impact the findings with regard to the gain from using dynamical forecast models, because they can potentially remove existing model biases or correct errors in the dynamic range. To quantify that impact, further research is needed with locally calibrated hydrological models that are used over regions with sufficiently long discharge observations.

The findings from this study are in line with earlier studies over Europe that show limited improvement when using a dynamical forecast model compared to an ESP-based forecast (e.g., Arnal et al. 2018; Greuell et al. 2018). These experiments are based on a single



GCM–HM combination, but they show a similar tendency as observed in the EDgE project.

#### *d. Multimodel forecasting*

##### 1) UNCERTAINTY

This study aims to identify the advantages of a multimodel hydrological forecasting system compared to single model forecasts. Moreover, Fig. 3 demonstrates that the positive impact of dynamical forecast models is different for each hydrological model. For PCR-GLOBWB, for example, an ESP-based forecast would be useful for end users. For Noah-MP, it would be a CanCM4-based forecast, and ECMWF-S4 for VIC and mHM would provide the best seasonal forecast. This could change when future generations of GCMs become available, that work at finer spatial or temporal resolutions or include improved process descriptions. Now we observe that the GCM improvements are linked to the HM used, while one would expect that when GCMs become more and more skillful this relationship will no longer exist. In single model forecasts this degree of uncertainty cannot be captured, resulting in a likely underestimation of the uncertainty in the forecast results. Having the additional information on the uncertainty compared to other GCM–HM combinations will be very valuable and is an advantage of a multimodel over a single-model system. Even when a single-model system is highly calibrated and uses the best dynamical forecast model as forcing, this could still result in an underestimation of the uncertainty. Moreover, it is difficult to identify the weaknesses of a single GCM–HM system, whereas the multimodel has the advantage that the weaknesses are identified in the forecast evaluation (e.g., Fig. 3). Post-processing the discharge forecast could artificially inflate the forecast uncertainty; however, it will be difficult to capture behavior, processes, or uncertainties that are not captured by the hydrological model structure. Figure 5 clearly indicates that currently the hydrological models provide the majority of the uncertainty, which cannot be captured with a single-model forecast (Kumar et al. 2013).

##### 2) CONSISTENCY

A consistent set of input geophysical properties is applied across all four hydrological model parameterizations to ensure the consistency of model establishment and to some degree the model forecasts. The individual models use different transfer functions to go from the basic information on soil, geology, and vegetation to their model-specific parameters. The use of consistent physiographic information for the surface parameterization removes the potential for discrepancies in the forecast interpretation as a result of large differences in land surface

parameterization (Samaniego et al. 2017b). By aligning the model's parameterizations we remove one source of model uncertainty, which could otherwise lead to significant differences in model forecasts. The use of consistent hydrological model parameterization mainly benefits the consistency in simulations of the land surface fluxes. The potential root-zone soil moisture storage and groundwater storage can especially be strongly affected by minor differences in the parameterization, making it difficult for end users to compare the individual hydrological models in extreme hydrological conditions. To further reduce the discrepancies between the different hydrological models, we use a single mRM to route the grid-specific generated runoff through an upscaled and unique river network.

By using the multimodel setup, we have effectively added sources of uncertainty compared to more traditional systems that only use one hydrological model in combination with one GCM. The improved consistency in the hydrological models will reduce some of this additional ensemble spread in our probabilistic forecast, by using a consistent parameterization for the soils and land cover for different models.

##### 3) SPATIAL RESOLUTION

The authors are confident that with the growing computational power and the expansion of high-resolution modeling (Wood et al. 2011; Bierkens et al. 2015), future forecasting systems will be producing seasonal forecasts at finer spatial resolutions ranging from 1 to 10 km. This advancement will not only increase local representativeness, but it will also increase the value of the seasonal forecasts for end users. The additional information requires that users receive additional information on the value and reliability of these forecasts (Taylor et al. 2015). In the EDgE project, we achieve this goal by providing much needed expert knowledge on the skill of the seasonal forecast. This additional skill information can inform users on the reliability of the forecast of high-resolution multimodel forecasts.

