



Multi-model hydrological seasonal forecasts for Europe in the C3S climate data store

Peter Berg¹, Denica Bozhinova¹, Yiheng Du¹, Yesheatesfa Hundecha¹, Tobias Lagander¹, Ursula McKnight¹, Lisanne Nauta², Spyridon Paparrizos², Ilias Pechlivanidis¹, and Wei Yang¹

¹Swedish Meteorological and Hydrological Institute, Folkborgsvägen 17, 601 76 Norrköping, Sweden

²Earth Systems and Global Change Group, Wageningen University and Research, the Netherlands

Correspondence: Peter Berg (peter.berg@smhi.se)

Abstract. Seasonal hydrological forecasts play a crucial role in water resource management and climate change adaptation, supporting a wide range of planning and decision-making activities. Access to tools, products, data and services is essential in helping societies to better prepare for and adapt to conditions in a changing climate. The demand for forecasts at monthly or longer timescales is increasing, and has resulted in several meteorological prediction systems being available. Notably, hydrological seasonal forecasts have shown high skill for longer lead-times than their meteorological counterparts, which makes them potentially useful for planning purposes for water resources management and risk reduction. Under the umbrella of the Copernicus Climate Change Services (C3S), an operational service has been established to provide operational hydrological seasonal forecasts for Europe. The service provides historical re-forecasts and operational monthly initialized forecasts of daily and monthly mean river discharge, runoff, snow water equivalent and soil wetness index up to a six-month horizon. A total of six hydrological models (E-HYPEcatch00, E-HYPEcatch01, E-HYPEcatch06, E-HYPEcatch09, E-HYPEgrid, and VIC-WUR) are running in the service, with forcing from two meteorological seasonal forecast systems (ECMWF-s51 and CMCC-s35). The output resolution is at catchment delineations for the E-HYPEcatch models, and at the 5 km drainage network of the Lisflood hydrological model used in the European Flood Awareness System (EFAS) for the two gridded models (E-HYPEgrid and VIC-WUR). This paper describes the production chain and the included model components, with evaluations of performance of the separate models, and the skill of the seasonal forecasts calculated for a 23 year long re-forecast period. All data are freely available from the Climate Data Store (CDS), with updated 6-month forecasts issued each month ([doi:10.24381/84dbf36e](https://doi.org/10.24381/84dbf36e) and [10.24381/ebac3954](https://doi.org/10.24381/ebac3954)).

1 Introduction

Hydrological predictions at the seasonal scale, typically defined for the range of one month to one year, are useful for planning of water resources and the mitigation of natural disasters. These forecasts are potentially very useful for applications in reservoir management (Anghileri et al., 2016; Wanders and Wood, 2016; Foster et al., 2018; Giuliani et al., 2020). White et al. (2022)



noted the potential usefulness of sub-seasonal to seasonal forecasts, while discussing also the great challenges in making these systems understood and actually put into use. Making seasonal forecasts available for users across disciplines is an important step forward in promoting the products' use and uptake (Cantone et al., 2023; Crochemore et al., 2024; Pechlivanidis et al., 2025).

Hydrological forecasting is challenging at the seasonal scale, but has an advantage over the seasonal meteorological forecasts as the initial conditions of the hydrological model have a substantial contribution to the skills in a longer forecast horizon (Crochemore et al., 2020). For example, the Ensemble Streamflow Prediction method (ESP), one of the traditional approaches to forecast hydrology, shows advantages over other methods, e.g. streamflow climatology, by considering the initial hydrological conditions (Harrigan et al., 2018; Crochemore et al., 2020; Girons-Lopez et al., 2021). The role of the initial hydrological conditions depends on the characteristics of the river catchment, e.g. catchment's physiography, the dominant processes (e.g. snow or precipitation), human influence, and its response time (Harrigan et al., 2018; Wanders et al., 2019; Pechlivanidis et al., 2020; Du et al., 2023). Further, added value of dynamical meteorological forecasts over historical assessments (e.g., ESP) has been demonstrated without clear indications of one method being superior to the other in general Crochemore et al. (2020); Musuuza et al. (2023); Girons-Lopez et al. (2025).

The C3S operational service for the water sector, developed by the main contractor at the Swedish Meteorological and Hydrological Institute, and sub-contractor Wageningen University and Research, was launched in 2022 with a multi-model hydrological seasonal forecast for Europe. This first version of the service consisted of a data set of re-forecast data set (Berg et al., 2022b), and monthly updated forecasts (Berg et al., 2022a), with the initial limitation to only monthly mean river discharge and the single ECMWF meteorological seasonal forecasting system SEAS5 (Johnson et al., 2019). The hydrological seasonal forecast data of that service has recently been updated with a new operational system with updated and extended data sets, which are presented in this paper. Whereas most openly available operational forecasts of this type are using a single models (e.g. available in the European and Global Flood Awareness Systems (known as EFAS and GloFAS respectively) (Arnal et al., 2018), the World-Wide HYPE (WWH) service (Arheimer et al., 2020; WWH, 2019)), the new C3S service contains multiple hydrological models at the European domain, and uses two sources of meteorological forecasts. The service therefore generates a unique data set for exploring the use of hydrological seasonal forecasts in a range of water dependent sectors, such as water allocation, flood management, ecological status, industrial water use and agriculture.

This paper presents features of the historical and re-forecast data from Service (2025b), including evaluation of the meteorological forcing and the individual hydrological models, as well as skill assessment of the hydrological seasonal re-forecasts. The paper ends with an outlook of possible future development paths.

2 Data and models

2.1 Observed river discharge data

Timeseries of river discharge from over 3000 stations were downloaded from the Global Runoff Data Centre (GRDC), the European Water Archive (EWA) and the Baltex Hydrological Data Centre (BHDC) databases, and filtered following a the data

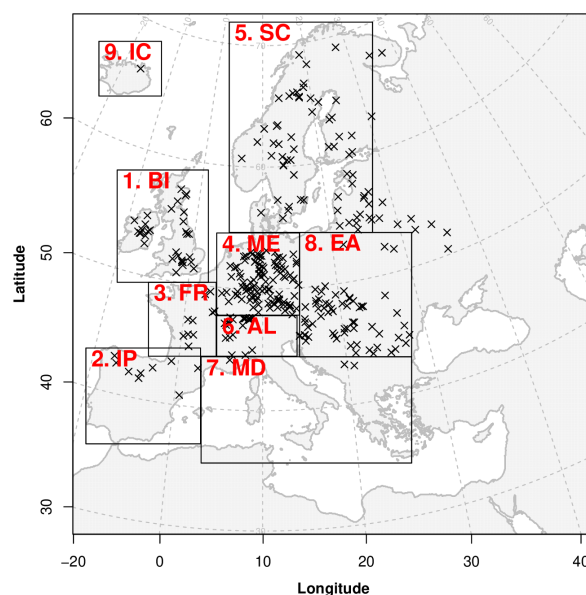


Figure 1. Map of the study region, with boxes for calculating the regional statistics, and marked locations (crosses) of the hydrological validation stations.

quality procedure (Donnelly et al., 2016). The time series were assessed regarding their geographical position and upstream area to ensure that they fit with the delineations of the E-HYPE model and the EFAS drainage network. Further, the time series were visually assessed for erroneous data, such as gaps and influence from upstream regulation. A large part of the stations was used for the calibration of the E-HYPE models, and 332 independent validation stations are retained for the current evaluation, see Fig. 1.

2.2 Meteorological forcing

The EFAS-Meteo data set (Ntegeka et al., 2013) was used as observational meteorological forcing for model calibration and forecast initialization. The version used is from May 2019, and recent releases are shared by JRC (Joint Research Centre) under the name EMO-5 (Thiemig et al., 2022). The dataset covers most of Europe, with a spatial resolution of 5 km a daily temporal resolution, while the temporal coverage is from 1990 until present. The ERA5 reanalysis (Hersbach et al., 2020) is providing the additional meteorological variables required by the VIC-WUR model, i.e. surface solar radiation, surface thermal radiation, surface wind components, surface pressure and dewpoint temperature.

The meteorological seasonal forecasts are taken from two different systems available from CDS (C3S, 2018): the ECMWF-s51 (also known as SEAS5 (Johnson et al., 2019)), and CMCC-s35 (also known as CMCC-SPS3.5 (Gualdi et al., 2020)). Both systems operate on the strategy of initializations at the first of the month, with 25 and 40 members in the re-forecast period for ECMWF-s51 and CMCC-s35, respectively. Both systems provide 51 members for the forecast period. The original



atmospheric resolution is originally of about 80 km for ECMWF-s51 and 50 km for CMCC-s35, but in the common format data set in CDS (C3S, 2018) they both provide remapped data at the 1.0 degree grid (approximately 100 km), which is the basis for all simulations presented here. While the data set has been in queue for review for publication in the CDS, the CMCCC-s35 data set has been replaced by a newer version (CMCC-s40), which is not included in the operational forecast production as described here.

2.2.1 Downscaling and bias adjustment

All forcing data are first bilinearly remapped to the 5 km resolution of the EFAS-Meteo dataset, which is the same as the drainage network map of the gridded hydrological models. Bias adjustment is performed for the main drivers of the hydrological models, i.e. daily mean precipitation and temperature. The bias adjustment method is a flavor of the empirical quantile mapping family, and is based on the core functionality of the MIDAS (Multi-scale bias AdjuStment) method (Berg et al., 2022c). It is therefore similar to the recently published data set of monthly resolution bias adjusted SEAS5 data by Weber et al. (2026). The empirical quantile mapping method essentially re-scales the values of the model time series, such that their cumulative density function resembles that of the reference data. For multi-member seasonal forecasts, it is essential to retain differences between the individual members. Therefore, the bias adjustment is calibrated on the joint cumulative density function of all 25 members of the re-forecast data set in the calibration period 1990–2016. This ensures that the distribution across all members remains close to the reference data set, but the individual members' deviations are retained for any given forecast. The joint cumulative density function of all data and model members is reshaped to an array of dimensions $(N_{days} * N_{years}, 25)$, where $N_{days} = 185$ and $N_{years} = 23$, and then averaged over the second dimension.

