HYDROLOGICAL FORECASTS AND PROJECTIONS FOR IMPROVED DECISION-MAKING IN THE WATER SECTOR IN EUROPE

Luis Samaniego, Stephan Thober, Niko Wanders, Ming Pan, Oldrich Rakovec, Justin Sheffield, Eric F. Wood, Christel Prudhomme, Gwyn Rees, Helen Houghton-Carr, Matthew Fry, Katie Smith, Glenn Watts, Hege Hisdal, Teodoro Estrela, Carlo Buontempo, Andreas Marx, and Rohini Kumar

A high-resolution multimodel ensemble of state-of-the-art climate and hydrological models has been developed to deliver hydrometeorological change metrics codesigned with key water sector stakeholders in Europe.

Existing water-oriented decision-support systems are either designed as early warning systems to provide forecasts of hydrological floods and droughts, or as monitoring platforms aiming to provide information on the current state of variables of interest such as streamflow or soil moisture. Such systems normally target national, continental, and global scales with examples of systems developed for Australia (Emerton et al. 2016), Africa (Sheffield et al. 2014), Europe [European Flood Awareness System (EFAS); Thielen et al. 2009], and North America (Demargne et al. 2014).

While the multiplication of water-focused climate services undeniably helps downstream

decision-making, a number of areas of improvements can be identified. First, only few systems operate at different prediction horizons, most focusing on a single function targeting monitoring, short- to medium-range forecasting, and seasonal or climate time scales. This means that users need to refer to different systems and services depending on their planning scale, each generally associated with different and inconsistent types of information and delivered services, hence requiring users to develop different application tools for each independent one, and to take much care when interpreting their different outcomes, as they might not provide exactly the same information. Two notable exceptions are the European Flood Awareness System (Thielen et al. 2009; Arnal et al. 2018) and the Global Flood Awareness System (Alfieri et al. 2013), both components of the Copernicus Emergency Management Service, which are currently the only operational forecasting systems providing ensemble streamflow forecasting and flood early warning at both medium range and seasonal time frames for Europe and the world, respectively; however, none currently offer predictions beyond a few months, limiting their use to short- and medium-term planning.

Second, existing continental or global-scale systems [e.g., for Africa (Sheffield et al. 2014), North America (Lawrimore et al. 2002), or Europe (Horion et al. 2012)] typically operate at resolutions of 0.25° or coarser (note the European Drought Observatory has a multiscale approach) and do not quantify the uncertainty associated with the monitoring and forecasting hydrometeorological chain. At national scale, monitoring systems are typically available at a high spatial resolution [e.g., the German Drought Monitor at 4-km scale (www.ufz.de/droughtmonitor; Zink et al. 2016) or the U.K. Drought portal at 5-km scale (https://eip.ceh.ac.uk/apps/droughts/)] and provide timely information for decision-making and the general public, while high-resolution national forecasting services are normally based on a single hydrological model, not capturing some of the important uncertainty. This limits robust local decision-making as single deterministic estimates are often given instead of probabilistic ones.

Third, most climate service portals, especially when designed for climate projections, only focus on a few specific climate-related variables. For example, the Royal Netherlands Meteorological Institute Climate Explorer (climexp.knmi.nl/) is a web application to visualize and analyze global climate data, but does not include hydrological-derived indices relevant for the water sector (see, e.g., National Research Council 2001). This can limit downstream applications because the information relevant for local decision-making is absent and would require further processing (and associated resources) by users, thus, hampering potential uptake.

The project "End-to-end Demonstrator for improved decision-making in the water sector in Europe" (EDgE; https://youtu.be/PqoRi6eSM2w) was designed to address three specific gaps in existing climate services for a more user-focused delivery for the water sector: 1) development of a high-resolution (5 km), multimodel system [using two land surface models (LSMs) and two hydrological models (HM) established with a common set of land surface properties across continental Europe] where both uncertainty in the atmospheric forcing and hydrological impact are accounted for and summarized to users; 2) delivery of a consistent and comparable multiscale sets of seasonal forecasts and climate impact projections based on the same hydroclimate modeling chain; and 3) provision of 36 climate and water indicators [so-called sectoral climate impact indicators (SCIIs)] and a unique web information service codesigned with over 30 different public- and private-sector stakeholders from different hydroclimatic regions and water-related industries across Europe, to facilitate the uptake of the service. The rationale of EDgE is simple: better-informed operational and strategic planning decisions can be made only with a timely, coherent, and codesigned water-oriented information system (Lourenço et al. 2015). EDgE was one of two proof of concepts (PoCs) for the water sector for the European Copernicus Climate Change Service Sectoral Information Systems program.

In this paper, we present briefly the technical features of the EDgE modeling chain, discuss the continental-scale application of the underlying model components, including the forcing data, and discuss the codesign and interpretation of impact indicators. Although the stakeholder feedback process and the development of visualization tools were fundamental parts of the PoC, they are not covered here. For completeness, note that EDgE was a proof-of-concept project and does not provide a real-time service; however, all indicators estimated under both seasonal forecast and climate impact prediction mode are accessible online (http://edge.climate.copernicus.eu).

CLIMATE DATA PROCESSING. Daily temperature and precipitation from the E-OBS 25-km gridded product (v12) and the underlying station data (Haylock et al. 2008) were used as historical meteorological forcing data for the period from 1950 to 2015. In addition, gridded daily wind speed was

obtained from the European Flood Alert System forcing (Thielen et al. 2009), made available by the European Centre for Medium-Range Weather Forecasts (ECMWF) for the period 1990–2014. Gridded historical observations were used to develop the hydrological model historical simulations, to bias correct the future climate prediction forcings and to generate the hydrological initial conditions for the seasonal hindcasts (i.e., forecasts of past periods).

For the climate change simulations, daily temperature and precipitation from five bias-corrected global climate models (GCMs) from the Coupled Model Intercomparison Project, phase 5 (CMIP5; HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M, and NorESM1-M), were used to drive the HM/LSMs during the period from 1950 to 2099 under two representative concentration pathways [RCPs; RCP2.6 and RCP8.5; see climate projections (CP) mode in Fig. 1]. This dataset was made available by the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP; Warszawski et al. 2014) at a spatial resolution of 0.5° and was selected as it benefited already from a trend-preserving bias correction (Hempel et al. 2013; see further detail in appendix A).

For the seasonal forecast simulations, daily temperature and precipitation hindcasts from four GCMs run in seasonal forecast mode (SF-GCM) were used to drive the HM/LSMs [see seasonal forecast (SF) mode in Fig. 1]. These comprise two models from the North American Multi-Model Ensemble [NMME; Canadian

Climate Model, version 4 (CanCM4)and the Geophysical Fluid Dynamics Laboratory Forecast-Oriented Low Ocean Resolution model (GFDL-FLOR)] and two European models from the Copernicus Climate Change Service [C3S; ECMWF System 4 (ECMWF-S4) and the Météo-France modeling system, version 5 (code name LFPW)]. The number of NMME SF-GCMs were selected to counterweight those provided by the ECMWF. The number of realizations among SF-GCMs were chosen so that each member has a similar weight in the multimodel ensemble. In total, the EDgE SF-GCM multimodel ensemble contained 52 realizations composed of 10 members from CanCM4, 12 members from GFDL-FLOR, and 15 members each from ECMWF-S4 and LFPW. Daily hindcasts starting on the first day of each month within the hindcast period 1993–2011 were used. The SF-GCMs forcings were downscaled from their native spatial resolution (of 1° for NMME models and 0.75° for C3S models) to the hydrological model resolution (see section below) without previous drift or bias correction, the relatively short hindcast period (1993–2011) being considered not sufficiently long to train a robust bias correction algorithm.

The spatial resolution of all forcing data were considered too coarse for deriving water sector indicators relevant for water managers and practitioners at local and regional levels in Europe. Instead, a spatial resolution of 5×5 km^2 was selected to derive all EDgE products as a trade-off between a spatial resolution
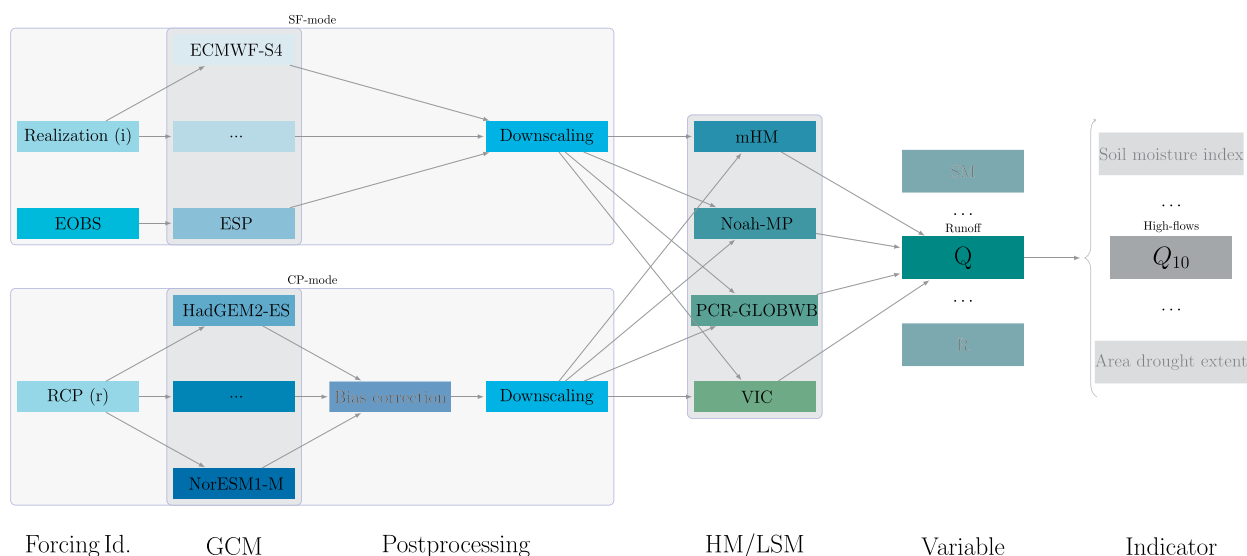
FIG. 1. EDgE modeling chain for seasonal forecast (SF) and climate prediction (CP) modes, shown here for the high-flow indicator Q_{10}. Both chains use four hydrological models to compute values for the terrestrial Essential Climate Variables (tECVs), which are the basis to estimate the sectorial climate impact indicators (SCIIs) requested by focus groups. For the SF mode the number of climate realizations can vary across models ($i = 10$–15). For the CP mode, the representative climate pathway (RCP) are set to RCP2.6 and RCP8.5 ($r = 1,2$). ESP is included as a benchmark for the dynamic SF models.

that is informative for practitioners across Europe and a scale at which it is still feasible to estimate water-related variables using current computational facilities and geophysical forcing information. The forcing data were hence downscaled from their native resolution to the common 5×5 km^2 resolution prior to their use as input to the hydrological modeling chain. Details of the statistical downscaling technique is provided in appendix A.

HYDROLOGICAL MODELING. Hydrological and land surface models (HM/LSMs) are the backbone of the EDgE modeling chain (Fig. 1); they comprise the mesoscale Hydrological Model (mHM); the PCRaster Global Water Balance model 2 (PCR-GLOBWB); the Variable Infiltration Capacity model (VIC); and the Noah land surface model with multiparameterization options (Noah-MP). All four models are process based, simulating canopy interception, snow accumulation and melting, infiltration, evapotranspiration, and runoff generation. They were selected based on the diversity of their underlying process representations and their widespread use in hydrological applications to capture as much as possible the structural uncertainty within the hydrological modeling component. A summary of the land surface data used to parameterize these models up can be found in the appendix B.