##### 4) COMMUNICATION

Communicating the degree of confidence and uncertainty in multimodel seasonal forecasts is a difficult challenge in MME forecasts. Forecasts ideally include sufficient models, from different model families, to fully capture the uncertainty in different hydrological processes, but on the other hand should be able to make careful selections of reliable models to get the best forecast. Some models are known to perform poorly under certain hydrological conditions, which is subjective expert knowledge that is difficult to include in the uncertainty estimates. Therefore, it remains one of the key challenges to communicate MME forecast uncertainty to the end users. Within EDgE we have

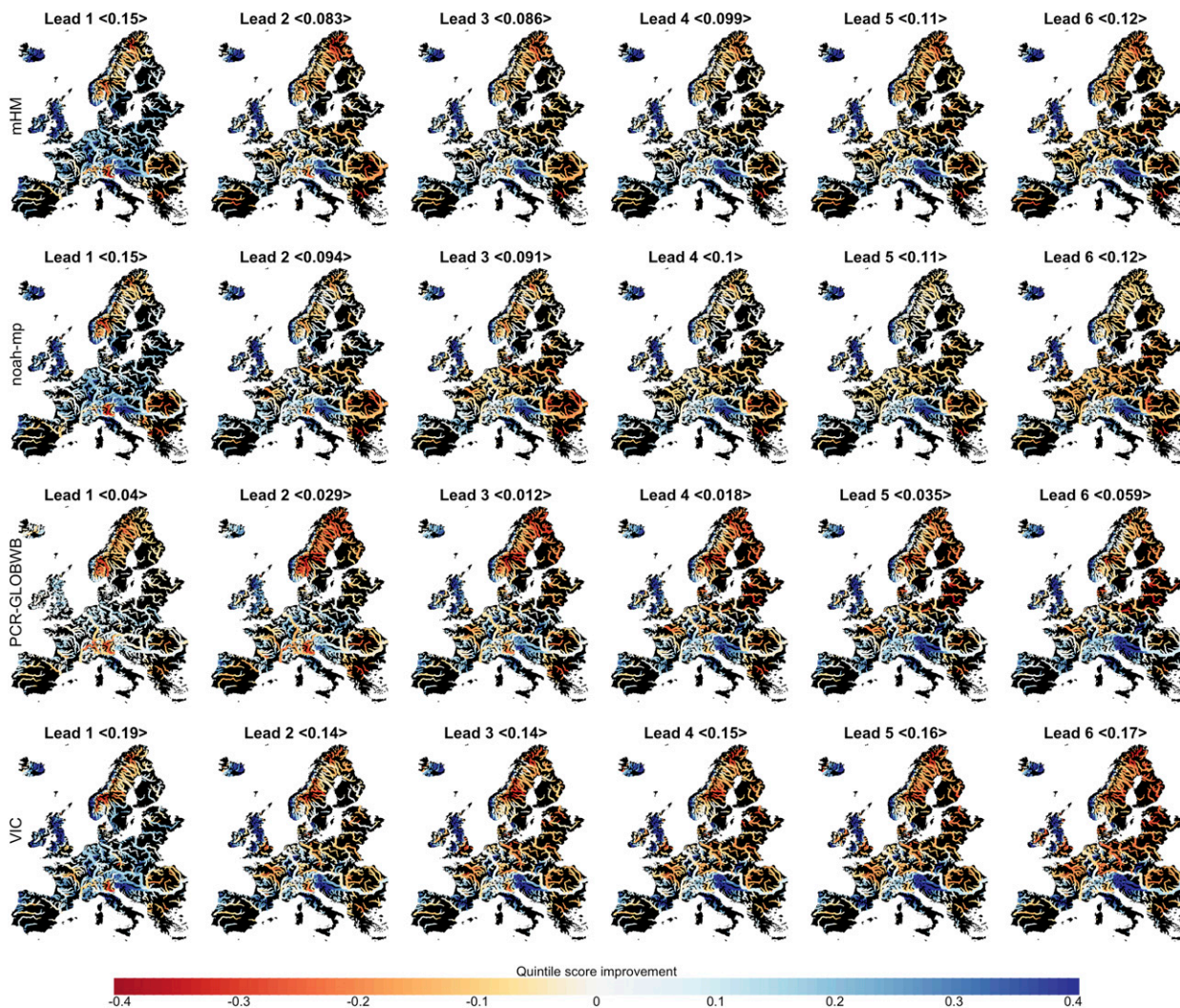


FIG. A1. Pan-European multidynamical forecast model skill as depicted by the QS separated for the four different hydrological models. The header indicates the lead and pan-European improvement in the average QS compared to the baseline ESP forecast (Fig. 3). The QSS are averaged over all months of the year. Positive values indicate an increase in forecast skill, whereas negative values indicate a deterioration in the forecast skill compared to the ESP baseline.

communicated the expert assessment of the uncertainty using a traffic light, which indicates the level of confidence of the forecast, based on expert knowledge and performance metric (<http://edge.climate.copernicus.eu/Apps/#seasonal>). Green indicates a high reliability of the forecast, orange indicates a fair forecast, and a poor forecast is indicated by a red color. This way of communicating the forecast uncertainty has been positively received by our forecast users within the EDgE project.