2.3 Hydrological models

2.3.1 E-HYPE

The process-based hydrological model HYPE (HYdrological Predictions for the Environment; (Lindström et al., 2010)), which simulates components of the catchment water cycle and water quality at a daily or hourly time step is employed. The model is semi-distributed, in which a river basin may be subdivided into multiple subcatchments, which are further subdivided into hydrological response units (HRUs) based on soil type and land use classes. It has conceptual routines for most of the major land surface and subsurface processes (e.g., snow/ice accumulation and melting, evapotranspiration, surface and macropore flow, soil moisture, discharge generation, groundwater fluctuation, aquifer recharge/discharge, irrigation, abstractions and routing through rivers, lakes and reservoirs, solid matter and dissolved nutrient pools) that are controlled by a number of parameters that are often linked to physiography to account for spatial variability and are estimated through calibration. The model parameters can generally be categorized into HRU-based and general parameters. The HRU-based parameters are either soil or land use type dependent. Unique parameter values are estimated for each soil or land use type and are applied throughout the model domain. The general parameters, such as the routing parameters, are subcatchment scale parameters and may be assigned constant values throughout the model domain or estimated separately for different parameter regions. They can also be estimated



using a regionalization approach as functions of catchment characteristics. A range of physiographic data sources were used in setting up the model (Donnelly et al., 2016). The meteorological forcing consists of daily mean surface air temperature and precipitation.

E-HYPE is here employed in three main setups: the original subcatchment based model, E-HYPEcatch00 (Hundecha et al., 2016), a multi-model version, whose members were derived through optimum resampling of parameters from their feasible ranges (Hundecha et al., 2020) with three members - E-HYPEcatch01, E-HYPEcatch06, and E-HYPEcatch09 - and the gridded model E-HYPEgrid at the 5 km EFAS drainage network. The models have somewhat different calibration strategies, as outlined in (Hundecha et al., 2016) for E-HYPEcatch00, and in (Hundecha et al., 2020) for the other catchment based models.

E-HYPEgrid was set up at a 5 km grid resolution over Europe with the same drainage network as in the EFAS system. The following data were obtained from the Lisflood model applied in EFAS: flow direction, mean elevation, fraction of water bodies, lake-mask, and stream channel length within each grid cell as well as the total upstream draining area to each grid cell, which is treated as a subcatchment as in the standard E-HYPEcatch00 model. Hydrological response units are defined within each grid cell as a combination of land use and soil type. Further, in order to exploit E-HYPE's capability of modeling outflows from lakes, big lakes are treated as subcatchments by their own, with grid cells along their peripheries split into two parts: a portion of the lake and the part that belongs to the land surface. Once the hydrological processes are computed, the computed variables are remapped to the original grid cells using the area weighting method. One exception is the treatment of lake areas since E-HYPE includes a conceptualized lake model with detailed natural and regulation lake processes. To accommodate for this, lake areas were added as additional computation grid cell during the E-HYPEgrid setup and calibration phase using European and global lake and reservoir databases (Hundecha et al., 2016). Large lakes in E-HYPEgrid cover one or several EFAS grid cells completely, and surrounding grid cells partially, while smaller lakes and elongated reservoirs behind river dams cover grid cells partially only. Since computation of the outflow from the lake is only performed for the discharge point of the lake, this point is representing the whole lake and any partial land computation cell is replaced by the lake cell, following the reasoning that the water flows through the lake unit. Remaining lake-only grid cells will have NaN in the output grids. To deliver E-HYPEgrid results on the original EFAS grid, the lake values replace the land values when the lake is only partial, following the reasoning that the water flows through the lake unit. Model parameters for E-HYPEgrid were derived from the regional relationships established in the calibration of E-HYPEcatch00, with only minor adjustments to account for the different drainage network.

2.3.2 VIC-WUR

The Variable Infiltration Capacity (VIC) model (Liang et al., 1994) is a semi-distributed macroscale hydrological model and forms the fundamentals of the VIC-WUR model (version 2.0.0) which is an extension of the VIC-5 model, including additional modules to account for anthropogenic influences (Hamman et al., 2018; Droppers et al., 2020). Model parameters are taken from the global calibration performed by Nijssen et al. (2001). The infiltration parameter, thickness of the second soil layer and soil hydraulic parameters has been calibrated for each climate zone to match the total annual flow volume and shape of the mean monthly hydrograph. Each parameter is transferred to the global basins based on the respective climate zone.



Here, the model has been setup in natural flow conditions and water balance mode with three thermal nodes, three soil layers and no snow bands. Sub-daily time steps of 6 h are used to solve the energy and water balance for each cell individually, i.e. there is no water or energy exchange between cells. Cell heterogeneity of soil and canopy characteristics (Lawrence and Chase, 2007) are extracted from a gridded parameter file and used by the model. There is surface runoff whenever precipitation occurs under saturated top soil conditions, while precipitation under unsaturated conditions infiltrates into the soil layers. The upper two soil layers interact with the canopy layer and provide water for transpiration, whereas the third and deepest soil layer is assumed to be outside the root zone and accounts for subsurface runoff (baseflow). VIC-WUR uses an internal routing module (Lohmann et al., 1996, 1998) to simulate discharge based on the unit hydrographs of both stream inflow and outflow. The unit hydrographs are computed from the EFAS flow direction, channel length, mean elevation (slope) and the SCS (US Soil Conservation Service) dimensionless unit hydrograph using a flow velocity of 1.5 ms^{-1} and diffusion value of 2000 ms^{-2} .

2.4 Hydrological historical simulation, initialization and seasonal (re)forecasting

The hydrological models were initially forced with the EFAS-Meteo for precipitation and temperature, at daily resolution for the E-HYPE models, and in the case of VIC-WUR disaggregated to 6 h resolution and additionally with ERA5 outputs for solar radiation, thermal radiation, wind speed, surface pressure and humidity. The historical forcing is used to generate: (1) the historical reference simulation for assessing the historical model performance, and (2) the initial model states on the forecast issue date (that is the 1st day of each month) for generating the (re-)forecasts.

The (re-)forecasts are based on bias adjusted daily precipitation and temperature of seasonal meteorological forecasts from ECMWF-s51 and CMCC-s35 as input to the hydrological models, using 25 members from each system for the re-forecast simulations, and 51 members for the forecasts. VIC-WUR is in addition forced by solar radiation, thermal radiation, wind speed, surface pressure and humidity forcing from the forecasting systems. Vapor pressure was calculated by substituting dewpoint temperature into the Clausius–Clapeyron equation. Both EFAS-Meteo and ERA5 daily data were disaggregated to 6-hour intervals, while ERA5 and the forecast data (extracted at 6 h resolution) were bilinearly remapped to the 5 km model grid. The mean elevation, 2 m temperature and mean sea level pressure are used to calculate the surface pressure forcing using the barometric formula. The daily bias adjusted values for temperature are disaggregated according to the 6-hour values by adding the difference between the adjusted and original daily mean to each time step and for each day separately. All other required forcing variables are computed as described previously.

2.5 Technical description of the dataset

The production chain is set up such that the input and output calculations are performed in a harmonized way for all hydrological ensemble members, both for the pre- and post-processing steps. Each hydrological model is contained inside a block where each model provider provides scripts for reading the common NetCDF input files, performing model simulations, and finally producing output data files in a common format for post-processing in the common work chain.

The data set is organized in three main parts: historical, re-forecast, and forecast. All three parts are accessed through the same CDS data page (Service, 2025b), and contain the following main data:



- Historical hydrological simulations for the period December 1992 until present.
- Six-month long hydrological re-forecasts for the two meteorological systems, including bias adjusted meteorological data, monthly initialized for the period January 1993 to December 2016.
- Six-month long hydrological forecasts for the two meteorological systems, including bias adjusted meteorological data, monthly initialized for the period January 2017 (ECMWF-s51) or October 2022 (CMCC-s35) until present.
- Skill scores calculated for the re-forecasts at a monthly time step and for each hydrological model combined with each meteorological forcing data.

Each of the hydrological models produce outputs for the variables river discharge, runoff water equivalent, soil wetness index, and snow water equivalent. The bias adjusted and downscaled precipitation and temperature forcing data are available for the forecast and re-forecast data sets. The time stamps in the NetCDF files of the (re)forecasts are organized by three metadata information: (i) "time" - the time stamp of the initialization date of the forecast, (ii) "step" - the number of hours into the forecast, and (iii) "valid_time" - the actual time of the step into the forecast. The coordinates of the gridded models are available in the metadata of each files as longitude and latitude variables. For the catchment delineated models, the defining polygons of each catchment are available from Isberg (2017), where the value of the 'id' variable maps to the 'SUBID' variable of the shapefile.

3 Methods

3.1 Historical hydrological model evaluation

Three standard model efficiency measures are employed: the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), the Kling-Gupta Efficiency (KGE; Gupta et al., 2009), and the percentage model bias (Pbias). In addition, the model performance is assessed using hydrological signatures focusing on the model's capability to reproduce extremes of daily flow. This is assessed by evaluating how well the model captures the 5th and 95th percentile of the long-term daily flow. For reference, NSE=0 is equivalent to using the mean flow as a predictor, while a value of KGE=-0.41 has a similar interpretation (Knoben et al., 2019). These limits are specifically marked in the maps of the result section.

Because the calibration strategies vary between the hydrological models, we identify a common set of validation stations that are independent from the different calibrations. The VIC-WUR model was globally calibrated and has no dependency on the current river discharge data. However, the various E-HYPE models have used somewhat different sets of calibration stations, leaving in total a set of 323 independent validation stations across Europe. These validation stations are used for the main model assessment presented here, using data from the time period 1991–2016.