The mHM (Samaniego et al. 2010; Kumar et al. 2013b) is a grid-based distributed hydrological model equipped with a multiscale parameter regionalization scheme, developed with a special focus on running seamlessly at multiple spatial resolutions ranging from 1 to 50 km (Kumar et al. 2013a; Rakovec et al. 2016; Samaniego et al. 2017b), and ready to be implemented in an operational setting (Kauffeldt et al. 2016). PCR-GLOBWB (van Beek et al. 2011; Wanders and Wada 2015; Sutanudjaja et al. 2018) is a grid-based hydrological and water resources model, developed to represent the terrestrial water cycle at global and continental scales, with a special emphasis on including human water uses. VIC (Liang et al. 1994; Cherkauer et al. 2003) is a macroscale hydrological model that solves full water and energy balances to represent the land surface hydrology and near-surface atmospheric fluxes. The VIC model has been implemented in catchment to global-scale applications for understanding catchment behavior, extreme hydrological events, hydrological predictability, and climate change impacts (a.o., Sheffield and Wood 2008; Sheffield et al. 2014; Yuan et al. 2015). The Noah-MP model provides several upgrades of the Noah LSM (Niu et al. 2011), which was originally developed as the land surface scheme for numerical weather prediction (Ek et al. 2003). In this study, we use the same process parameterizations as in Cuntz et al. (2016).

Within EDgE, all four HMs/LSMs were established using the same high-resolution (500 m) morphologic, land-cover, and soil databases (appendix B). Differences between models originate only from different process representations. All HMs/LSMs are set up at a spatial resolution of 5 km, simulate daily water fluxes and states, and were calibrated with standard procedures described in appendix C. These models were subsequently evaluated in hundreds of river basins across Europe that cover a wide range of hydroclimatological regimes. For more details, refer to Fig. A1.

Another hallmark of the hydrological modeling chain is the use of a common river routing scheme to minimize predictive uncertainty from inconsistencies in the channel network. The gridded runoff fields generated by the four HMs/LSMs are routed through the same 5-km river network using the multiscale Routing Model (mRM; Thober et al. 2019) that was originally developed for mHM (Samaniego et al. 2010). The mRM has the ability to simultaneously route cell-generated runoff to multiple outlets, allowing streamflow to be generated over the entire domain simultaneously (Fig. 2). The key characteristic of mRM is its capacity to estimate streamflow at various spatial resolutions without recalibration of river-routing parameters within the employed Muskingum scheme. This simplification of the Saint-Venant equations is justified in all hydrological models used in this project because it accounts for wave advection and attenuation, which are the governing fluvial processes in the study area [see details in Thober et al. (2019) and references therein].

SCII: STAKEHOLDER CODESIGN AND ESTIMATION. The general graphical representation of the hydrometeorological modeling chains used to generate terrestrial essential climate variables (tECVs; Sessa and Dolman 2008) for climate predictions and seasonal forecasts is shown in Fig. 1. Four tECVs (Table 1) were stored from each simulation for the historical, climate prediction and seasonal forecast modes. These are streamflow Q, soil moisture in the top 2 m (SM), snow water equivalent (SWE), and groundwater recharge R.

In the CP chain, all four HMs/LSMs were forced with the downscaled GCM data for the period from 1950 to 2099 under RCP2.6 and RCP8.5. The period 1971–2000 is selected to represent present-day conditions because it is often used within the IPCC SR1.5 (Hoeg-Guldberg et al. 2019). The present-day
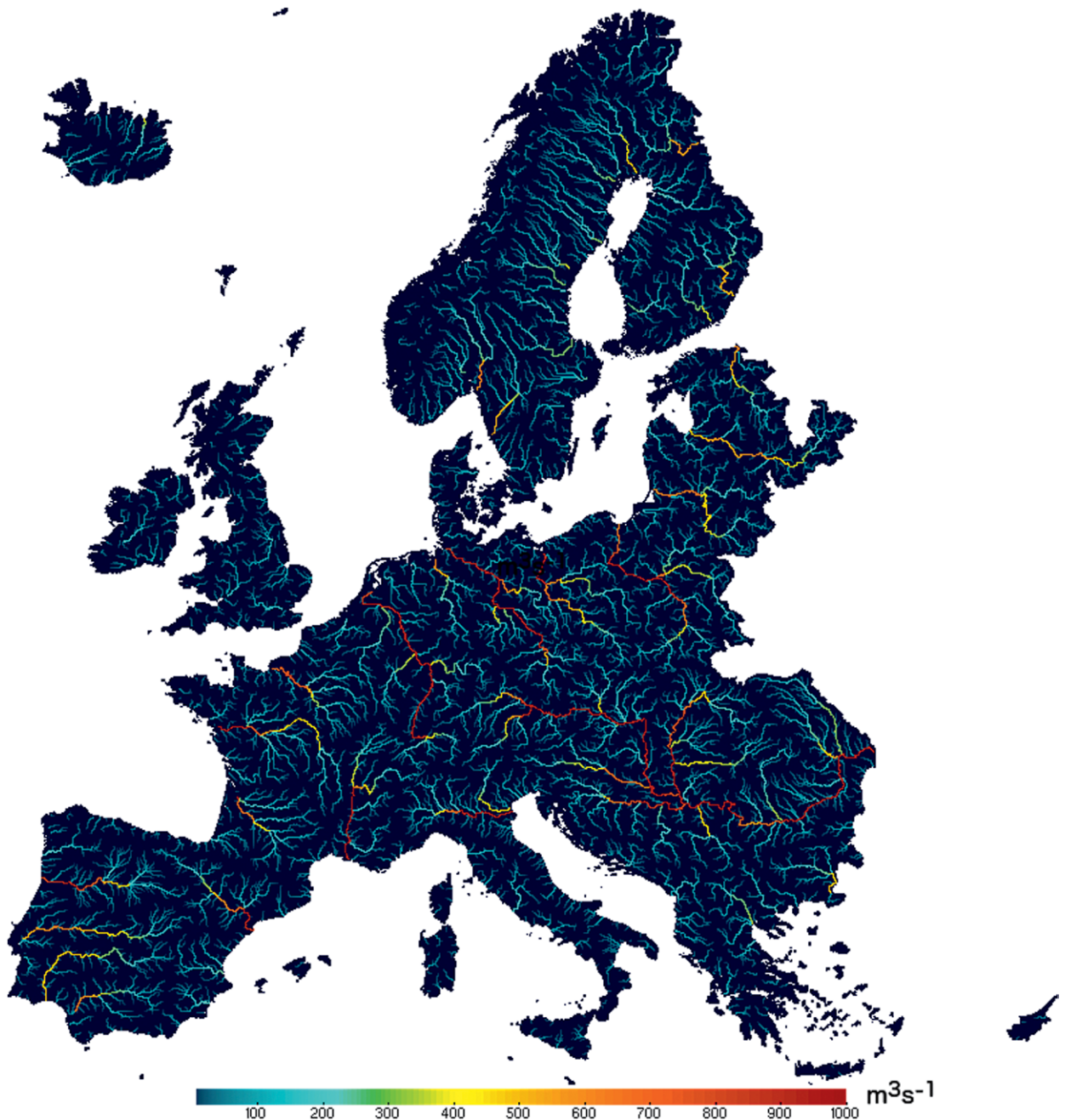
Fig. 2. Simulated mean daily streamflow from 1950 to 2011 for the EDgE domain at 5-km spatial resolution. Simulated daily runoff was obtained with the mHM model forced with downscaled E-OBS forcing data. River routing was carried out with the mRM algorithm.

conditions are based on the historical HMs/LSMs simulations driven by the GCM datasets. The workflow of the operationalization of the CP chain includes the following steps (Fig. 3a): (i) obtain GCM projections, (ii) perform bias correction (in this case, this step was not necessary because the ISI-MIP forcing data are already bias corrected) and downscaling, (iii) run the multihydrometeorological modeling chain to generate an ensemble of target variables (tECVs), (iv) generate an ensemble of indicators (SCIIs) based on the tECVs, (v) estimate uncertainty in the ensemble, and (vi) export ensemble outputs/indicators for visualization. This workflow also includes occasional reconfigurations, such as recalibration of the HMs/LSMs, modification of the SCIIs according to user needs, and updates of the web service.

In the SF chain, all four HMs/LSMs were forced monthly with the downscaled SF-GCM 7-month

TABLE 1. Sectoral climate impact indicators (SCIIs) derived from terrestrial essential climate variables (tECVs): streamflow Q ($m^3 s^{-1}$), top-2-m soil moisture as fraction of saturation (SM; $m m^{-1}$), groundwater recharge R ($mm day^{-1}$), snow water equivalent (SWE) (m), and meteorological forcing data: potential evapotranspiration (PET) ($mm day^{-1}$), precipitation P ($mm day^{-1}$), daily average temperature T (°C). The letter X denotes any of the tECVs, and X_p denotes the value of X that is equaled or exceeded p% of the time over a time horizon. Lead time is denoted by ℓ.

Index type	Statistic	Time	tECV	Notes
Climate predictions				
Relative change	Daily X_{10}	30 years	Q, R	Relative change in high values X_{10} with respect to (w.r.t) the reference period
Relative change	Median annual X_{max}	30 years	Q	Peak values w.r.t the reference period
Relative change	Daily X_{90}	30 years	Q, R	Low values w.r.t the reference period
Relative change	Daily X_{95}	30 years	Q, R	w.r.t the reference period
Relative change	Monthly mean X	1 month	Q, R, PET, P, SWE, T	w.r.t the reference period; for each calendar month
Relative change	Seasonal-mean X	3 months	Q, R, PET, P, SWE, T	w.r.t the reference period; seasons considered: DJF, MAM, JJA, SON
Relative change	Annual-mean X	1 year	Q, R, PET, P, SWE, T	w.r.t the reference period
Percentile index	Monthly $F(X)$	Monthly	SM, R	$F(X)$ indicates the percentage of the time that X at a given location and point in time will take a value less than or equal to X
Duration	$F(X) < 0.2$	Monthly	SM	Number of consecutive months over a 30-yr window in which the $F(X) < 0.2$, which indicates the onset of a moderate drought
Relative change	Area $F(X) < 0.2$	Monthly	SM	Relative change of area of a basin $F(X) < 0.2$ w.r.t. a reference period
Seasonal forecast				
Probabilistic	Monthly quintiles X_p	1,…,6	Q, SM, R	Percentage of realizations of monthly forecasted X for every quintile category and lead time; cutoffs: 20th, 40th, 60th, and 80th percentiles
Probabilistic	Above X_{10}	1,…,6	Q, SM, R	High values; percentage of realizations above the reference monthly X_{10} for each ℓ
Probabilistic	Below X_{90}	1,…,6	Q, SM, R	Low values; percentage of realizations above the reference monthly X_{90} for each ℓ

hindcast data representative of the period 1993–2011. The SF chain differs from the CP chain because it requires initialization of state variables for each monthly simulation. The workflow of the SF chain is depicted in Fig. 3b, with monthly operations (blue) carried out on a routine basis, and occasional operations (red). The main steps of the operational SF chain are (i) update initial state variables through restart files obtained from the respective E-OBS reference historical run, (ii) obtain and downscale SF-GCM data, (iii) run the multihydrometeorological modeling chain to generate an ensemble of tECVs, (iv) estimate an ensemble of SCIIs based on the tECVs, (v) estimate the ensemble forecast uncertainty and skill, and (vi) export ensemble outputs/indicators for visualization. Occasional tasks include the creation of historical initial states for the HMs/LSMs and the recalibration of the HMs/LSMs.