### 5. Conclusions

This study presents the first pan-European, high-resolution, multimodel seasonal hydrological forecasting

system, using an ensemble of dynamical meteorological forecasts, a climatological benchmark forecast, and distinctly behaving hydrologic/land surface models. We show that the benchmark climatological forecasts have a considerable skill for the first 2 or 3 months, after which the skills reduce. The skill in the benchmark forecasts is completely driven by the initial hydrological conditions, which determine the forecast skill of the hydrological models using climatological meteorological forecasts. Large differences between the benchmark skills of the hydrological models is observed, which is caused by the model’s dependency on the initial hydrological conditions.

We show that the dynamical model forecasts outperform the climatological benchmark for large parts of



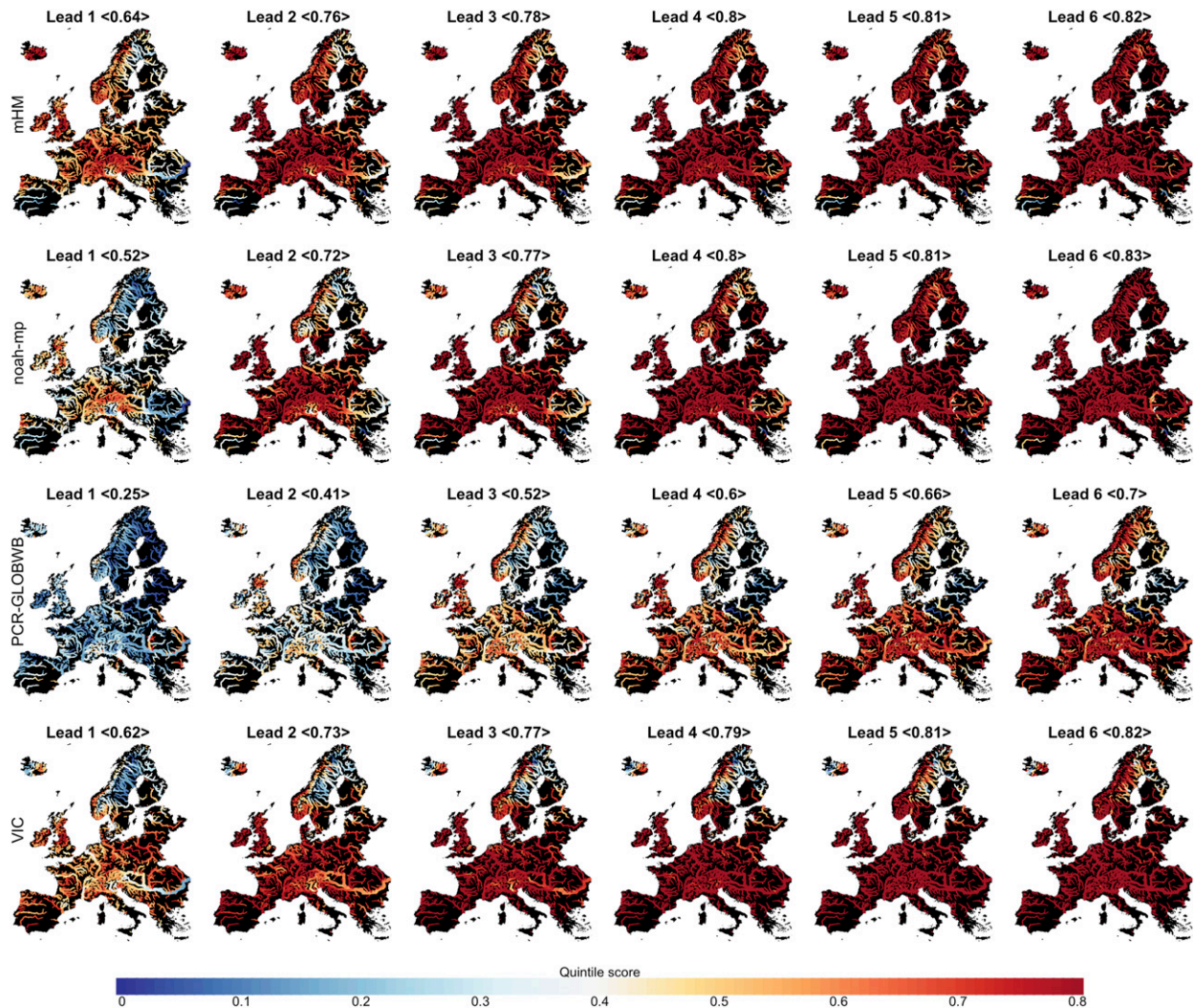


FIG. A2. Pan-European ESP forecast skill as depicted by the QS for the four different hydrological models. The header indicates the lead and pan-European average QS. The QSs are averaged over all months of the year. Using the quintile formulation and the quintile forecast produced in EDgE, QSs of 0.8 indicate no skill, and a value of 0.0 indicates perfect forecasts.

Europe. The improvement in forecast skill is highly dependent on the initial skill of the benchmark forecast. The largest improvements are found for western and central Europe, while a degradation in the forecast skill is found for Scandinavia as a result of the impact of snow.

A large part of the forecast uncertainty comes from the hydrological models that make up 55%–60% of the total uncertainty in the dynamical forecasts. The spatial patterns in hydrological model uncertainty change with increasing lead times, however, the European spatial-average contribution remains stable.

The strong contribution of hydrological models to the forecast uncertainty shows the added value of multimodel seasonal hydrological forecast systems over single-model

forecasts. The large differences between the models indicate that the use of a single model could significantly overestimate the skill and underestimate the uncertainty of the hydrological forecasts.

The fine spatial resolution, combined with the large model ensemble used within EDgE, provides a unique support system for forecast decision-making. The detailed information on the uncertainty (contributions) and the skill compared to climatological benchmarks informs end users on the added value of the system, compared to the benchmark systems.