3.2 Forecast performance and skill assessment

200 The hydrological seasonal re-forecasts are evaluated with respect to their performance against the historical simulation (also known as “pseudo-observations” or perfect forecast) per initialization month and lead month (Pechlivanidis et al., 2020). Evaluation against the pseudo-observations allows analysis of the forecasting skill over all modelled locations and does not require historical observed river discharge records. Nevertheless, such an evaluation eventually leads to a theoretical, and not to the actual, accuracy, which is subject to the model’s capability of representing reality. An evaluation in the modelled reality
 205 is independent from the hydrological model’s imperfection. The evaluation of the re-forecasts is performed on monthly means, by calendar definition, for the 1993–2016 period. For the forecast evaluation, a set of 25 members are used for each system ECMWF-s51 and CMCC-s35 (for CMCC-s35, the members 0–24 are used), and calculations are carried out for each variable, initialization month and all six lead months for the period 1993–2016. The same forecast performance and skill scores are calculated for both the hydrological and the meteorological variables.

210 Performance, which measures how closely the forecast aligns with the reference simulation, was evaluated using four metrics: the fair Continuous Rank Probability Score (CRPS), the tercile fair Ranked Probability Score (tercile RPS), and the fair Brier Score for lower and upper terciles (BSbn and BSun respectively). By using four different metrics, the aim is to assess the models’ predictability with regard to overall performance, anomalies to the terciles, and low and high extremes. The forecast skill assessment is performed by comparing the performances between the two systems (ECMWF-s51-based and CMCC-s35-based forecasts) to the benchmark (climatology), using the the four scores above with their corresponding skill metrics: CRPSS, RPSS, BSSbn and BSSun. All skill scores range between $-\infty$ and 1, with 1 being the optimum. All metrics are calculated using two R-packages: easyVerification MeteoSwiss (2023), and SpecsVerification Siegert (2020). The functions used are: FairRpss, FairCpss, and FairBrierSs(ens, ens.ref, obs, tau = NA). Resulting scores are written to netcdf files and provided
 215 in the re-forecast data entry (Service, 2025b).

220 4 Results

4.1 Evaluation of the bias adjustment of meteorological forecasts

Bias adjustment is applied to precipitation and temperature forecasts for each of the meteorological data sets. The bias differ between the two datasets (not shown), but the bias adjustment performs similarly in both cases. We exemplify the performance assessment of the bias adjustment method with some samples of the evaluation of the ECMWF-s51 re-forecasts. The adjustment
 225 step consists also of a statistical downscaling from the model grid of 1 degree to the EFAS-Meteo grid of 5 km. For temperature, this is implicitly adjusting for differences in elevation between the model and the reference data as a constant “lapse rate” offset. Figure 2 and Fig. 3 show the performance of the bias adjustment method for the ensemble forecast for precipitation and temperature, respectively, and for different regions of Europe as presented in Fig. 1. Precipitation (Fig. 2) is generally underestimated for the extreme statistics, especially during the summer months in several of the regions, such as IP, FR, ME,
 230 AL, MD and EA. The mean statistic is both over- and underestimated for different regions and time of the year. After bias

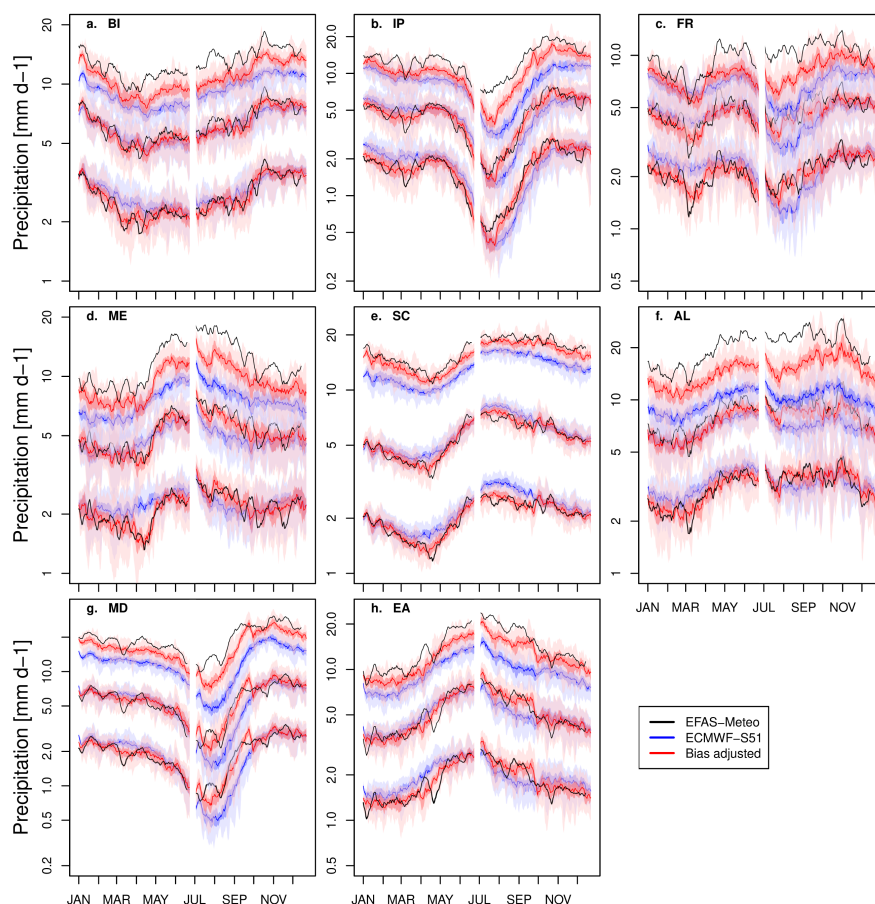


Figure 2. Regionally averaged climatological precipitation statistics of the original and the bias adjusted ECMWF-s51 seasonal forecasts, initialized in January and July. The presented statistics are the mean, the 95th, and the 99th percentile, in order from bottom to the top of each panel. The model data is presented as the ensemble median (line), the envelope between the 25th and 75th percentiles (dark shading) and the envelope of the min and max values (light shading). A rolling mean of 12 d has been applied to the EFAS-Meteo reference data (only one possible realisation), and 6 d to the multi-member model data. A white section separates the two forecasts.

adjustment, the mean and P95 statistics are generally in good agreement with observations, and the ensemble range is well retained. The more extreme P99 statistic is improved after bias adjustment, but still somewhat underestimated in most regions. A consequence of the application of the bias adjustment at the monthly calendar scale can be seen clearly for the mediterranean region, where there is a sharp transition from September to October. This causes a somewhat unphysical step as the re-forecasts go from underestimation of the precipitation intensity to almost no bias from October for the mean climate.

Temperature bias (Fig. 3) is sometimes quite large, especially for the 1st percentile with bias of several degrees Celsius for many regions in summer. However, these deviations are partly due to a general offset in regions of strong orography, and is to a large degree due to lapse rate related offsets between the model and observational data sets. The bias adjustment is correcting

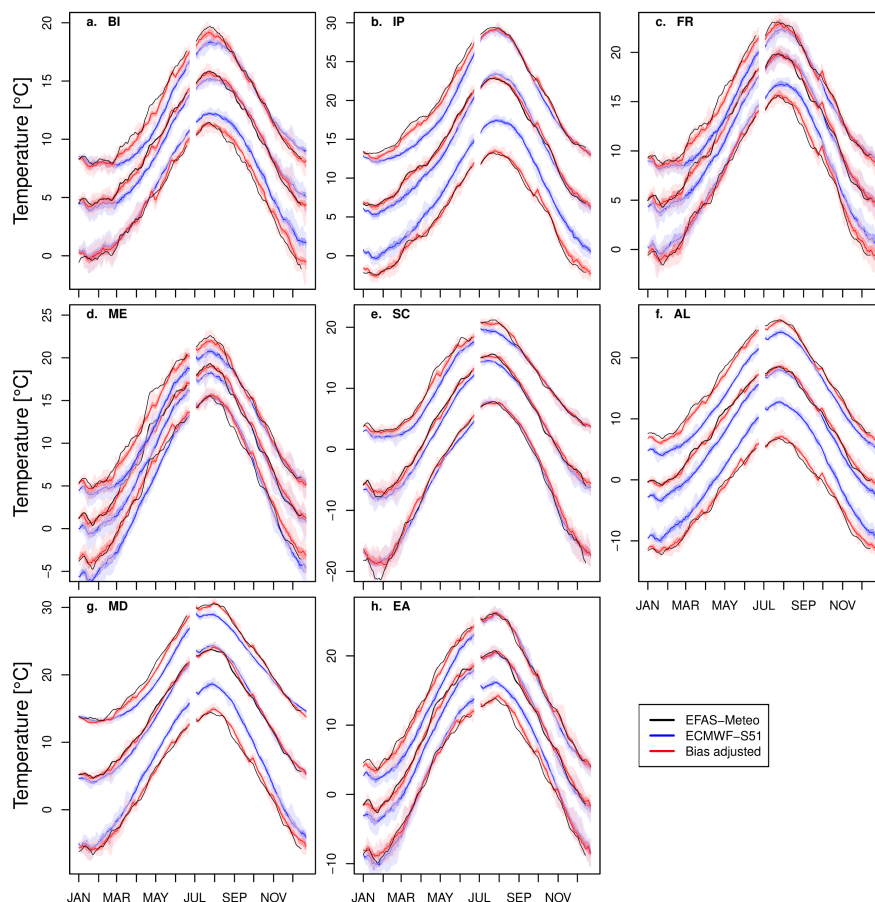


Figure 3. Same as Fig. 2, but for temperature and the statistics 1st percentile, mean, and 99th percentile, from bottom to top of each panel.

for such offsets along with other model bias, and the adjusted data agrees well with observations. The range of the ensemble
 240 members is retained well also for temperature after bias adjustment.

We next examine how the skill of the meteorological data is affected by the bias adjustment (Fig. 4). To exemplify this, we
 present CRPSS for monthly mean temperature and precipitation forecasts, and its difference without and with bias adjustment.
 Overall, we observe that the skill of the original data deteriorates rather fast with time, and there is essentially no skill after the
 first lead month (not shown). The bias adjusted forecasts have the same skill reduction as a function of lead time, but have higher
 245 skill scores than the original data, concluding that the bias adjustment has a positive impact on the skill of the meteorological
 forecasts. In particular, the differences are significant in the Mediterranean and overall southern Europe, including also Norway
 and Iceland. These are important observations highlighting the need of bias adjustment of the seasonal meteorological forecasts
 prior to be used in impact (i.e. water resources) analyses.