The EDgE SCIIs for the water sector were identified from an open survey conducted at the onset of the project (see appendix E). The users—which included members from consultancy, academia, NGOs, water user associations, local, regional, and national authorities—were asked for the information that they would require from a climate service delivering SF and long-term CP. The end users requested SF-based indices for lead times from 1 to 6 months, and CP-based indices for each decade up to 2100. To restrict the number of projections, two RCPs were chosen, which define the lowest (RCP2.6) and highest (RCP8.5) emission scenarios. End users were also interested in obtaining information regarding relative changes compared to the baseline 1971–2000.

The selected SCIIs for the CP and SF modeling chains are listed in Table 1. SCIIs based on forcing
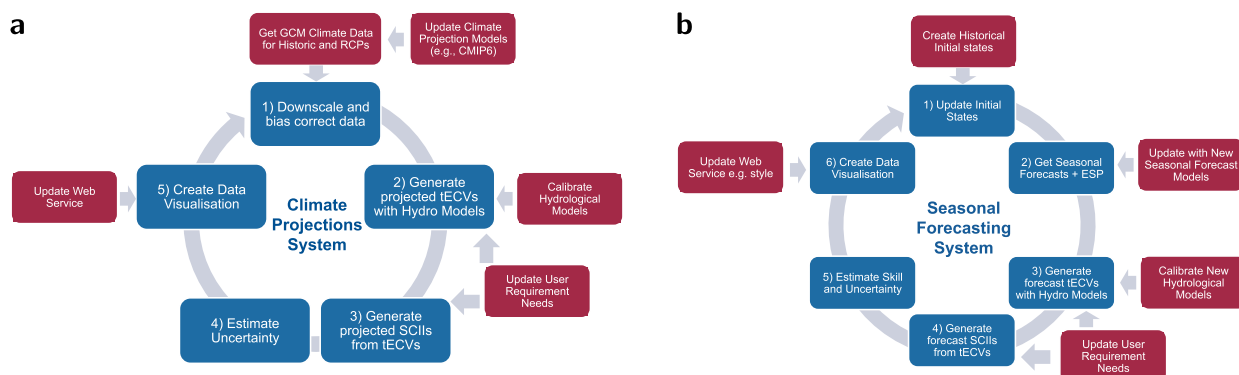
FIG. 3. Workflow of the (a) climate projection and (b) seasonal sorecast model chains. Both chains include occasional (red) and routine tasks (blue).

variables (i.e., precipitation, temperature, and potential evapotranspiration) were derived directly from the bias-corrected and downscaled forcing datasets.

The majority of CP-based SCIIs denote the changes in a given hydroclimatic variable (e.g., high flow as Q_{10}) over a future 30-yr period with respect to their historical reference values. A reference value was established for every GCM–HM or GCM–LSM combination using the model simulations for 1971–2000. Subsequently, SCIIs were estimated for every 30-yr period, starting from 2011 until 2095 with a gap of 5 years between each 30-yr periods (i.e., 2011–40, 2016–45,…, 2066–95). Some of the CP-based SCIIs related to soil moisture and groundwater recharge are expressed as quantile indices of monthly values for the entire period 1971–2099 (Samaniego et al. 2013). Figure 4 shows the projected changes in one of the

SCIIs (high-flow indicator Q_{10}) for different future periods using the multimodel ensemble consisting of five GCMs and four HMs/LSMs. This figure shows, for example, the level of detail made available by the EDgE simulations. Significantly drier hydrological conditions are expected across the Mediterranean region. Extended analyses of the implications of global warming on floods, low flows, and soil moisture droughts based on EDgE water projections are available in Thober et al. (2018), Marx et al. (2018), and Samaniego et al. (2018), respectively.

The SF-based SCIIs are expressed as the percentage of realizations that detect a reference-based indicator at a given month and lead time (varying from 1 to 6 months) to assess the accuracy of the available seasonal forecast data. The reference is based on the E-OBS historical simulations conducted
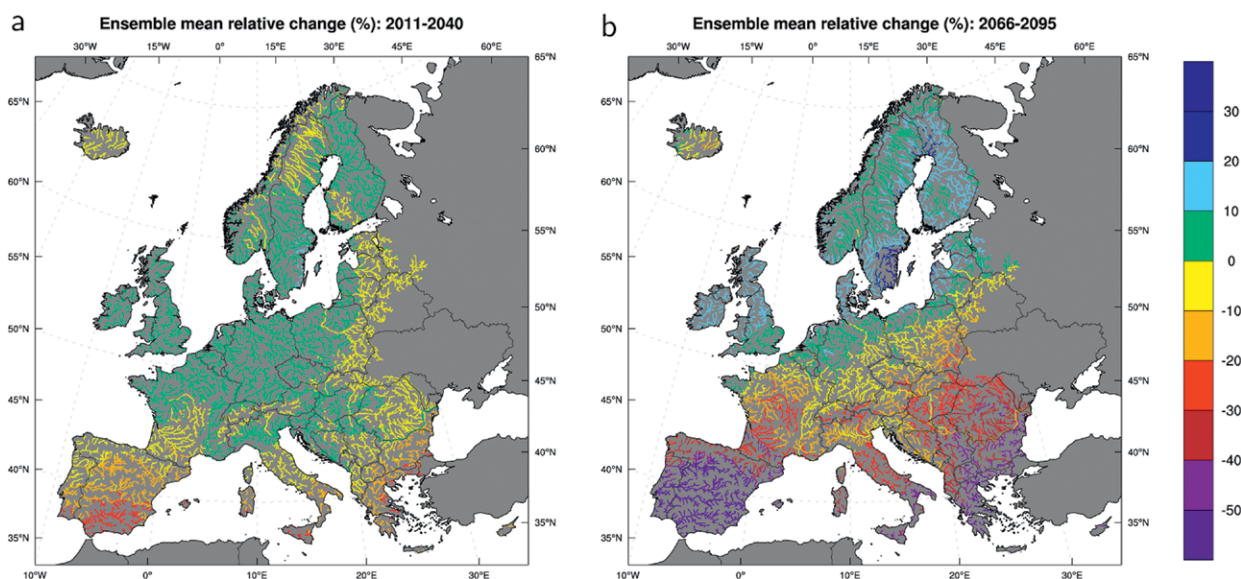
FIG. 4. Multimodel ensemble mean of the projected changes in a high-streamflow indicator Q_{10} for two future time periods: (a) 2011–40 and (b) 2066–99, both under the RCP8.5 scenario. Note that Q_{10} is the daily streamflow equaled or exceeded 10% of the time over a 30-yr window. The historical reference period is from 1971 to 2000.
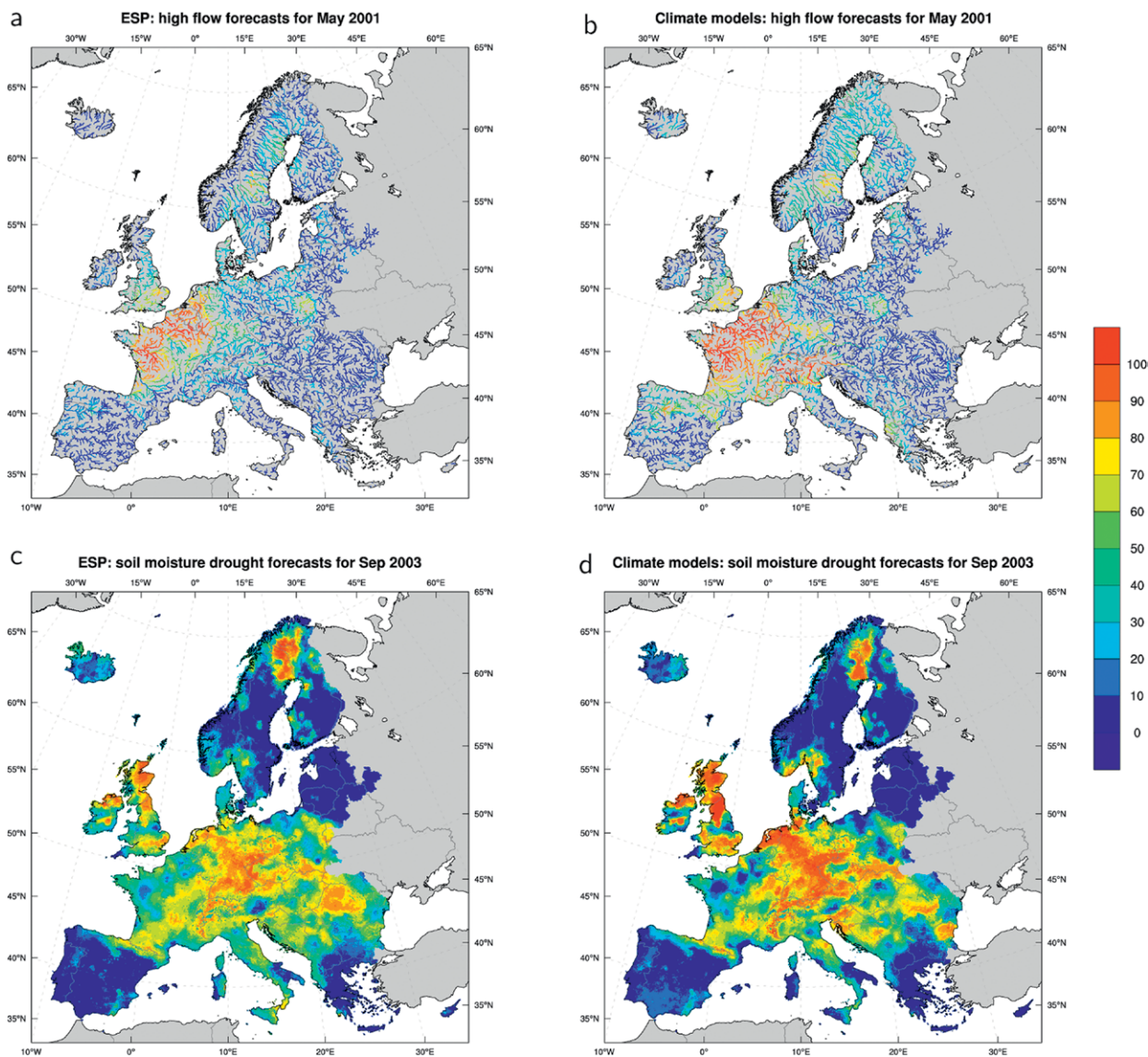
FIG. 5. Multimodel ensemble mean percentage of forecast realizations detecting the reference run–based high flow Q_{10} and soil moisture drought indicators at 1-month lead time for the (a),(c) ensemble streamflow predictions (ESP) and (b),(d) climate model (GCM) forced hydrological model runs. The reference runs are based on hydrological model simulations driven by the observed meteorological forcing data (here, the E-OBS dataset).

for every HMs/LSM separately. For example, the high flow indicator quantifies the percentage of ensemble realizations above the reference monthly streamflow that is exceeded 10% of the time Q_{10} for each lead time. The value of Q_{10} is derived for each calendar month from the relevant reference model run. Similarly, the low flow indicator was based on the assessment of the streamflow forecasts falling below the reference monthly streamflow that is exceeded 90% of the time. Another category of SF-SCIIs are the counts of monthly forecasts falling within each of the quantile intervals (≤20%, 20%–40%, 40%–60%, 60%–80%, ≥80%), with the five level limits derived

for each calendar month from the relevant historic reference run.