This work shows the potential benefits of multimodel forecasting operational systems for operational seasonal forecasts. This will come at some considerable computational cost, but shows improved understanding and



information obtained from seasonal hydrological forecasts in this study. In addition, we suggest other seasonal forecast systems should use dynamical meteorological forecasts in combination with a climatological benchmark to have an accurate estimate of the skill and uncertainty compared to the climatological default forecasts.

**Acknowledgments.** This study has been funded by the Copernicus Climate Change Service. The European Centre for Medium Range Weather Forecasts implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. We thank all the colleagues who contributed to the EDgE project (<http://edge.climate.copernicus.eu/>). N.W. acknowledges the funding from NWO Rubicon 825.15.003 and NWO 016.Veni.181.049. E.W. was supported through Grant NA150AR4310075 (Assessing Phase 2 NMME). We acknowledge the E-OBS dataset from the EU FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>).

Author contribution: N.W., S.T., M.P., and R.K. produced the seasonal forecast and reference simulations for the four different hydrological models. S.T. produced the downscaled seasonal meteorological forecast data and provided the code for the stand-alone routing model. R.K. and L.S. provided the baseline information for the hydrological model parameterization. N.W. and E.W. designed the skill and uncertainty methodology, and N.W. and M.P. performed the skill and uncertainty assessment for all hydrological models. N.W. and S.T. took the lead in the preparation of the manuscript, and all authors contributed to the further writing of the manuscript.

Code and data availability: Large parts of the code developed in the EDgE project are publicly available. All forecast data are available at <http://edge.climate.copernicus.eu/Apps/#seasonal>. The codes for the hydrological models are available from: PCR-GLOBWB ([https://github.com/UU-Hydro/PCR-GLOBWB\\_model](https://github.com/UU-Hydro/PCR-GLOBWB_model)), VIC (<https://vic.readthedocs.io/en/master/>), mHM (<http://www.ufz.de/index.php?en=40114>), Noah-MP (<https://ral.ucar.edu/solutions/products/noah-multiparameterization-land-surface-model-noah-mp-lsm>). The reference E-OBS meteorological forcing can be obtained from <https://www.ecad.eu/download/ensembles/download.php>. The dynamical seasonal forecast from the CanCM4 and FLOR can be obtained from <https://www.earthsystemgrid.org/>. The ECMWF-S4 and LFPW forecast are only available through ECMWF (<https://www.ecmwf.int/>). Finally, all forecast and indicators can be directly downloaded from <http://edge.climate.copernicus.eu/>.