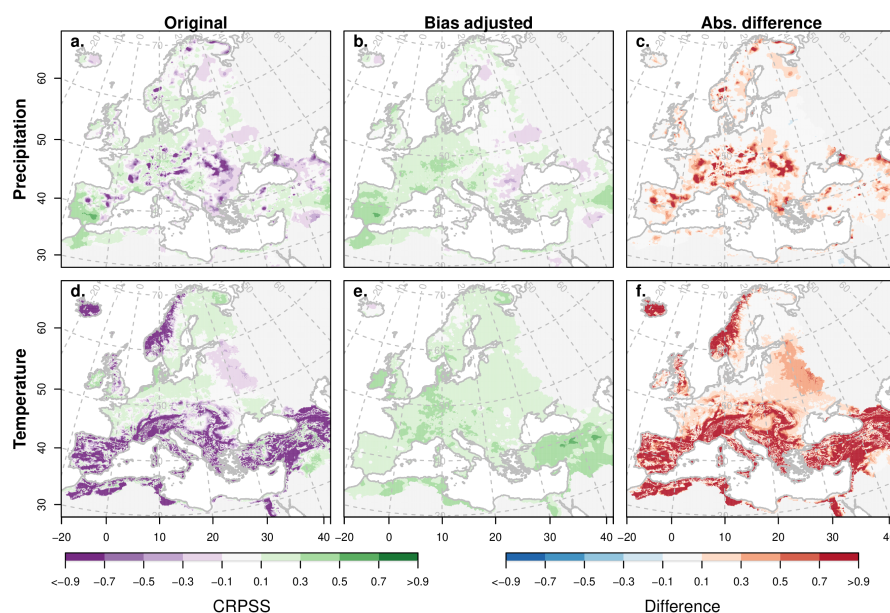


Figure 4. CRPSS values for the first lead month of the January ECMWF-s51 forecasts for precipitation (a–c) and temperature (d–f). The left panels (a, d) show the skill in the original meteorological forecasts, the middle panels (b, e) show the forecast skill after bias adjustment, and the right panels (c, f) show their difference in skill (original minus bias adjusted forecasts).

4.2 Historical hydrological performance assessment

250 The hydrological models are validated for river discharge on independent validation stations in the historical simulations between 1991 and 2016. Figure 5 provides an overview of the scores across all validation stations of each hydrological model. All models show fairly good validation scores of KGE and NSE, with median values around 0.5. E-HYPEcatch00 and E-HYPEgrid, which share the same basic regionally calibrated parameters, perform the best, while the other E-HYPEcatch models have slightly lower KGE and NSE scores. The VIC-WUR has yet a bit lower performance for KGE, but more notably,

255 it has lower NSE values, especially encompassing negative values within the 50% central range of the stations. The percentage bias is better centred around zero for the models E-HYPEcatch01, E-HYPEcatch06, and E-HYPEcatch09, which were calibrated more directly with the EFAS-Meteo data set. The E-HYPEcatch00, and thereby E-HYPEgrid, have only been slightly adjusted from an earlier calibration using the HydroGFD data set (Berg et al., 2021), which explains some of the larger bias. VIC-WUR is again standing out with a large bias, which probably relates to the very different calibration strategy.

260 Figures 6, 7, and 8 show results for each validation station. There are clear patterns among all models regarding the stations with a low performance, i.e. below zero for NSE and below -0.41 for KGE. There are strong similarities in the performance for all E-HYPE models, with some clustering of performance for the E-HYPEcatch00 and E-HYPEgrid, versus the other E-HYPEcatch models. The stations in eastern Europe tend to have lower scores for all models, else there is no clear geographical patterns for the performance in KGE and NSE. In contrast, Pbias (Fig. 8) show clear patterns of underestimation in the Nordic

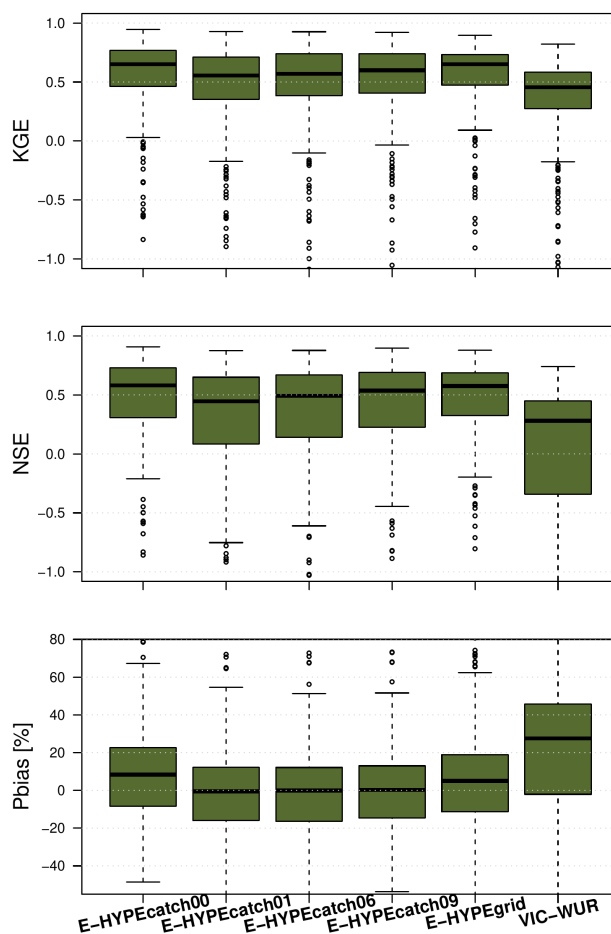


Figure 5. Historical hydrological model performance assessment for river discharge in terms of NSE, KGE and Pbias across all 352 stations presented in box-whisker plots.

265 countries and the British Isles, while there is overestimation in central Europe for all models. The E-HYPEcatch01,06,09 models show lower bias, mainly in central Europe, affecting the overall score in Fig. 5.

The low flows, here represented in Fig. 9 by the 5th percentile of the daily mean river discharge over the evaluation period, is well aligned with observations for VIC-WUR, E-HYPE00 and E-HYPEgrid. The E-HYPEcatch01,06,09 models are rather consistently underestimating the low flows, and have some larger outliers compared to the other models.

270 The high flows, here represented in Fig. 10 by the 95th percentile of the daily mean river discharge over the evaluation period, are well represented only by VIC-WUR. Among the E-HYPE models, only E-HYPEgrid manages to simulate the higher flows, and all the catchment-based models are rather strongly underestimating high flows. It is possible that the underestimation by

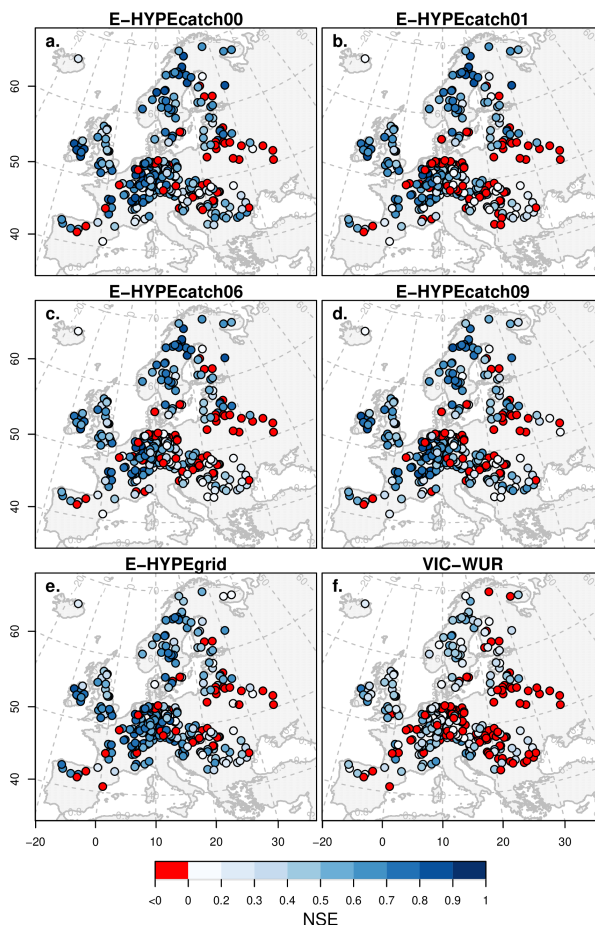


Figure 6. Historical performance assessment for river discharge in terms of NSE for the single hydrological models. Results are presented for the 352 stations that are independent of the calibration stations.

the catchment delineated models (operating at 215 km^2 on average) is due to the process in which several 5 km grid boxes of the forcing data are averaged to produce the model forcing for a given catchment. This will inherently reduce the highest precipitation intensities, which has an influence of also reducing the peak discharges.

Although there are no reliable observational reference, we present also a comparison between the different models' simulation of soil wetness index and snow water equivalent, as this is useful for explaining some of the skill scores of the seasonal forecasts presented in Section 4.3. Because the performance is similar between all E-HYPE models, we present only results from E-HYPEgrid, which shares the same map projection as VIC-WUR.