Figure 5 presents the results of two SF-based SCIIs related to high river flow and soil moisture drought (i.e., quantile level ≤ 20%) at 1-month lead time. It shows two extreme events: a flood occurring in May 2001 in France and the Alps regions (Figs. 5a,b); and the 2003 soil moisture drought in central Europe (Figs. 5c,d). The climatological-based Ensemble Streamflow Prediction (ESP) (Day 1985) was used to provide a benchmark of the GCM-based seasonal forecasts. In general, the GCM-based seasonal forecast showed a larger agreement with observations
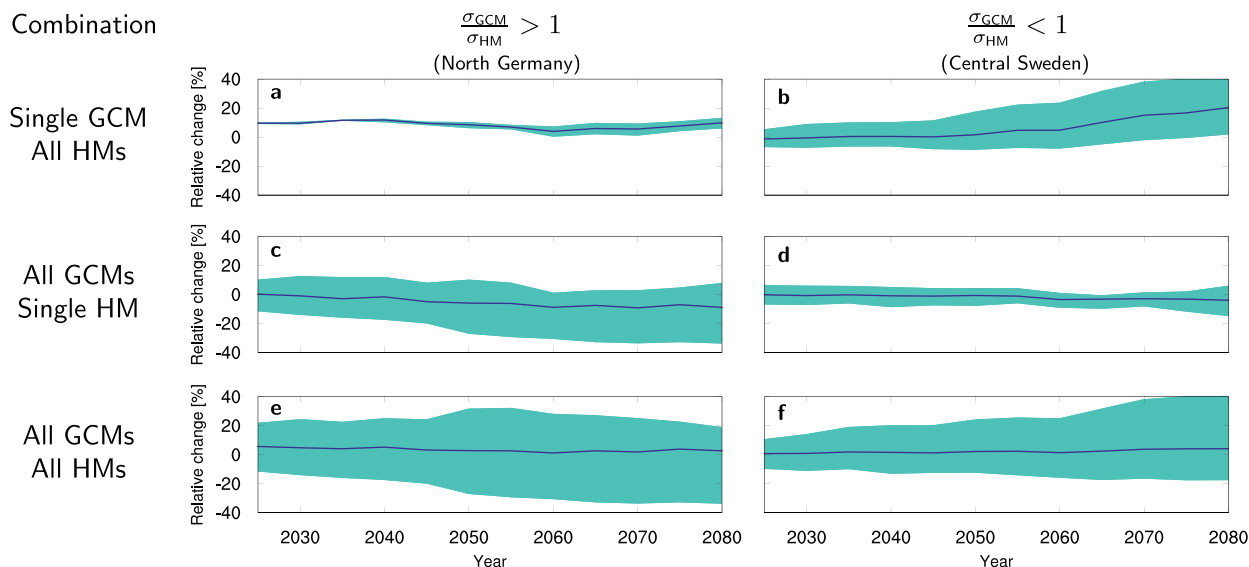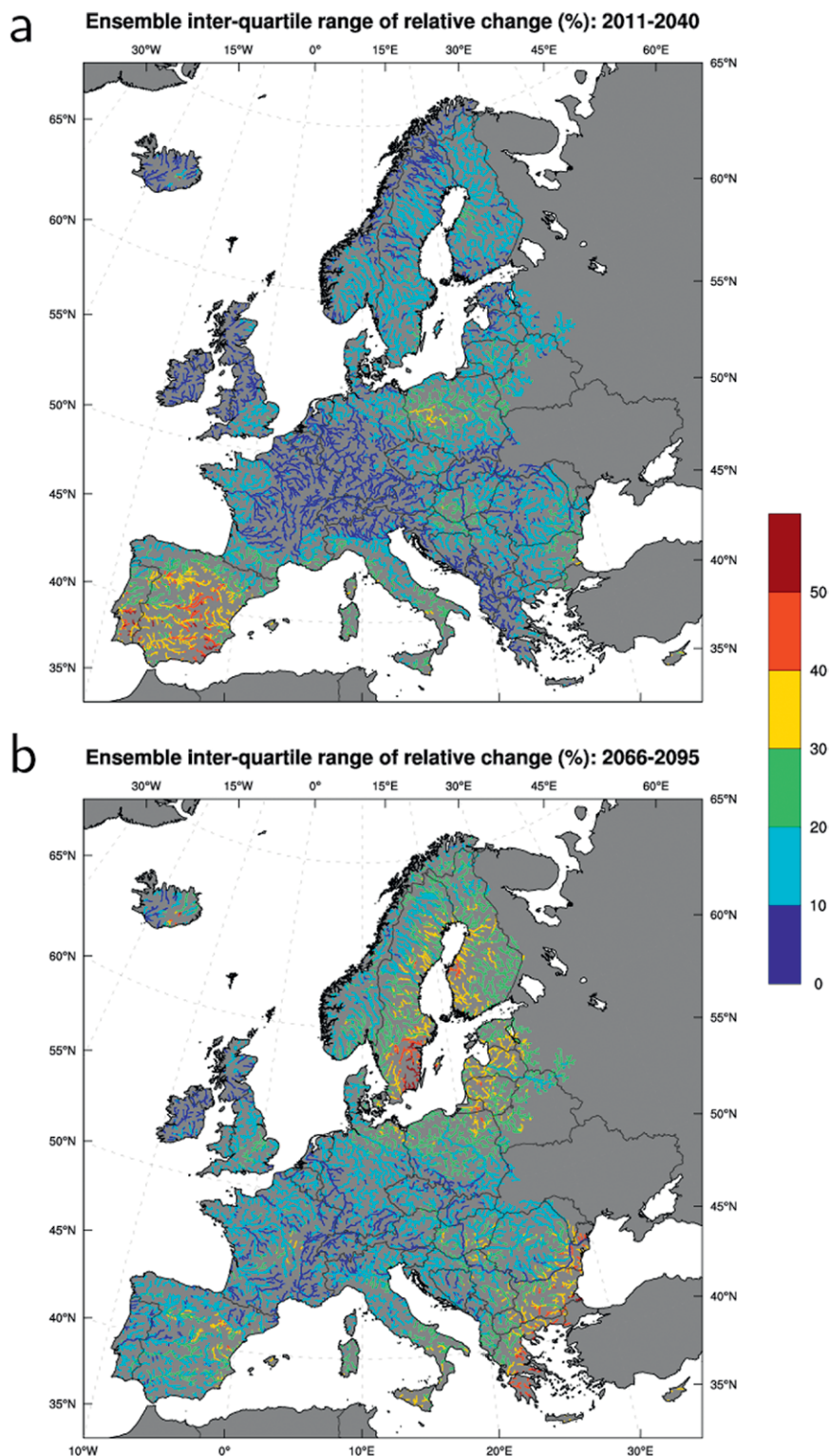
Fig. 6. Uncertainty of the multimodel ensemble of the relative change of high-streamflow indicator Q_{10} for consecutive 30-yr periods with respect to the reference period for different combinations of GCMs and HMs/ LSMs and for two distinctive locations in central Europe (≈52.12°N, 9.38°E) and Scandinavia (≈62.98°N, 16.72°E) characterized by a larger and smaller ratios of GCM/HM uncertainty (σ_{GCM}/σ_{HM}), respectively. The reference runs are based on hydrological model simulations driven by the historical forcing data of the respective GCM.

than the ESP-based forecast. A detailed analysis of the seasonal forecasting skill and uncertainty associated with EDgE seasonal streamflow hindcasts is reported by Wanders et al. (2019).

It was clear from the consultation with users that information on the uncertainty of climate projections and the skill of seasonal forecasts is critical for outputs to be used to their full potential (Taylor et al. 2015). This is especially true for large ensembles, like the EDgE seasonal forecast and climate projection chains, that contain many combinations of models, seasons, and geographical locations. Within this project, skill and uncertainty information were combined with expert knowledge to provide end users with both quantitative and qualitative information, designed to facilitate interpretation. The simulation quality was determined for each individual hydrometeorological model combination and initialization months, and was conducted for all parts of the geographical domain independently. In the following two sections we analyze the uncertainty and the skill of both modeling chains.

UNCERTAINTIES IN CLIMATE PROJECTIONS. Due to project limitations, uncertainty analysis did not include that from parameter data, downscaling method, and geophysical data. Consequently, the term uncertainty here refers to the spread among the hydrometeorological ensemble members, calculated over all meteorological and hydrological model combinations.

The ability of GCMs and HMs/LSMs to capture the predictive uncertainty of key water fluxes and state variables has been extensively discussed in the literature (e.g., Prudhomme et al. 2014; Giuntoli et al. 2015, and sources therein). The propagation of forcing data and meteorological model uncertainties and its dependency on RCP and time horizon were shown by Samaniego et al. (2017a) for streamflow for a few selected basins. In EDgE, the assessment was extended to all indicators shown in Table 1, and estimated over a large domain and at the high resolution of the EDgE modeling framework. Figure 6 provides clear evidence, when investigating the impact of climate change on water-related variables, of the importance of a multihydrological model ensemble over a single hydrological model to better capture the uncertainty propagation in water fluxes and state variables. There are regions where GCM forcing uncertainty clearly dominates the HM/LSM structural uncertainty for the high-streamflow indicator (see Figs. 6a,c,e), but there are other locations in which the opposite is true (Figs. 6b,d,f). Notably, the spatial distribution of uncertainty varies between all indicators. Consequently, it is not possible to establish a priori the optimal size, or membership, of ensemble GCMs and ensemble HMs/ LSMs. Ideally, the ensemble size should be as large as possible to be able to identify the uncertainties originating from either GCMs or HMs/LSMs. These results clearly indicate that a single model or a subensemble may work well in a given location but perform poorly in

a Ensemble inter-quartile range of relative change (%): 2011-2040

b Ensemble inter-quartile range of relative change (%): 2066-2095

FIG. 7. Ensemble interquartile range of the relative change of high-streamflow indicator Q_{10} for two 30-yr periods with respect to the reference period: (a) 2011–40 and (b) 2066–99, both under the RCP8.5 scenario. Southern Europe exhibits, for example, a significant decrease in uncertainty by the end of the century. The historical reference period is from 1971 to 2000.

another, and hence should not be recommended for an operational system.