## APPENDIX

### Detailed Forecast Skill Information

Dynamical model forecasts are shown in Fig. A1, and the ensemble streamflow prediction baseline is shown in Fig. A2.

### REFERENCES

- Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS—Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, <https://doi.org/10.5194/hess-17-1161-2013>.
- Arnal, L., H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger, 2018: Skillful seasonal forecasts of streamflow over Europe? *Hydrol. Earth Syst. Sci.*, **22**, 2057–2072, <https://doi.org/10.5194/hess-22-2057-2018>.
- Bierkens, M. F. P., 2015: Global hydrology 2015: State, trends, and directions. *Water Resour. Res.*, **51**, 4923–4947, <https://doi.org/10.1002/2015WR017173>.
- , and Coauthors, 2015: Hyper-resolution global hydrological modelling: What is next? *Hydrol. Processes*, **29**, 310–320, <https://doi.org/10.1002/hyp.10391>.
- Bohn, T. J., B. Livneh, J. W. Oyster, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteorol.*, **176**, 38–49, <https://doi.org/10.1016/j.agrformet.2013.03.003>.
- Brier, G., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Chaney, N. W., P. Metcalfe, and E. F. Wood, 2016: HydroBlocks: A field-scale resolving land surface model for application over continental extents. *Hydrol. Processes*, **30**, 3543–3559, <https://doi.org/10.1002/hyp.10891>.
- Clark, M. P., and Coauthors, 2015: A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour. Res.*, **51**, 2498–2514, <https://doi.org/10.1002/2015WR017198>.
- Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian, 2007: Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.*, **30**, 1371–1386, <https://doi.org/10.1016/j.advwatres.2006.11.014>.
- Greuell, W., W. H. P. Franssen, and R. W. A. Hutjes, 2016: Seasonal streamflow forecasts for Europe – Part II. Explanation of the skill. *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2016-604>.
- , —, H. Biemans, and R. W. A. Hutjes, 2018: Seasonal streamflow forecasts for Europe – Part I: Hindcast verification with pseudo- and real observations. *Hydrol. Earth Syst. Sci.*, **22**, 3453–3472, <https://doi.org/10.5194/hess-22-3453-2018>.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, <https://doi.org/10.1029/2008JD010201>.
- Kirtman, B. P., and Coauthors, 2014: The North American Multi-model Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull.*

- Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kumar, R., B. Livneh, and L. Samaniego, 2013: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme. *Water Resour. Res.*, **49**, 5700–5714, <https://doi.org/10.1002/wrcr.20431>.
- Kundzewicz, Z. W., and Z. Kaczmarek, 2000: Coping with hydrological extremes. *Water Int.*, **25**, 66–75, <https://doi.org/10.1080/02508060008686798>.
- Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global Planet. Change*, **13**, 195–206, [https://doi.org/10.1016/0921-8181\(95\)00046-1](https://doi.org/10.1016/0921-8181(95)00046-1).
- Marx, A., and Coauthors, 2018: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3c. *Hydrol. Earth Syst. Sci.*, **22**, 1017–1032, <https://doi.org/10.5194/hess-22-1017-2018>.
- Merryfield, W. J., and Coauthors, 2013: The Canadian Seasonal to Interannual Prediction System. Part I: Models and initialization. *Mon. Wea. Rev.*, **141**, 2910–2945, <https://doi.org/10.1175/MWR-D-12-00216.1>.
- Météo-France, 2015: Météo-France seasonal forecast system 5 for Eurosp. Tech. Doc., 38 pp., <https://www.umr-cnrm.fr/IMG/pdf/system5-technical.pdf>.
- Mizukami, N., M. P. Clark, A. J. Newman, A. W. Wood, E. D. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego, 2017: Towards seamless large-domain parameter estimation for hydrologic models. *Water Resour. Res.*, **53**, 8020–8040, <https://doi.org/10.1002/2017WR020401>.
- Mo, K. C., and D. P. Lettenmaier, 2014: Objective drought classification using multiple land surface models. *J. Hydrometeorol.*, **15**, 990–1010, <https://doi.org/10.1175/JHM-D-13-071.1>.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (System 4). ECMWF Tech. Memo. 656, 49 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2011/11209-new-ecmwf-seasonal-forecast-system-system-4.pdf>.
- Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, <https://doi.org/10.1029/2010JD015139>.
- Rakovec, O., R. Kumar, S. Attinger, and L. Samaniego, 2016: Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resour. Res.*, **52**, 7779–7792, <https://doi.org/10.1002/2016WR019430>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Samaniego, L., R. Kumar, and S. Attinger, 2010: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.*, **46**, W05523, <https://doi.org/10.1029/2008WR007327>.
- , and Coauthors, 2017a: Mesoscale hydrologic model, version 5.8. Zenodo, <https://doi.org/10.5281/zenodo.1069203>.
- , and Coauthors, 2017b: Toward seamless hydrologic predictions across spatial scales. *Hydrol. Earth Syst. Sci.*, **21**, 4323–4346, <https://doi.org/10.5194/hess-21-4323-2017>.
- , and Coauthors, 2017c: 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**, 435–449, <https://doi.org/10.1007/s10584-016-1778-y>.
- , and Coauthors, 2018: Anthropogenic warming exacerbates European soil moisture droughts. *Nat. Climate Change*, **8**, 421–426, <https://doi.org/10.1038/s41558-018-0138-5>.
- Sheffield, J., and Coauthors, 2014: A drought monitoring and forecasting system for sub-Saharan African water resources and food security. *Bull. Amer. Meteor. Soc.*, **95**, 861–882, <https://doi.org/10.1175/BAMS-D-12-00124.1>.
- Shi, X., A. W. Wood, and D. P. Lettenmaier, 2008: How essential is hydrologic model calibration to seasonal streamflow forecasting? *J. Hydrometeorol.*, **9**, 1350–1363, <https://doi.org/10.1175/2008JHM1001.1>.
- Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier, 2013: On the sources of global land surface hydrologic predictability. *Hydrol. Earth Syst. Sci.*, **17**, 2781–2796, <https://doi.org/10.5194/hess-17-2781-2013>.
- Sutanudjaja, E. H., and Coauthors, 2018: PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geosci. Model Dev.*, **11**, 2429–2453, <https://doi.org/10.5194/gmd-11-2429-2018>.
- Taylor, A. L., S. Dessai, and W. B. de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc. London*, **373A**, 20140454, <https://doi.org/10.1098/rsta.2014.0454>.
- Thielen, J., J. Bartholmes, M. Ramos, and A. de Roo, 2009: The European Flood Alert System – Part 1: Concept and development. *Hydrol. Earth Syst. Sci.*, **13**, 125–140, <https://doi.org/10.5194/hess-13-125-2009>.
- Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego, 2015: Seasonal soil moisture drought prediction over Europe using the North American Multi-Model Ensemble (NMME). *J. Hydrometeorol.*, **16**, 2329–2344, <https://doi.org/10.1175/JHM-D-15-0053.1>.
- , and Coauthors, 2018: Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming. *Environ. Res. Lett.*, **13**, 014003, <https://doi.org/10.1088/1748-9326/aa9e35>.
- Todini, E., 2007: A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. *Hydrol. Earth Syst. Sci.*, **11**, 1645–1659, <https://doi.org/10.5194/hess-11-1645-2007>.
- Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, <https://doi.org/10.1175/JCLI-D-14-00158.1>.
- Wanders, N., and E. F. Wood, 2016: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environ. Res. Lett.*, **11**, 094007, <https://doi.org/10.1088/1748-9326/11/9/094007>.
- , D. Karssenberg, A. de Roo, S. M. de Jong, and M. F. P. Bierkens, 2014: The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol. Earth Syst. Sci.*, **18**, 2343–2357, <https://doi.org/10.5194/hess-18-2343-2014>.
- Wood, A. W., and D. P. Lettenmaier, 2008: An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophys. Res. Lett.*, **35**, L14401, <https://doi.org/10.1029/2008GL034648>.
- , A. Kumar, and D. P. Lettenmaier, 2005: A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States. *J. Geophys. Res.*, **110**, D04105, <https://doi.org/10.1029/2004JD004508>.
- Wood, E. F., and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring

- earth's terrestrial water. *Water Resour. Res.*, **47**, W05301, <https://doi.org/10.1029/2010WR010090>.
- Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.*, **117**, D03110, <https://doi.org/10.1029/2011JD016051>.
- Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.*, **116**, D12110, <https://doi.org/10.1029/2010JD015140>.
- Yuan, X., E. F. Wood, and Z. Ma, 2015: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *Wiley Interdiscip. Rev.: Water*, **2**, 523–536, <https://doi.org/10.1002/wat2.1088>.