Figure 11 presents the seasonal mean soil wetness index for the two gridded models. There are clear differences between the models, with VIC-WUR being much wetter in central Europe and the Mediterranean, with mean values above 0.6 and rather smooth spatial gradients. E-HYPEgrid has a pattern of soil wetness index more closely related to the orography, and clearly

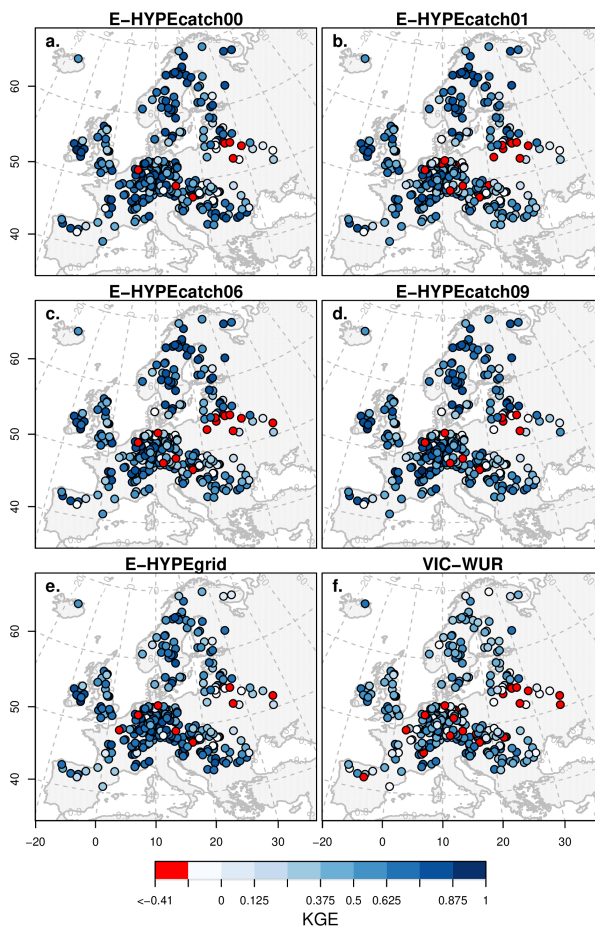


Figure 7. Historical performance assessment for river discharge in terms of KGE for the single hydrological models. Results are presented for the 352 stations that are independent of the calibration stations.

drier conditions below 0.5 for the Iberian Peninsula for most of the year, in autumn extending with values below 0.5 to central and southern Europe. The Nordic countries are wetter in E-HYPEgrid compared to VIC-WUR.

285 The simulated snow water equivalent is presented for the period March to June, when the peak snow amounts typically occur, before the onset of the melting period, in Fig. 12. The VIC-WUR models show greater snow depths in Scandinavia and north-east Europe in March and April, at over twice the amounts observed e.g. in Sweden (Wern, 2015), although the timing of the snowmelt is similar for that region. E-HYPEgrid is also somewhat overestimating the snow amounts compared to observations in Wern (2015). It seems that the VIC-WUR model melts snow at a faster rate than the E-HYPEgrid model. This
 290 can for example be seen in the east of the domain and the Carpathians in April and May, where the snow amounts disappear one or two months earlier in VIC-WUR compared to E-HYPEgrid. The two models also differ regarding the peak snow amounts in the Alps, where VIC-WUR peaks in March, while E-HYPEgrid peaks in April. Note that the coarse monthly resolution of

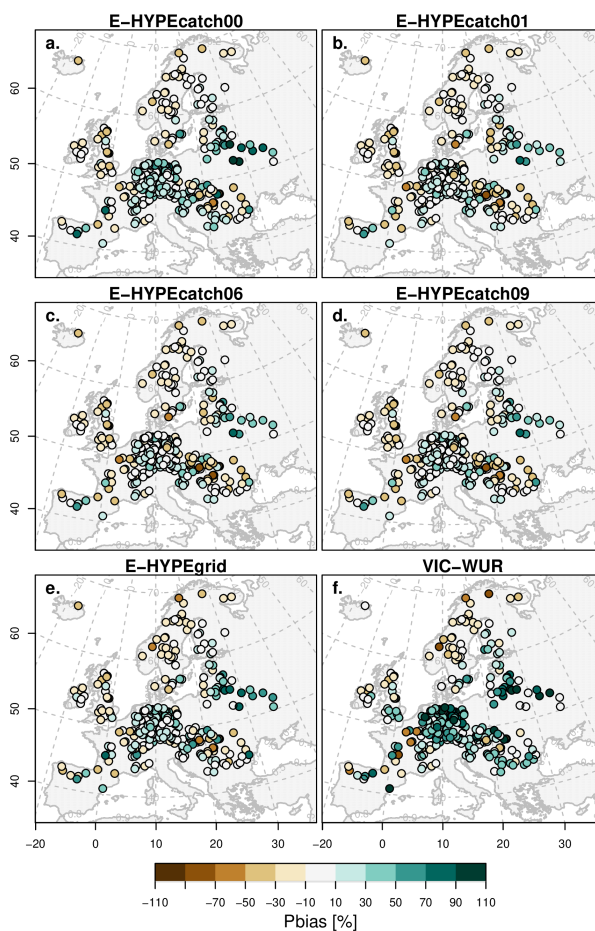


Figure 8. Historical performance assessment for river discharge in terms of Pbias for the single hydrological models. Results are presented for the 352 stations that are independent of the calibration stations.

the maps prevents a more detailed assessment, and more information can be assessed in a temporally finer resolution analysis using the daily outputs. The snow models differ between VIC-WUR and E-HYPE, which can of course explain the differences.
 295 However, differences may also arise from how the models are forced with meteorological data, with E-HYPE forced with daily mean temperature and precipitation, and VIC-WUR with 6 h values and disaggregated daily radiation fields.

4.3 Skill evaluation

Seasonal forecast skills are calculated based on the monthly mean of each hydrological variable, and evaluated as a function of initialization month and lead time. The same analysis is carried out for each hydrological model, with each meteorological
 300 forecast system. Moreover, the hydro-meteorological variables are assessed separately using the four skill scores presented in Sect. 3.2. This leads to a large set of skill values, which are all available as NetCDF files in the re-forecast dataset (Service,

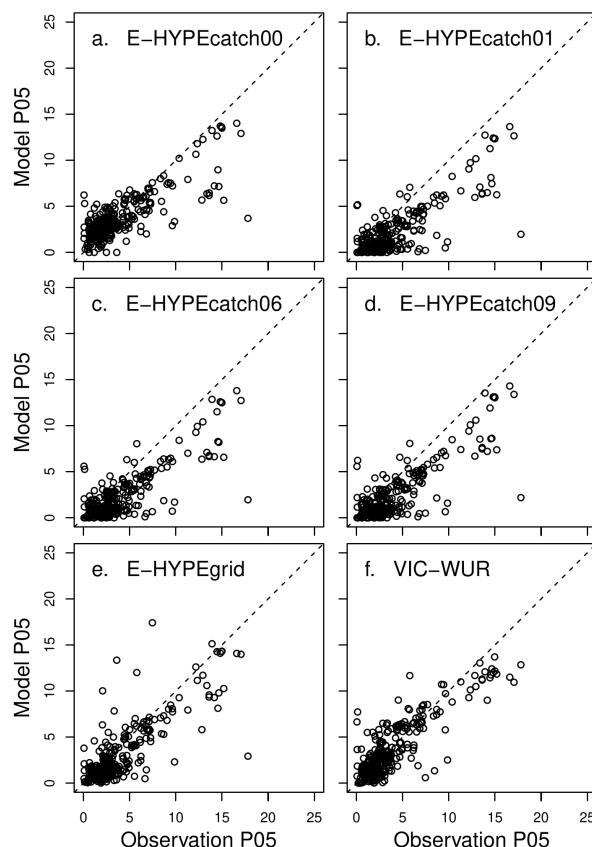


Figure 9. Scatter plots of modelled versus observations of the 5th percentile of the daily mean river discharge (low flows) at each station.

2025b). Here, a selection of the forecast skill assessments is presented to provide an overview of the results. We present line plots for the whole domain to summarize the results (see Fig. 13 and Fig. 16), which are calculated by taking the median skill for all catchments or grid points. Note that the geographical balance between regular grid points and the irregular catchment delineations can lead to some systematic deviations in the median statistics between the model setups.

First, it is noted that the four different skill scores give a similar pattern of the skill and its deterioration with lead time, although the absolute magnitude can differ to some extent (not shown). The different skill scores represent certain aspects of the forecast attributes, here only CRPSS is presented focusing on the overall forecast performance. Further, the skills for the two meteorological forecast systems are very similar, although the ECMWF-s51 forecasts tend to have slightly higher median skill scores as an average of the entire domain.

Figure 13 shows the median CRPSS for river discharge for each initialization month and lead month. The E-HYPEcatch models perform similarly, with only slight variations in the scores among different models. E-HYPEgrid performs closely to the other E-HYPEcatch models, which is expected as it is based on the calibration of E-HYPEcatch00. Note also that the median

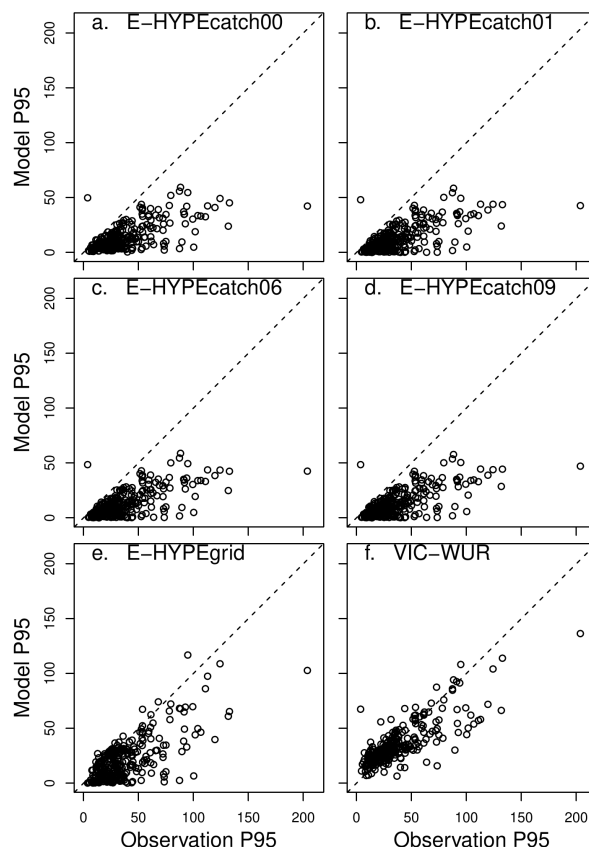


Figure 10. Scatter plots of modelled versus observations of the 95th percentile of the daily mean river discharge (high flows) at each station.

skill may not be fully compatible between the gridded and catchment based models due to the varying size of the catchments.