The uncertainty of the climate projections under a given RCP were estimated as in Samaniego et al. (2017a), which used the long-term mean of the interquartile range for a given SCII as the uncertainty metric, so that the contributions stemming from GCMs and HMs/LSMs could be disentangled. The 40-member ensemble interquartile range of the relative change of the high-streamflow indicator Q_{10} shows significant regional changes within a given RCP scenario for two future periods (2011–40 and 2066–99 in Figs. 7a and 7b, respectively). Southern Europe, in particular, exhibits a significant decrease in uncertainty by the end of the century. This suggests that GCM projections and HMs/LSMs tend to have a consistent estimate of the projected changes. The same method applied to subsamples of the ensemble can be used to systematically quantify uncertainties stemming from different origins (i.e., uncertainties from hydrological model structure and input uncertainty—given that all HMs/LSMs use an identical set of underlying physiographical land surface characteristics), using, for example, the full ensemble interquartile range spread as the benchmark uncertainty. Users of the online platform can experiment with different combinations to assess the quality of a given combination of GCMs and LSMs/
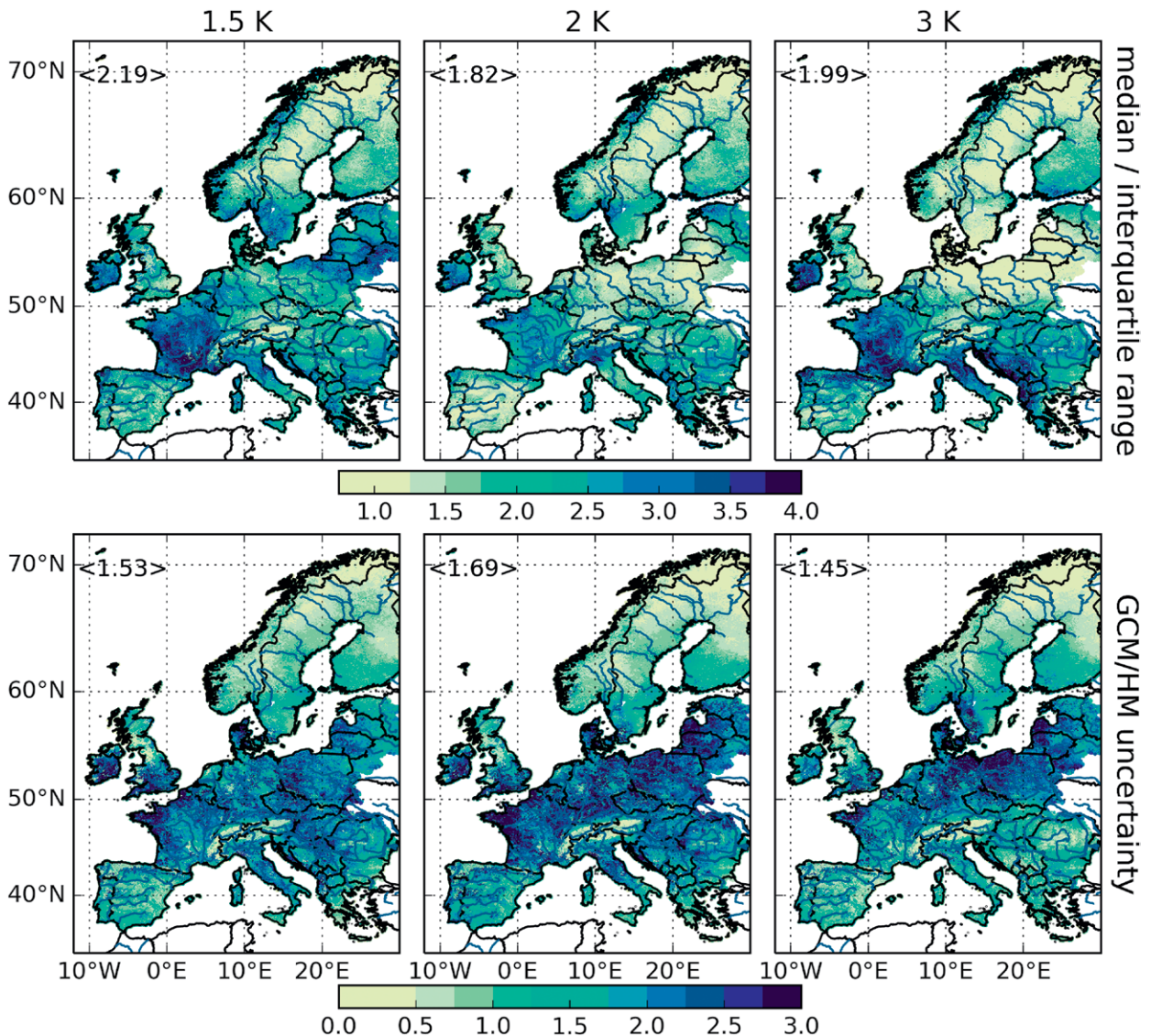
FIG. 8. Uncertainty in the estimated number of drought months per year expressed as (top) signal-to-noise ratio and (bottom) the ratio of the uncertainty contribution of global climate models with respect to hydrological models [using the method described in Samaniego et al. (2017a)]. The columns correspond to a global warming level of (left) 1.5, (center) 2, and (right) 3 K. The number in the brackets in each panel denotes the mean value over space.

HMs. Note, however, that a smaller ensemble might not capture the full uncertainty of the hydrological and meteorological combinations.

A detailed analysis of the contribution by GCMs and HMs/LSMs to the total uncertainty has been evaluated within the High-Resolution Climate Indicators for 1.5 Degree Global Warming project (HOKLIM; www.ufz.de/hoklim), which employed EDgE simulation data. It was found that the total uncertainty in modeled variables is dominated by the choice of hydrological models in Alpine and semiarid regions for both floods and low flows (Thober et al. 2018; Marx et al. 2018). In these regions, different

representations of snow processes and soil water redistribution in the HMs/LSMs have an impact on the projected climate change signal comparable to the different meteorological forcing data from the GCMs.

A further analysis of the GCMs and HMs/LSMs uncertainty contribution to the soil moisture drought duration (duration in Table 1) under different global warming levels reveals that the ensemble spread is dominated by the GCMs in comparison to the HMs/LSMs (Fig. 8, lower panels), with the ratio of GCM and HM/LSM uncertainties (σ_{GCM}/σ_{HM}) being particularly high in the continental and Atlantic region. In these humid regions, the variability of precipitation

among GCMs has the largest effect on simulated soil moisture drought development. In Scandinavia, the Southern Alps, and parts of the Mediterranean, the HMs/LSMs contribute at least as much to the uncertainty as GCMs. In cold regions, this might be related to the importance of snow processes, where the rate of snowmelt and accumulation varies substantially among the HMs/LSMs. In arid regions such as the Mediterranean, the soil water restriction on evapotranspiration also varies among HMs/LSMs. This process greatly influences how fast soils dry and thus drought development. These results provide substantial evidence against the assumption that the uncertainty of derived SCIIs is equally distributed between atmospheric (GCM) and land surface/hydrological models (HMs/LSMs), in accordance with the findings of Samaniego et al. (2017a) for low-flow duration.

The spatial distribution of the signal-to-noise ratio for drought duration calculated as the median divided by the interquartile range is shown in the upper panels in Fig. 8. Low signal-to-noise ratios are generally found in Scandinavia, Germany, and Poland, irrespective of the amount of global warming. Note that the spatial distribution of GCMs/HMs/LSMs uncertainty contribution does not correlate with that of the signal-to-noise ratio (cf. Fig. 8, top and bottom panels). For example, for a global warming of 3 K (Fig. 8, right column), the signal-to-noise ratio is low in Scandinavia, where uncertainty is dominated by HMs/LSMs, but it is also low in Poland, where uncertainty is dominated by GCMs. In other words, low confidence in future projections (i.e., low signal-to-noise ratio) can be created by both HMs/LSMs and GCMs, and be undifferentiated.

SKILL OF SEASONAL FORECASTS. The seasonal forecast skill was calculated employing one of the most commonly used skill scores for seasonal forecasting in meteorology and hydrology: the Brier score (BS; Brier 1950). The BS uses categorical forecast thresholds to determine the quality of the forecast compared to a reference simulation. E-OBS-based simulations were used as reference. Quintile classes were defined for the qualification of the streamflow skill whereas a threshold soil moisture value was used to discriminate soil moisture drought events (a soil moisture value that is exceeded 80% of the time in a given calendar month and location). Forecasts that hit the reference class of the E-OBS reference simulation were considered as "skillful," whereas those that did not are denoted as "unskillful." A lower BS indicates a better forecast. The spread of the BS values is estimated as a standard error: $\varepsilon = t_\alpha (n-1) S/\sqrt{n}$. Here, n denotes the ensemble size, S, the standard deviation of the ensemble statistic (BS), and t_α denotes the Student's t critical value for $1 - \alpha$ confidence interval (95%) and $n - 1$ degrees of freedom. Consequently, the uncertainty of the mean Brier score $\overline{BS}$ at a given location is $\overline{BS} \pm \varepsilon$. A forecast with large standard error is deemed highly uncertain because the accuracy of the forecast is low.

The ESP approach (Day 1985) is an often-used benchmark in seasonal hydrological forecasting (e.g., Thober et al. 2015; Wanders et al. 2019) and provides the forecast skill that can be obtained from the initial hydrological conditions. The method was implemented in EDgE using a hindcast starting at month m and year y, generated from 15 years randomly drawn from the E-OBS forcing data during the period 1993–2011, starting at the calendar month m. The sample size was selected to resemble the maximum number of ensemble members of the seasonal forecast models, so that the forecast quality cannot be influenced by different ensemble size. For shorter lead times and in regions that have a long hydrological memory, the ESP can provide a highly skillful forecast because the impact of the initial hydrological conditions dominates the seasonal predictability (Wanders et al. 2019). For longer lead times, the ESP tends to become close to hydroclimatology and the performance of ESP-based forecasts is comparable to that of GCM-based forecasts. Users can evaluate the added value of the dynamical (GCM based) seasonal forecast by comparing it with ESP.

The full ensemble of GCMs and HMs/LSMs exhibits a BS of 0.14 over the entire domain for seasonal soil moisture droughts at 1-month lead time (Fig. 9a). The standard error of the BS values for the full ensemble is 0.02 on average. In comparison, the average values using ESP is slightly higher than that of the full ensemble (0.16), but has a standard error that is on average 5 times higher than that of the full ensemble. Notably, the spatial patterns in BS are comparable among the full ensemble and ESP. This indicates that forecasting skill does not only depend on the meteorological input, but also on other factors such as the persistence of initial hydrologic conditions (Wanders et al. 2019). Notably, the skill of the full ensemble (SF-GCM + LSM/HM) is consistently higher than that of ESP at almost all locations in Europe (BS tends to have lower values). This is remarkable because the analysis is favorable to ESP for two reasons: 1) the SF-GCMs were not bias corrected prior to being used as forcing data to the hydrological models, and 2) the reference values for the BS estimates are based on HMs/LSMs

simulations using E-OBS forcing. As a consequence, the ESP forecasts and the reference values share the same climatology, whereas the SF-GCMs have their own climatologies. The single best performing SF-GCM, ECWMF-S4, provides a minor improvement with respect to the full ensemble, but has a threefold standard error as the latter (Figs. 9e,g). Similarly, the skill of individual ECMWF-S4/HMs/LSMs combinations is slightly higher than that of the full ensemble and all of these outperform ESP (see Fig. A2 in appendix D for individual combinations). However, the uncertainty for the individual model combinations cannot be estimated (because its sample size for the HMs/LSMs is one in this case), which lowers the credibility of their skill. Overall, any combination of SF-GCMs/HMs/LSMs provides a higher forecasting skill than ESP for soil moisture droughts at 1-month lead time. The full ensemble exhibits the highest forecasting skill with respect to both bias and uncertainty among all possible combinations. This may not be true everywhere, which is why users can choose the SF-GCMs/HMs/LSMs that provides the best ensemble forecast for a given location and time in the web interface of the demonstrator.