315 VIC-WUR stands out compared to the E-HYPE models, with generally higher initial skills for river discharge for the winter half-year (October to March), and comparable or slightly lower values during the summer half-year (April to September). However, the E-HYPE models show greater persistency in the skill than VIC-WUR in the long lead months indicating that the former models can represent processes of longer memory (hence achieving a slower river response) (Crochemore et al., 2020; Hurkmans et al., 2023). Figure 14 shows the spatial distribution of river discharge forecast skill from E-HYPEgrid and

320 VIC-WUR initialized in January for the first two lead months. For continental Europe, the skills are similar, albeit generally slightly higher for VIC-WUR than for E-HYPEgrid. However, the main feature of these maps is the near perfect skill for VIC-WUR in large parts of Scandinavia and the Alps in the first two lead months (Fig. 14b and d). The forecasts initialized in July, presented in Fig. 15, show that the pattern of high skill observed for January's initialization does not remain, pointing toward snow processes with frozen conditions during winter in the VIC-WUR model; see also the deviations in snow patterns

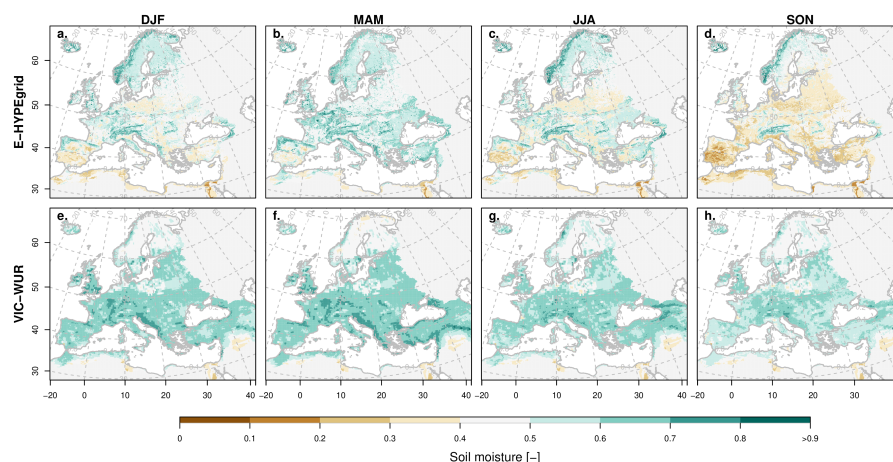


Figure 11. Seasonal climatology (1993–2016) of soil wetness index for the E-HYPEgrid and VIC-WUR models. Note that the catchment based models have similar climatology to that of the E-HYPEgrid model.

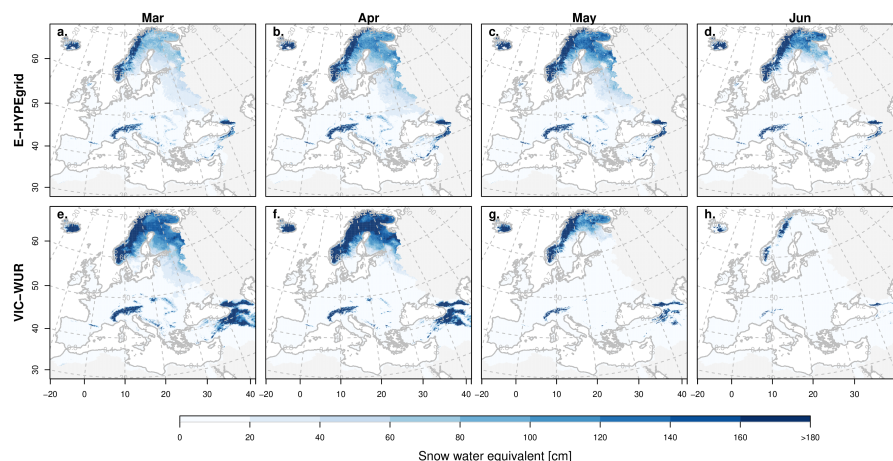


Figure 12. Monthly climatology (1993–2016) of snow water equivalent for the E-HYPEgrid and VIC-WUR models during the main European peak snow period from March to June, leading up to the melting period. The catchment based models have similar climatology to that of the E-HYPEgrid model. Climatology is calculated as the mean of the last month of each season, i.e. the accumulated snowpack.

presented in Fig. 12. Further, runoff shows very similar skill scores as the river discharge, and therefore the results are not shown here.

The skill scores for soil wetness index, shown in the left half of Fig. 16 is again largely consistent across the different E-HYPE models. Again, VIC-WUR stands out with higher skills than the E-HYPE models in the winter months, however lower skills are observed in spring and early summer. Note that for VIC-WUR, the median skill even turns negative in the long lead months for model initializations from December to July.

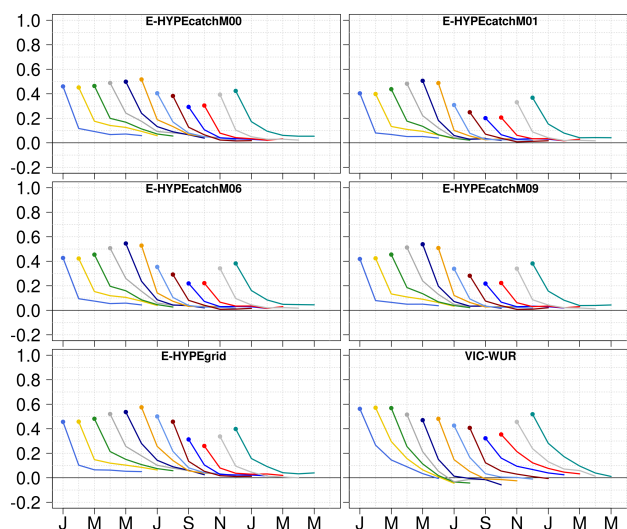


Figure 13. Median CRPSS values for river discharge based on all catchment or grid points in the domain. The hydrological models are forced with the bias adjusted CMCC-s35 forecasts. The initialization month is marked with a dot, while subsequent months with a line.

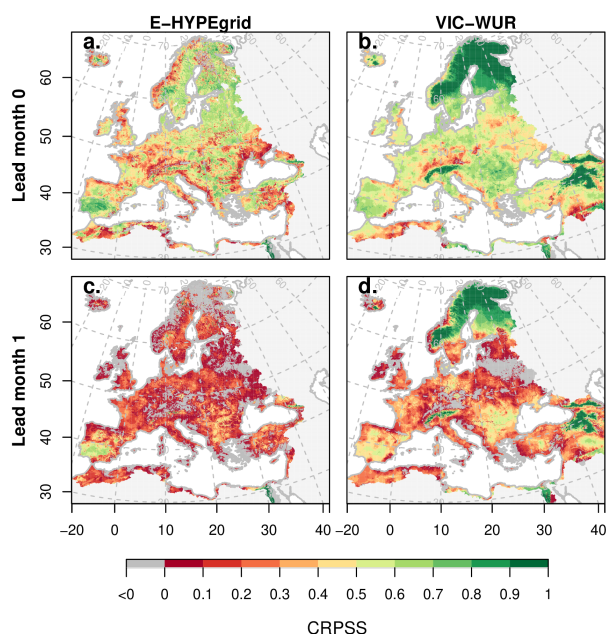


Figure 14. CRPSS values for river discharge for the first two lead months. E-HYPEgrid and VIC-WUR are forced with the bias adjusted CMCC-s25 forecasts initialized in January.

When considering the skill scores for snow water equivalent, shown in the right half of Fig. 16, the issue with snow in VIC-WUR becomes more apparent, with very low - mostly zero - skills. The different E-HYPE models have similar skill also for this variable.

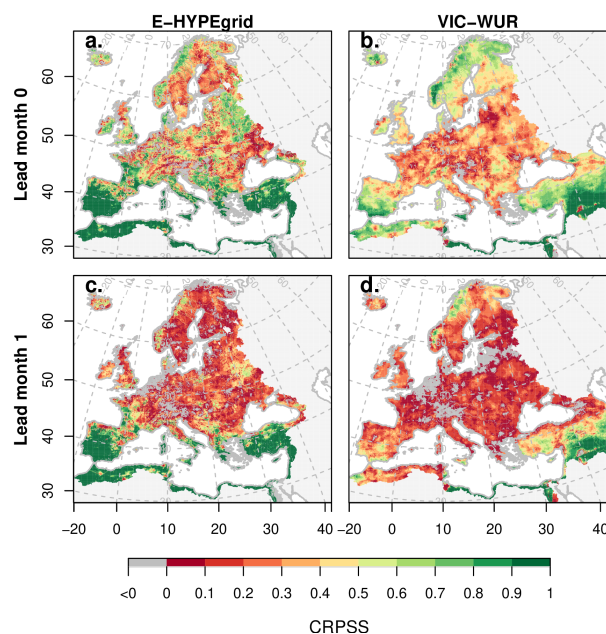


Figure 15. CRPSS values for river discharge for the first two lead months. E-HYPEgrid and VIC-WUR are forced with the bias adjusted CMCC-s25 forecasts initialized in July.

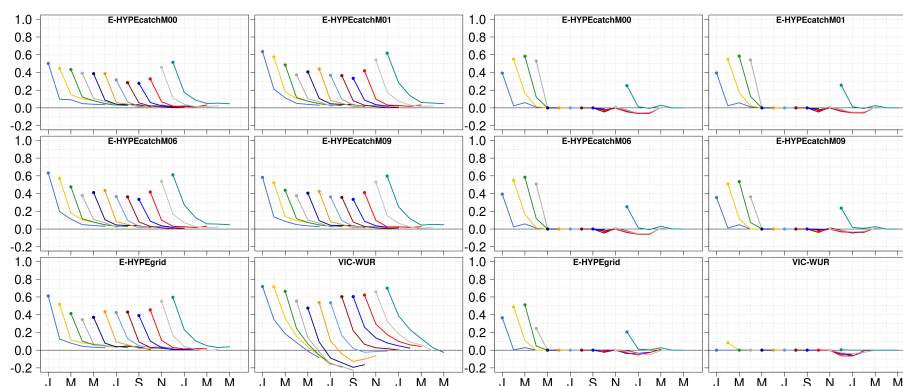


Figure 16. Median CRPSS values for soil wetness index (left) and snow water equivalent (right) based on all catchment or grid points in the domain. The hydrological models are forced with the bias adjusted CMCC-s35 forecasts. The initialization month is marked with a dot, while subsequent months with a line.