Improvement of skill using SF-GCM-driven forecasts instead of ESP can be measured using the Brier skill score (BSS), with a
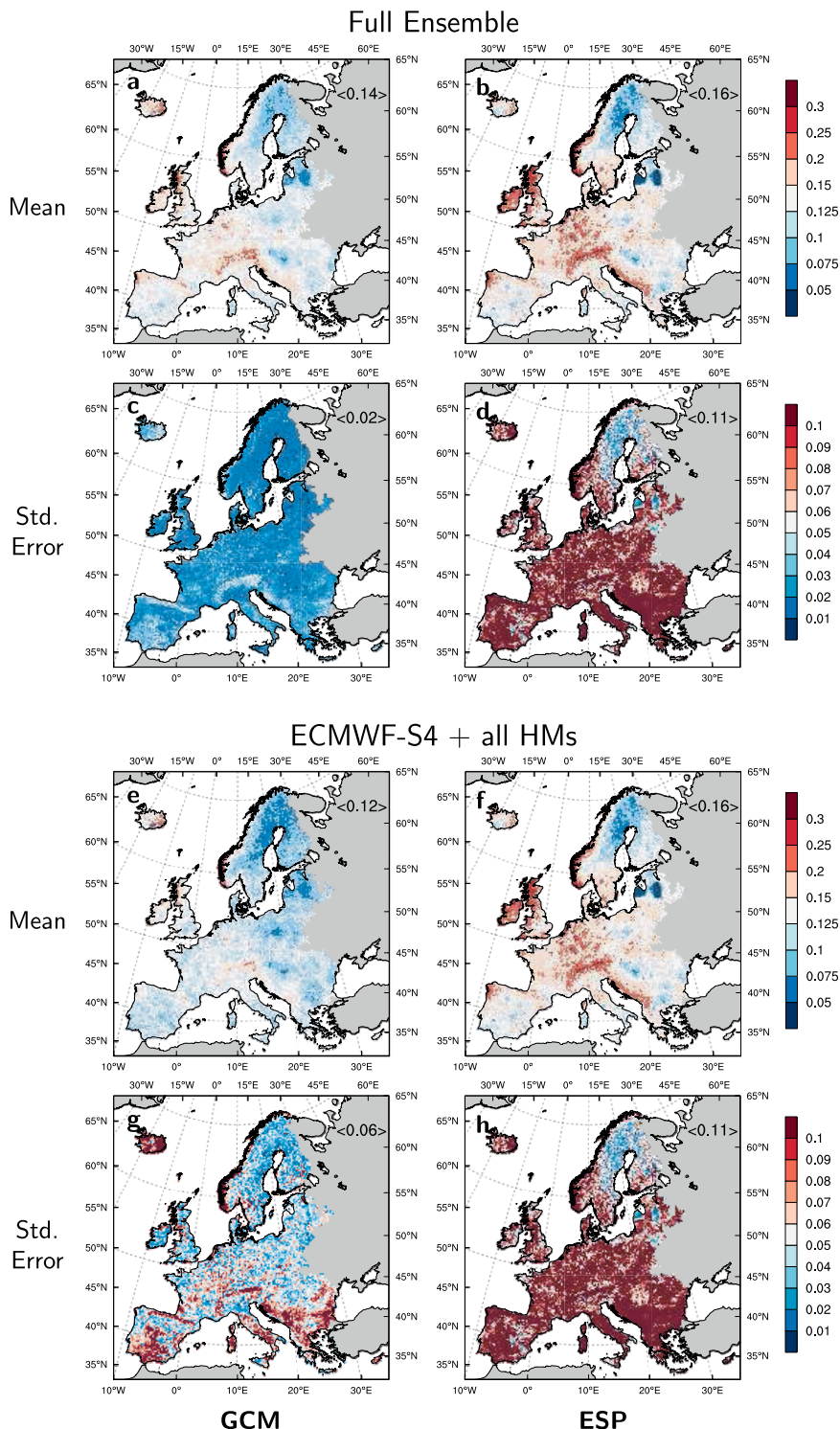
FIG. 9. Effects of the subensemble selection on the Brier score (BS) skill in SF mode for the first soil moisture quantile indicator (drought events). (a),(c) Ensemble mean and standard error of BS obtained with the full ensemble, which includes all SF-GCMs and all HMs/LSMs, respectively. (e),(g) As in (a) and (c), but for the subensemble of one SF–GCM (ECMWF-S4) with all HMs/LSMs. (b),(d),(f),(h) Corresponding statistics of BS obtained with ESP instead of the SF–GCM model combination. The number in the brackets in each panel denotes the mean value over space.
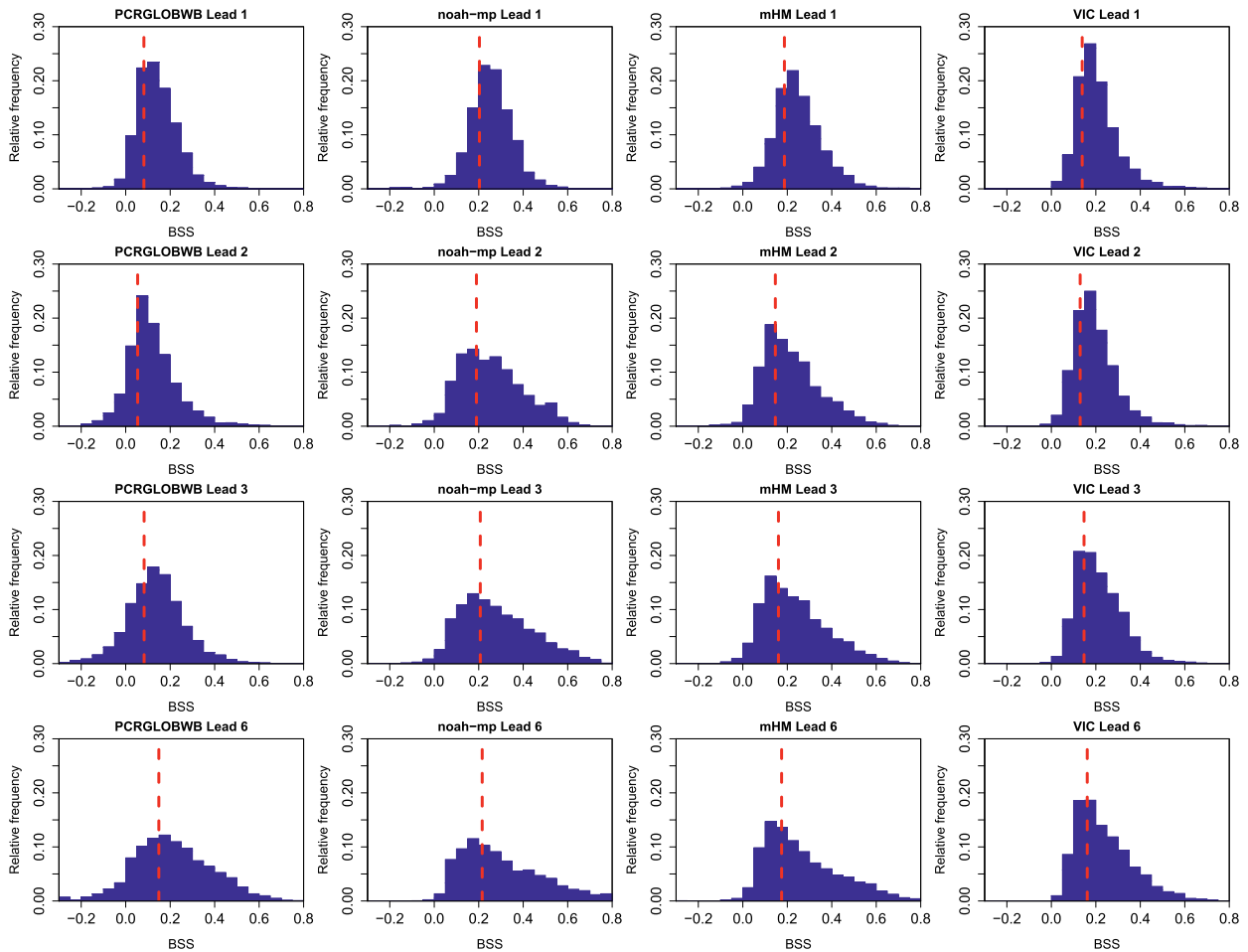
FIG. 10. Brier skill score (BSS) for dynamical seasonal forecasts over Europe. BSS is calculated for each gauging station individually (shown are hydrological forecasts and projections for improved decision-making in the water sector in Europe) for each lead time and model separately, using ESP as reference (BS_{ESP}). No separation has been made with respect to the forecast season. The red dashed line indicates the median of the distribution over 465 streamflow stations during the hindcast period. Positive BSS values indicate an improvement in the dynamical forecasts compared to the climatological forecast.

BSS > 0 showing an improvement. BSS was calculated at every location and for every SCII, and is available on the online demonstrator (EDgE 2017). Figure 10 gives the histogram of the BSS for streamflow at 465 selected gauging locations, and Fig. A1 shows the LSM/HM model performance. Results show a strong relationship between the LSM/HM model performance and BSS, with high-performing hydrological models being associated with high median BSS values and a histogram skewed toward the right. For example, the median BSS are around 0.2 for mHM and Noah-MP but are slightly less than 0.2 for VIC and less than 0.1 for PCR-GLOBWB (except for 6-month lead-time forecasts). This could be related to the initial model skill, but is also linked to the impact of the initial hydrological conditions. PCR-GLOBWB tends to show a long hydrological memory (Wanders et al. 2019), which limits the impact of the dynamical

forecast improvement. Hydrological models that respond rapidly to precipitation or temperature changes are more likely to benefit from accurate dynamical seasonal forecasts and thus show a strong improvement in the BSS. This suggests that future hydrological models with more accurate representations of the observed hydrology and applied in regions with shorter hydrological response times are likely to profit more from SF-GCMs.

CONCLUSIONS AND OUTLOOK. The EDgE project was one of the two proof of concepts (PoCs) commissioned as precursors to the development of a fully operational system for the water sector in Europe. Three highlights of the EDgE approach are 1) the unprecedented high resolution and consistency of inputs of the multimodel hydrological simulations at time scales of seasonal forecasts and climate

projections, 2) the systematic uncertainty estimation for 36 codesigned water impact indicators, and 3) the delivery of a high quality water information service tailored to the needs of end users. These characteristics of EDgE are preconditions for users to make informed decisions and therefore constitute the key for improved decision-making. The results shown here also highlight the relevance and value of having a multihydrological model ensemble capable of capturing the total uncertainty of the prediction chain.

Operationalization of the modeling chain would be straightforward because it was designed to be upgradable and scalable. Including new hydrological models or updated versions of Noah-MP, mHM, VIC, and PCR-GLOBWB would be possible due to the flexibility of the operational framework. Similarly, adding new climate models for the CP and the SF modeling chains would only be limited by the storage capacity and computational power available.

Skillful seasonal forecasts will depend on the quality of the initial conditions, the performance of the hydrological models against observations, the employed spatial resolution, and skill of the SF-GCM model. One limitation of this PoC is that no bias correction is applied to the SF-GCM data because of the short hindcast period of 19 years. A longer hindcast period should be used during an operationalization phase. As shown in this PoC, model cross-validation at gauged locations is an excellent diagnostic tool to assess model deficiencies (i.e., model parameterization and/or structure). Ideally, a diagnostic tool should be part of the operational system giving updated information on a regular basis, with updates on skill assessments regularly conveyed to end users.

Next steps within the development of this water information system should focus on anthropogenic influences that alter the natural course and water balance of the hydrologic cycle in all hydrological models. These will greatly improve the quality of seasonal forecasts and the usability of the climate projections. For this reason, high-resolution data of water bodies, dam systems, and water distribution infrastructure in Europe should be further assembled. For climate projections, it is crucial to include dynamic land-cover/land-use models coupled with hydrological models. Further work should also provide indicators related to water quality and river temperature as needed by end users. All hydrological and land surface models in EDgE used the same underlying static datasets (e.g., soil and land cover). These were, however, processed differently for the individual models. Applying a seamless parameterization following Samaniego et al.

(2010, 2017b), would help to further increase the consistency among hydrological model simulations. All these factors would improve the quality and realism of the impact indicators.

APPENDIX A: BIAS CORRECTION AND DOWNSCALING. The bias correction technique proposed by Hempel et al. (2013) corrects systematic deviations of simulations from historical observations, but preserves the absolute warming signal for temperature and the relative warming signal for precipitation. Daily variability around the monthly means has been adjusted by a quantile mapping assuming a normal and gamma distribution for temperature and precipitation residuals, respectively (Hempel et al. 2013).

Coarse GCM daily values (CP and SF modes) of precipitation, daily mean temperature, daily maximum and minimum temperature were downscaled from their native resolution ($\ell_2 = 1°, 0.75°$, or $0.5°$) to $\ell_1 = 5$ km using the external drift kriging (EDK) algorithm. EDK is an interpolation technique that provides the best linear unbiased estimation at unknown locations (Kitanidis 1997). It also includes a drift governed by the terrain elevation. The advantage of this procedure over other estimation approaches is that it can account for the finescale orographic effects in interpolated precipitation and temperature fields. In this case, EDK can be considered as a simple but unbiased form of statistical downscaling because it uses coarser predictors and terrain characteristics

TABLE A1. Physiographic information used for the EDgE Project.