5 Discussion

335 The availability of seasonal re-forecasts has been important for providing adequate data for both understanding the prediction system's performance but also for the forecast bias adjustment, which as it has been shown here, is a central process prior to forcing them in impact models. However, given that every seasonal prediction system comes with its set of process hypothesis



and spatial resolution, and hence inherited biases, any change in the system causes downstream effects, including the need for a new bias adjustment. For the currently presented hydrological seasonal forecasts, introducing a new meteorological forecast system means that the bias adjustment parameters, all hydrological re-forecasts and resulting climatology, and skill calculations need to be performed all over again. This is a substantial effort, which also produces large data amounts; the current presented data set amounts to over 50 TB in total. Therefore, a hydrological operational system is vulnerable to frequent updates to the forcing data, and time delays for updating the production. In a new contract with C3S (C3S2_411_lot1), a larger set of meteorological systems are planned to include redundancy for discontinued systems, as well as a more streamlined and automated bias adjustment calibration procedure to update the operational system - to make a more robust and dependable multi-model forecast service. The contract runs until 2028, when updated forecast systems for both Europe and global domains will be operational.

The significant impact of initial conditions of the hydrological models puts extra emphasis in getting a consistent and high quality initial hydrological state. This is currently performed by a continuous simulation with observational meteorological inputs through the EFAS-Meteo data set. This gridded product is, however, sensitive to the station inputs that vary with time. Therefore, the initial state may differ in quality for different time periods. We believe that the service could greatly benefit from improved and more consistent initialization data, using e.g. regional reanalysis such as CERRA (Copernicus European Regional ReAnalysis; Schimanke et al., 2021) together with station observations and other Earth observation products. There is also potential in the assimilation of hydrological variables for hydrological seasonal forecasting applications (Musuuza et al., 2023).

While advancements in modelling and data processing have undoubtedly enhanced the quality of hydrological seasonal forecasts, it is equally crucial to emphasize the importance of making these forecasts accessible and useful to users. Simply improving the accuracy and quality of existing datasets is not sufficient; the true value of seasonal forecasts lies in their utility for informed decision-making and planning processes. Therefore, efforts should be directed towards ensuring that the provided information is not only accurate but also presented in a format that is understandable and actionable for stakeholders. This includes considerations such as user-friendly interfaces, clear communication of uncertainties, and tailored products that address user specific needs. By prioritizing the usability of data, its impact can be maximized in helping societies better prepare and adapt to changes.

6 Conclusions

A new data set for multi-model hydrological seasonal forecasts is presented based on an ensemble of meteorological seasonal forecast systems and hydrological models. In particular, an ensemble of six hydrological models is used (E-HYPEcatch00, E-HYPEcatch01, E-HYPEcatch06, and E-HYPEcatch09 at catchment resolution and E-HYPEgrid and VIC-WUR at 5 km grid resolution), each with forecasts for two meteorological prediction systems (ECMWF-s51 and CMCC-s35), for which the bias adjusted precipitation and temperature are included in the dataset. The data set is disseminated as three main components:



370 (i) historical simulations from 1993 to present, (ii) seasonal re-forecasts for the period 1993–2016, and (iii) seasonal forecasts from 2016 (ECMWF-s51) and 2022 (CMCC-s35) until present.

This paper describes the performance of the historical hydrological simulations, the bias adjustment of the meteorological forecasts, and the skill of the hydrological seasonal forecasts for river discharge, runoff, soil wetness index, and snow water equivalent.

375 The main conclusions are:

- The historical hydrological performance, assessed using the NSE, KGE, and percentage bias metrics, and the high and low river discharge signatures, shows generally good performance of the continental scale hydrological models, with a spread of results across the ensemble depending on the studied performance metric.
- The bias of the meteorological seasonal forecasts is efficiently reduced, while retaining the ensemble spread. The skill of the meteorological seasonal forecasts improves after the bias adjustment.
- The skill of the hydrological seasonal forecasts is similar for the hydrological models of the ensemble for most variables and skill scores. The main exceptions are for soil wetness index and snow water equivalent, for which there is a discrepancy between the VIC-WUR model and the different E-HYPE models.

380

7 Data availability

385 The data set is available from the Copernicus Climate Change Service's Climate Data Store, see Service (2025b), under the following doi:10.24381/84dbf36e . Operational forecasts with the same model setups (not presented here) are available from Service (2025a).

Author contributions. PB was the main project lead, including the development and execution of the production chain, the bias adjustment method and its evaluation, quality assurance of the data in general, as well as the design of the data sets. TL developed the code of the production chain and supervised the production. LN implemented the code for, and the evaluation of, the VIC-WUR model. YH performed the evaluation of the E-HYPE model ensemble. DB performed the skill evaluation of the bias adjustment. YD performed the skill evaluation of the seasonal forecasts. UMK was project managing and contributed with quality assurance of all documentation. SP was project managing and contributing to the VIC-WUR modelling efforts. WY performed the set up and evaluation of the bias adjustment and statistical downscaling. IP contributed with the skill definitions and interpretation and review of the manuscript. PB prepared the manuscript with contributions from all co-authors.

395

Competing interests. The authors declare that they have no conflicts of interest.



Acknowledgements. The production of the presented data set was funded by the Copernicus Climate Service (C3S) in contract C3S2_410_lot2. We would like to extend our gratitude to our C3S contract officer Shaun Harrigan for support and discussions, as well as the technical staff at ECMWF for support in setting up and optimising the production chain.



400 References

- Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., and Lettenmaier, D. P.: Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, *Water Res. Res.*, 52, 4209–4225, <https://doi.org/10.1002/2015WR017864>, 2016.
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, *Hydrology and Earth System Sciences*, 24, 535–559, <https://doi.org/10.5194/hess-24-535-2020>, 2020.
- 405 Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, *Hydrol. Earth Syst. Sci.*, 22, 2057–2072, <https://doi.org/10.5194/hess-22-2057-2018>, 2018.
- Berg, P., Almén, F., and Bozhinova, D.: HydroGFD3.0 (Hydrological Global Forcing Data): a 25 km global precipitation and temperature data set updated in near-real time, *Earth Syst. Sci. Data*, 13, 1531–1545, <https://doi.org/10.5194/essd-13-1531-2021>, 2021.
- 410 Berg, P., Bartosova, A., Capell, R., Hundecha, Y., Isberg, K., Musuuza, J., Näslund, V., Pechlivanidis, I., Rosberg, J., Photiadou, C., Berggren, L., Fahlesson, T., Nauta, L., Ludwig, F., Franssen, W., van der Linden, E., Wetterhall, F., Arnal, L., Barnard, C., Krzeminski, B., Ferrario, I., Mazzetti, C., and Prudhomme, C.: Multi-model seasonal forecasts of river discharge for Europe from January 2021 to present, Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.52f45864>, (Accessed on 10-03-2024), 2022a.
- 415 Berg, P., Bartosova, A., Capell, R., Hundecha, Y., Isberg, K., Musuuza, J., Näslund, V., Pechlivanidis, I., Rosberg, J., Photiadou, C., Berggren, L., Fahlesson, T., Nauta, L., Ludwig, F., Franssen, W., van der Linden, E., Wetterhall, F., Arnal, L., Barnard, C., Krzeminski, B., Ferrario, I., Mazzetti, C., and Prudhomme, C.: Multi-model seasonal reforecasts of river discharge for Europe, Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.13c18212>, (Accessed on 10-03-2024), 2022b.
- 420 Berg, P., Bosshard, T., Yang, W., and Zimmermann, K.: MidASv0.2.1 – Multi-scale bias Adjustment, *Geosci. Model Dev.*, 15, 6165–6180, <https://doi.org/10.5194/gmd-15-6165-2022>, 2022c.
- C3S: Seasonal forecast daily and subdaily data on single levels., Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.181d637e>, (Accessed on 10-03-2024), 2018.
- Cantone, C., Ivars Grape, H., El Habash, S., and Pechlivanidis, I. G.: A co-generation success story: Improving drinking water management through hydro-climate services, *Climate Services*, 31, 100399, <https://doi.org/10.1016/j.cliser.2023.100399>, 2023.
- 425 Crochemore, L., Ramos, M., and Pechlivanidis, I. G.: Can Continental Models Convey Useful Seasonal Hydrologic Information at the Catchment Scale?, *Water Resour. Res.*, 56, <https://doi.org/10.1029/2019WR025700>, 2020.
- Crochemore, L., Materia, S., Delpiazzo, E., Bagli, S., Borrelli, A., Bosello, F., Contreras, E., Valle, F. D., Gualdi, S., Herrero, J., Larosa, F., Lopez, R., Luzzi, V., Mazzoli, P., Montani, A., Moreno, I., Pavan, V., Pechlivanidis, I., Tomei, F., Villani, G., Photiadou, C., Polo, M. J., and Mysiak, J.: A Framework for Joint Verification and Evaluation of Seasonal Climate Services across Socioeconomic Sectors, *Bulletin of the American Meteorological Society*, 105, E1218–E1236, <https://doi.org/10.1175/BAMS-D-23-0026.1>, 2024.
- 430 Donnelly, C., Andersson, J. C., and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, *Hydrolo. Sci. J.*, 61, 255–273, <https://doi.org/10.1080/02626667.2015.1027710>, 2016.
- Droppers, B., Franssen, W. H. P., van Vliet, M. T. H., Nijssen, B., and Ludwig, F.: Simulating human impacts on global water resources using VIC-5, *Geosci. Model Dev.*, 13, 5029–5052, <https://doi.org/10.5194/gmd-13-5029-2020>, 2020.
- 435