Description	Dataset name	Data owner	Source
Elevation	EU-DEM	EEA	www.eea.europa.eu/data-and-maps/data/eu-dem
	GOTOPO30	USGS	https://lta.cr.usgs.gov/GTOPO30
Pan-European River and Catchment Database	CCM2 v2.1	EC -JRC	http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23
Soils texture	SoilGrids1km	ISRIC	www.isric.org/explore/soilgrids
Land cover	GlobCOVER v2	ESA	http://due.esrin.esa.int/page_globcover.php
	CLC00, CLC06, CLC12, CLC90 v18.4	Copernicus, ESA	http://land.copernicus.eu/pan-european/corine-land-cover
Hydrogeology	IHME1500 v11/	BGR IHME	www.bgr.bund.de/ihme1500
Leaf area index	GIMMS	UMD	https://iridl.ldeo.columbia.edu/SOURCES/.UMD/.GLCF/.GIMMS/.NDVIg/.global/.dataset_documentation.html
World Register of Dams	WRD	CIGB-ICOLD	www.icold-cigb.org/GB/world_register/world_register_of_dams.asp

within a variance minimization scheme. The spatial weights determining the EDK interpolation depend on the spatial variability of the meteorological fields, which is quantified by a semivariogram. The semivariograms for the daily precipitation and temperature were derived from daily E-OBS station data (Haylock et al. 2008) covering the entire domain following the procedure proposed by Zink et al. (2017). It should be noted that EDK does not modify the long-term trends, making this a suitable technique for climate change impact studies. After downscaling the precipitation and temperature fields, the Mountain Climate simulator tool (Bohn et al. 2013) was used to generate the 3-hourly forcing data of air temperature, downward shortwave and longwave radiation, specific humidity and surface pressure required for running the LSMs. Daily wind speed climatology was derived from the EFAS forcing dataset.

APPENDIX B: LAND SURFACE DATA. A key feature of EDgE is the use of consistent land surface data at a high spatial resolution of 500 m. A summary of all open-source data used is listed in Table A1. Terrain characteristics (e.g., elevation, slope, aspect, flow direction, and flow accumulation) were derived from the joined Europe-wide (EU) and Global (GOTOP30) Digital Elevation Model (DEM). The Global dataset was used for delineating river basins at those locations that were not covered by the EU-DEM dataset. All datasets were reprojected to the European Terrestrial Reference System (ETRS) 1989 Lambert azimuthal equal area coordinate reference system with a spatial resolution of 500 m for consistency. The spatial domain covers the entire drainage area of all rivers within the pan-EU territory.

APPENDIX C: HMS/LSMS CALIBRATION AND EVALUATION. Different parameter estimation strategies have been used for the individual HMs/LSMs, based on the expert knowledge of the different modeling teams involved: for example, 1) estimation of global transfer function parameters, 2) manual tuning of selected (sensitive) model parameters, and 3) bilinear interpolation based on coarser-resolution parameter sets. The first approach, which leads to a seamless parameterization of hydrological model parameters, as described in Samaniego et al. (2010, 2017b), was applied to mHM and PCR-GLOBWB. Given the resources and time constraints of the EDgE project, a manual calibration was applied for Noah-MP focusing on adjusting the surface evaporation resistance parameter, which was identified as highly sensitive by Cuntz et al. (2016). The parameters for the VIC model (Liang et al. 1994) were mapped from a global simulation (Sheffield and Wood 2007) so that they are consistent with the land surface parameters specified for the other models.

All models were driven with the downscaled 5×5 km^2 historical E-OBS data (Haylock et al. 2008) and evaluated against observed GRDC streamflow data for 357 basins (www.bafg.de/GRDC/). It should be noted that the GRDC basins mainly consist of large rivers, which are often heavily modified by anthropogenic activities and infrastructure (e.g., large hydropower dams). As a result, it is difficult for models describing naturalized streamflows to simulate GRDC data influenced by river regulations. Parameter estimation for each HMs/LSMs was conducted only on river basins without large hydroinfrastructure facilities because no dam
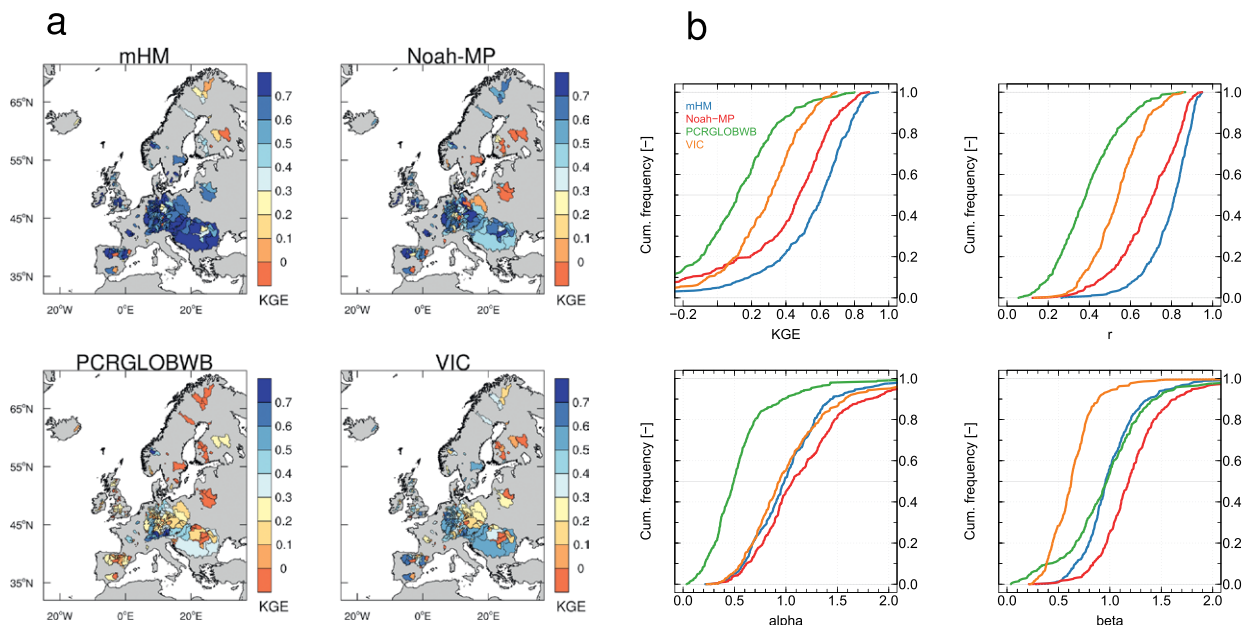
FIG. A1. Evaluation of the hydrological models using the observed daily streamflow over 357 European basins forced with the E-OBS meteorological data. (a) Spatial maps of the daily Kling–Gupta efficiency (KGE) for HMs/LSMs. (b) Cumulative frequency of the daily KGE measure and decomposition into its three components (correlation *r*, ratio of variability *α*, ratio of bias *β*). Model statistics are based on the 30-yr period covering 1966–95.

management information was available at the time of the analysis.

Figure A1 summarizes the model performance in terms of predicting daily streamflow data for all HMs/LSMs. In total, 357 diverse basins with a median area of around 1,700 km², and a complete streamflow record for a 30-yr period (1966–95), are evaluated using the Kling–Gupta efficiency (KGE; Gupta et al. 2009). Figure A1a presents the basinwise spatial evaluation and reveals that the model performance based on historical forcing data strongly depends on model type and region, which highlights the added value of using multiple hydrological models. All models have some difficulties in capturing streamflow dynamics in the northeastern part of the domain, where snowmelt processes are dominant. Figure A1b details quantitative estimates for KGE and its three components: correlation *r*, ratio of variability *α*, and ratio of mean *β*. The median KGE varies between 0.1 and 0.6. The mHM and PCR-GLOBWB models provide unbiased streamflow estimates at the majority of the basins, while Noah-MP and VIC tend to overestimate and underestimate the mean flows, respectively. In the majority of the 357 basins, the variability of observed streamflow flow is well captured by all models except for PCR-GLOBWB. Overall, mHM exhibits the best model performance followed by Noah-MP, VIC, and PCR-GLOBWB. Within the project, the stakeholders evaluated the model results in their

target basins and found that the all models exhibited reasonable performance.

APPENDIX D: PERFORMANCE OF ECMWF-S4 DRIVEN FORECASTS. The individual HMs/LSMs (i.e., mHM, Noah-MP, PCR-GLOBWB, and VIC) driven by ECMWF-S4 show unique spatial patterns of the seasonal soil moisture drought forecasting skill at 1-month lead time (Fig. A2). This analysis provides insights on how the individual models contribute to the forecasting skill of the full ensemble. The spatial distribution of BS values for the PCR-GLOBWB is very different from that of the other models (Fig. A2c). Notably, this model also has the highest forecasting skill both for ECMWF-S4 and ESP-based forecasts among all four models. It appears as a paradox that the model with the least ability to simulate observed streamflow receives the highest forecasting skill. PCR-GLOBWB shows a high underestimation of both the observed variability and correlation (*α* and *r* in Fig. A1). This indicates that this model has a high persistence, which implies that the ECWMF-S4 and the ESP-derived forecasts do not deviate too much from the E-OBS-based reference run. ECMWF-S4-based forecasts for mHM and Noah-MP outperform ESP and their forecasts are overall comparable (Figs. A2a–d). Among all models, VIC shows the least skill for both ECMWF-S4 and ESP-based forecasts (Figs. A2g,h). Notably, it also has the
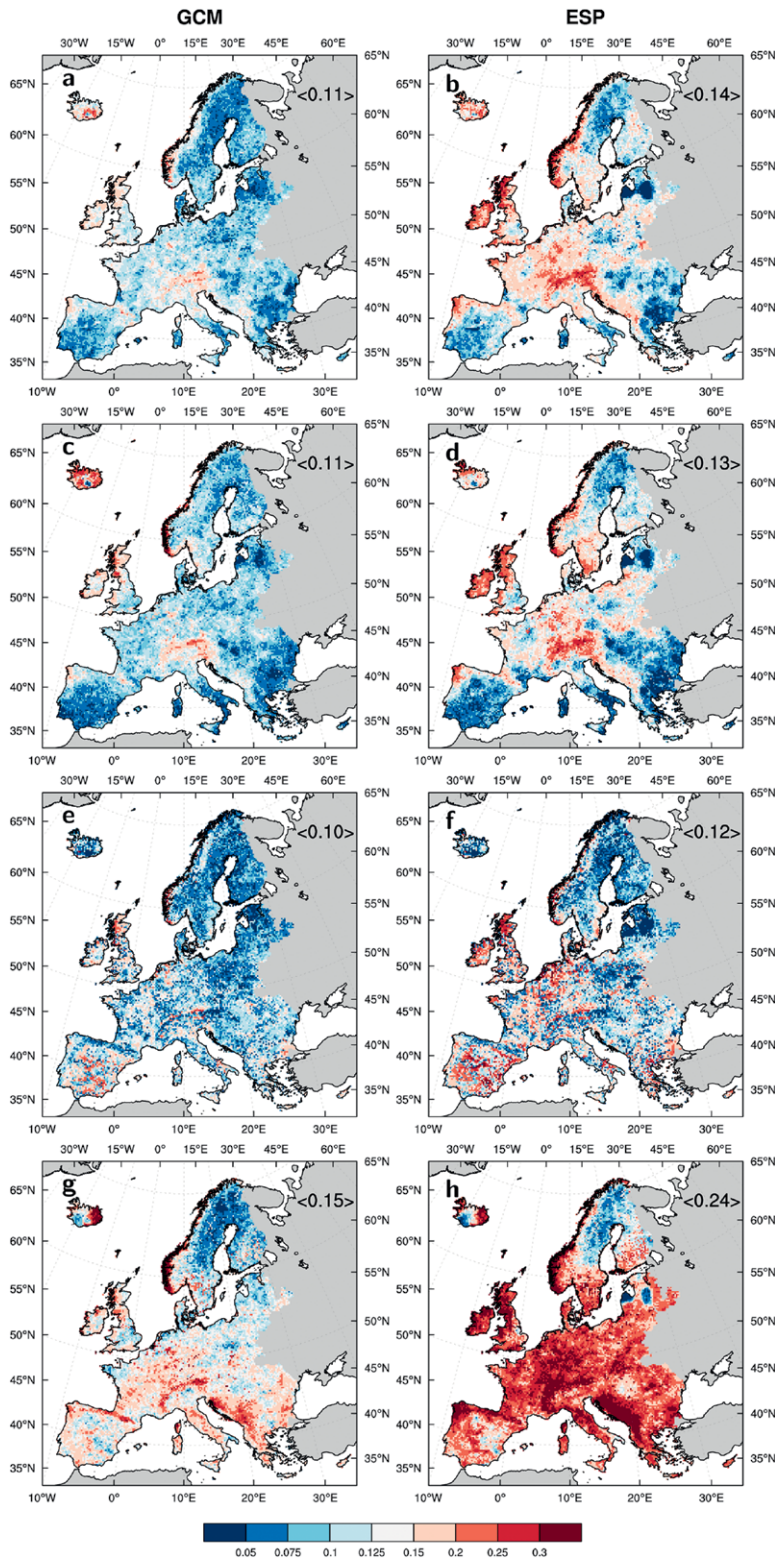
Fɪɢ. A2. Effects of the subensemble selection on the Brier score (BS) skill in SF mode for the first soil moisture quantile. (a),(c),(e),(g) BS obtained with ECMWF-S4 and the mHM, Noah-MP, PCR-GLOBWB, and VIC, respectively. (b),(d),(f),(h) BS obtained with ESP and the corresponding HM/LSM. The number in the brackets in each panel denotes the mean value over space.

largest difference between these two forecasts. In other words, VIC shows the highest skill improvements of 37% for ECMWF-S4 forecasts with respect to ESP-based ones. It is unknown which characteristics of VIC are causing this behavior. Among the four models considered in the ECMWF-S4 ensemble (Figs. 9a,c), VIC contributes the most to the relatively higher BS values of this ensemble with respect to the ESP ensemble. Overall, there is no model that outperforms the other models at all locations, which highlights the necessity of a multimodel approach.

APPENDIX E: STAKEHOLDER FOCUS GROUPS AND FEEDBACK. Stakeholder focus groups (FGs) in Norway, Spain, and the United Kingdom were formed as part of the EDgE project, comprising representatives of national government agencies, regional and local government authorities, international water and hydropower companies, agricultural sector, river basin authorities, consultancies, and academic sector. To examine user needs in detail, each focus group considered a different aspect of water information and decision-making: water supply in the United Kingdom; catchment planning and agriculture in Spain; hydropower generation and local authority planning in Norway. The number of active focus group members totaled 29: 11 in the United Kingdom, 6 in Norway, and 12 in Spain. Stakeholders in Germany were analyzed in a follow-up project (HOKLIM; www .ufz.de/hoklim).

Stakeholders welcomed the information provided by the EDgE demonstrator that combines seasonal forecasts and climate projections in a single platform, seeing it as a useful addition to the information they currently had access to. The international companies saw also a large potential in this system because it provides a consistent dataset across Europe. In

general, it was found that the seasonal forecasts need to have a better skill before they can be used operationally, although they would be used as additional information to climatology. However, the value of using multiple hydrological models to assess hydrological modeling uncertainty was generally appreciated by stakeholders. A European sectoral information system like EDgE was thought to be useful for countries that do not have national climate services providing seasonal forecasts and climate and hydrological projections, but it should have higher spatial resolution to replace currently available national systems.

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: Skilful seasonal forecasts of streamflow over Europe? *Hydrol. Earth Syst. Sci.*, **22**, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018.

Bohn, T. J., B. Livneh, J. W. Oyler, 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. Meteor.*, **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.

Cherkauer, K. A., L. C. Bowling, and D. P. Lettenmaier, 2003: Variable infiltration capacity cold land process model updates. *Global Planet. Change*, **38**, 151–159, https://doi.org/10.1016/S0921-8181(03)00025-0.

Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S. Thober, 2016: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *J. Geophys. Res. Atmos.*, **121**, 10 676–10 700, https://doi.org/10.1002/2016JD025097.

Day, G., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Res. Plann. Manage.*, **111**, 157–170, https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157).

Demargne, J., and Coauthors, 2014: The science of NOAA's operational hydrologic ensemble forecast service. *Bull. Amer. Meteor. Soc.*, **95**, 79–98, https://doi.org/10.1175/BAMS-D-12-00081.1.

EDgE, 2017: End-to-end Demonstrator for improved decision making in the water sector in Europe. EDgE, http://edge.climate.copernicus.eu.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, https://doi.org/10.1029/2002JD003296.

Emerton, R. E., and Coauthors, 2016: Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev.: Water*, **3**, 391–418, https://doi.org/10.1002/wat2.1137.

Giuntoli, I., J. P. Vidal, C. Prudhomme, and D. M. Hannah, 2015: Future hydrological extremes: The uncertainty from multiple global climate and global hydrological models. *Earth Syst. Dyn.*, **6**, 267–285, https://doi.org/10.5194/esd-6-267-2015.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003.

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.

Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek, 2013: A trend-preserving bias correction—The ISI-MIP approach. *Earth Syst. Dyn.*, **4**, 219–236, https://doi.org/10.5194/esd-4-219-2013.

Hoeg-Guldberg, O., and Coauthors, 2019: Impacts of 1.5°C global warming on natural and human systems. *Global Warming of 1.5°C*, V. Masson-Delmotte et al., Eds., World Meteorological Organization, 1–138.

Horion, S., A. Singleton, P. Barbosa, and J. Vogt, 2012: JRC experience on the development of drought information systems. Joint Research Centre Tech. Rep., 70 pp.

Kauffeldt, A., F. Wetterhall, F. Pappenberger, P. Salamon, and J. Thielen, 2016: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environ. Modell. Software*, **75**, 68–76, https://doi.org/10.1016/j.envsoft.2015.09.009.

Kitanidis, P. K., 1997: *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press, 249 pp.

Kumar, R., B. Livneh, and L. Samaniego, 2013a: 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.

——, L. Samaniego, and S. Attinger, 2013b: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resour. Res.*, **49**, 360–379, https://doi.org/10.1029/2012WR012195.

Lawrimore, J., R. R. Heim, M. Svoboda, V. Swail, and P. J. Englehart, 2002: Beginning a new era of drought monitoring across North America. *Bull. Amer. Meteor. Soc.*, **83**, 1191–1192, https://doi.org/10.1175/1520-0477-83.8.1191.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *J. Geophys. Res.*, **99**, 14 415–14 428, https://doi.org/10.1029/94JD00483.

Lourenço, T. C., R. Swart, H. Goosen, and R. Street, 2015: The rise of demand-driven climate services. *Nature*, **6**, 13–14, https://doi.org/10.1038/nclimate2836.

Marx, A., and Coauthors, 2018: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3°C. *Hydrol. Earth Syst. Sci.*, **22**, 1017–1032, https://doi.org/10.5194/hess-22-1017-2018.

National Research Council, 2001: *A Climate Services Vision: First Steps Toward the Future.* National Academies Press, 96 pp., https://doi.org/10.17226/10198.

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.

Prudhomme, C., and Coauthors, 2014: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Natl. Acad. Sci. USA*, **111**, 3262–3267, https://doi.org/10.1073/pnas.1222473110.

Rakovec, O., and Coauthors, 2016: Multiscale and multivariate evaluation of water fluxes and states over European river basins. *J. Hydrometeor.*, **17**, 287–307, https://doi.org/10.1175/JHM-D-15-0054.1.

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 M. Zink, 2013: Implications of parameter uncertainty on soil moisture drought analysis in Germany. *J. Hydrometeor.*, **14**, 47–68, https://doi.org/10.1175/JHM-D-12-075.1.

——, and Coauthors, 2017a: 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, 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, 2018: Anthropogenic warming exacerbates European soil moisture droughts. *Nat. Climate Change*, **4**, 421–426, https://doi.org/10.1038/s41558-018-0138-5.

Sessa, R., and H. Dolman, Eds., 2008: Terrestrial essential climate variables for climate change assessment, mitigation and adaptation. Food and Agriculture Organization of the United Nations Rep., 44 pp., www.fao.org/3/i0197e/i0197e00.htm.

Sheffield, J., and E. F. Wood, 2007: Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *J. Geophys. Res.*, **112**, D17115, https://doi.org/10.1029/2006JD008288.

——, and ——, 2008: Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J. Climate*, **21**, 432–458, https://doi.org/10.1175/2007JCLI1822.1.

——, and Coauthors, 2014: A drought monitoring and forecasting system for sub-Sahara African water resources and food security. *Bull. Amer. Meteor. Soc.*, **95**, 861–882, https://doi.org/10.1175/BAMS-D-12-00124.1.

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.*, **373A**, 20140454, https://doi.org/10.1098/rsta.2014.0454.

Thielen, J., J. Bartholmes, M. H. 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. Hydrometeor.*, **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 degree global warming. *Environ. Res. Lett.*, **13**, 014003, https://doi.org/10.1088/1748-9326/aa9e35.

——, M. Cuntz, M. Kelbling, R. Kumar, J. Mai, and L. Samaniego, 2019: The multiscale routing model mRM v1.0: Simple river routing at resolutions from 1 to 50 km. *Geosci. Model Dev.*, **12**, 2501–2521, https://doi.org/10.5194/gmd-12-2501-2019.

van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens, 2011: Global monthly water stress: 1. Water balance and water availability. *Water Resour. Res.*, **47**, W07517, https://doi.org/10.1029/2010WR009791.

Wanders, N., and Y. Wada, 2015: Human and climate impacts on the 21st century hydrological drought. *J. Hydrol.*, **526**, 208–220, https://doi.org/10.1016/j.jhydrol.2014.10.047.

——, S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood, 2019: Development and evaluation of a pan-European multi-model seasonal hydrological forecasting system. *J. Hydrometeor.*, **20**, 99–115, https://doi.org/10.1175/JHM-D-18-0040.1.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe, 2014: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. *Proc. Natl. Acad. Sci. USA*, **111**, 3228–3232, https://doi.org/10.1073/pnas.1312330110.

Yuan, X., J. K. Roundy, E. F. Wood, and J. Sheffield, 2015: Seasonal forecasting of global hydrologic extremes: System development and evaluation over GEWEX basins. *Bull. Amer. Meteor. Soc.*, **96**, 1895–1912, https://doi.org/10.1175/BAMS-D-14-00003.1.

Zink, M., L. Samaniego, R. Kumar, S. Thober, J. Mai, D. Schäfer, and A. Marx, 2016: The German drought monitor. *Environ. Res. Lett.*, **11**, 074002, https://doi.org/10.1088/1748-9326/11/7/074002.

——, R. Kumar, M. Cuntz, and L. Samaniego, 2017: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty. *Hydrol. Earth Syst. Sci.*, **21**, 1769–1790, https://doi.org/10.5194/hess-21-1769-2017.