- Du, Y., Clemenzi, I., and Pechlivanidis, I. G.: Hydrological regimes explain the seasonal predictability of streamflow extremes, *Environ. Res. Lett.*, 18, 094 060, <https://doi.org/10.1088/1748-9326/acf678>, 2023.
- Foster, K., Uvo, C. B., and Olsson, J.: The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers, *Hydrol. Earth Syst. Sci.*, 22, 2953–2970, <https://doi.org/10.5194/hess-22-2953-2018>, 2018.
- 440 Girons-Lopez, M., Crochemore, L., and Pechlivanidis, I. G.: Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden, *Hydrol. Earth Syst. Sci.*, 25, <https://doi.org/10.5194/hess-25-1189-2021>, 2021.
- Girons-Lopez, M., Bosshard, T., Crochemore, L., and Pechlivanidis, I.: Leveraging GCM-based forecasts for enhanced seasonal streamflow prediction in diverse hydrological regimes, *Journal of Hydrology*, 650, 132 504, <https://doi.org/10.1016/j.jhydrol.2024.132504>, 2025.
- Giuliani, M., Crochemore, L., Pechlivanidis, I., and Castelletti, A.: From skill to value: isolating the influence of end user behavior on
 445 seasonal forecast assessment, *Hydrol. Earth Syst. Sci.*, 24, 5891–5902, <https://doi.org/10.5194/hess-24-5891-2020>, 2020.
- Gualdi, S., Sanna, A., Borrelli, A., Cantelli, A., del Mar Chaves Montero, M., and Tibaldi, S.: The new CMCC Operational Seasonal Prediction System SPS3.5, Tech. Rep. CMCC Tech. Note RP0288, Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), <https://doi.org/10.25424/CMCC/SPS3.5>, 2020.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria:
 450 Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., and Mao, Y.: The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility, *Geosci. Model Dev.*, 11, 3481–3496, <https://doi.org/10.5194/gmd-11-3481-2018>, 2018.
- Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, *Hydrol.*
 455 *Earth Syst. Sci.*, 22, 2023–2039, <https://doi.org/10.5194/hess-22-2023-2018>, 2018.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut,
 460 J.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Hundecha, Y., Arheimer, B., Donnelly, C., and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, *J. Hydrol. Reg. Stud.*, 6, 90–111, <https://doi.org/10.1016/j.ejrh.2016.04.002>, 2016.
- Hundecha, Y., Arheimer, B., Berg, P., Capell, R., Musuuza, J., Pechlivanidis, I., and Photiadou, C.: Effect of model calibration strategy on climate projections of hydrological indicators at a continental scale, *Clim. Change*, 163, 1287–1306, <https://doi.org/10.1007/s10584-020-02874-4>, 2020.
- 465 Hurkmans, R. T. W. L., van den Hurk, B., Schmeits, M., Wetterhall, F., and Pechlivanidis, I. G.: Seasonal Streamflow Forecasting for Fresh Water Reservoir Management in the Netherlands: An Assessment of Multiple Prediction Systems, *Journal of Hydrometeorology*, 24, 1275–1290, <https://doi.org/10.1175/JHM-D-22-0107.1>, 2023.
- Isberg, K.: *EHYPE3_subbasins.zip*, Tech. rep., Zenodo, <https://doi.org/https://doi.org/10.5281/zenodo.581451>, 2017.
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremmer, D., Weisheimer, A.,
 470 Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, *Geosci. Model Dev.*, 12, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>, 2019.



- Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. and Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.
- Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), *J. Geophys. Res. Biogeosci.*, 112, <https://doi.org/10.1029/2006JG000168>, 2007.
- 475 Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res. Atm.*, 99, 14 415–14 428, <https://doi.org/10.1029/94JD00483>, 1994.
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, *Hydrol. Res.*, 41, 295–319, <https://doi.org/10.2166/nh.2010.007>, 2010.
- 480 Lohmann, D., Nolte-Holube, R., and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parametrization schemes, *Tellus A*, 48, 708–721, <https://doi.org/10.1034/j.1600-0870.1996.t01-3-00009.x>, 1996.
- Lohmann, D., Raschke, E., Nijssen, B., and Lettenmaier, D. P.: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, *Hydrol. Sci. J.*, 43, 131–141, <https://doi.org/10.1080/02626669809492107>, 1998.
- MeteoSwiss: easyVerification: Ensemble Forecast Verification for Large Data Sets, <https://CRAN.R-project.org/package=easyVerification>, r package version 0.4.5, 2023.
- 485 Musuuza, J. L., Crochemore, L., and Pechlivanidis, I. G.: Evaluation of Earth Observations and In Situ Data Assimilation for Seasonal Hydrological Forecasting, *Water Resources Research*, 59, <https://doi.org/10.1029/2022WR033655>, 2023.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- 490 Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., and Wood, E. F.: Predicting the Discharge of Global Rivers, *Journal of Climate*, 14, 3307 – 3323, [https://doi.org/10.1175/1520-0442\(2001\)014<3307:PTDOGR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2), 2001.
- Ntegeka, V., Salamon, P., Gomes, G., Sint, H., Lorini, V., Zambrano-Bigiarini, M., and Thielen, J.: EFAS-Meteo: A European daily high-resolution gridded meteorological data set for 1990–2011, Tech. rep., Report Eur, 2013.
- Pechlivanidis, I. G., Crochemore, L., Rosberg, J., and Bosshard, T.: What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts?, *Water Resour. Res.*, 56, <https://doi.org/10.1029/2019WR026987>, 2020.
- 495 Pechlivanidis, I. G., Du, Y., Bennett, J., Boucher, M.-A., Chang, A. Y. Y., Crochemore, L., Dasgupta, A., Baldassarre, G. D., Luterbacher, J., Pappenberger, F., Ramos, M.-H., Slater, L., Uhlenbrook, S., Wetterhall, F., Wood, A., Lavado-Casimiro, W., Yoshimura, K., Imhoff, R., van Oevelen, P. J., Cantone, C., Cattoën, C., Pimentel, R., and Werner, M.: Enhancing Research-to-Operations in Hydrological Forecasting: Innovations across Scales and Horizons, *Bulletin of the American Meteorological Society*, 106, E894 – E919, [https://doi.org/10.1175/BAMS-](https://doi.org/10.1175/BAMS-D-24-0322.1)
- 500 D-24-0322.1, 2025.
- Schimanke, S., Ridal, M., Le Moigne, P., Berggren, L., Undén, P., Randriamampianina, R., Andrea, U., Bazile, E., Bertelsen, A., Brousseau, P., Dahlgren, P., Edvinsson, L., El Said, A., Glinton, M., Hopsch, S., Isaksson, L., Mladek, R., Olsson, E., Verrelle, A., and Wang, Z.: CERRA sub-daily regional reanalysis data for Europe on model levels from 1984 to present, Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/cds.7c27fd20>, (Accessed on 10-03-2024), 2021.
- 505 Service, C. C. C.: Multi-model hydrological seasonal forecasts and historical simulations for Europe, Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/ebac3954>, (Accessed on 27-07-2026), 2025a.
- Service, C. C. C.: Multi-model hydrological seasonal reforecasts for Europe, Tech. rep., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), <https://doi.org/10.24381/84dbf36e>, (Accessed on 27-07-2026), 2025b.



- Siegert, S.: SpecsVerification: Forecast Verification Routines for Ensemble Forecasts of Weather and Climate, <https://CRAN.R-project.org/package=SpecsVerification>, r package version 0.5-3, 2020.
- Thiemig, V., Gomes, G. N., Skøien, J. O., Ziese, M., Rauthe-Schöch, A., Rustemeier, E., Rehfeldt, K., Walawender, J. P., Kolbe, C., Pichon, D., Schweim, C., and Salamon, P.: EMO-5: a high-resolution multi-variable gridded meteorological dataset for Europe, *Earth Syst. Sci. Data*, 14, 3249–3272, <https://doi.org/10.5194/essd-14-3249-2022>, 2022.
- Wanders, N. and Wood, E. F.: Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations, *Environ. Res. Lett.*, 11, 094 007, <https://doi.org/10.1088/1748-9326/11/9/094007>, 2016.
- Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., and Wood, E. F.: Development and Evaluation of a Pan-European Multimodel Seasonal Hydrological Forecasting System, *J. Hydrometeorol.*, 20, 99–115, <https://doi.org/10.1175/JHM-D-18-0040.1>, 2019.
- Weber, J. N., Lorenz, C., Schober, T. C., Wiegels, R., and Kunstmann, H.: SEAS5-BCSD: a bias-corrected and downscaled global seasonal forecast reference dataset for 1981–2024, *Earth System Science Data*, 18, 5259–5282, <https://doi.org/10.5194/essd-18-5259-2026>, 2026.
- Wern, L.: Snödjup i Sverige 1904/05 – 2013/14, Tech. rep., Swedish Meteorological and Hydrological Institute, meteorologi Nr 158, 2015.
- White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., Balogun, A. A., Bertram, D., Bluhm, S., Brayshaw, D. J., Browell, J., Büeler, D., Charlton-Perez, A., Chourio, X., Christel, I., Coelho, C. A. S., DeFlorio, M. J., Monache, L. D., Giuseppe, F. D., García-Solórzano, A. M., Gibson, P. B., Goddard, L., Romero, C. G., Graham, R. J., Graham, R. M., Grams, C. M., Halford, A., Huang, W. T. K., Jensen, K., Kilavi, M., Lawal, K. A., Lee, R. W., MacLeod, D., Manrique-Suñén, A., Martins, E. S. P. R., Maxwell, C. J., Merryfield, W. J., Ángel G. Muñoz, Olaniyan, E., Otieno, G., Oyedepo, J. A., Palma, L., Pechlivanidis, I. G., Pons, D., Ralph, F. M., Reis, D. S., Remenyi, T. A., Risbey, J. S., Robertson, D. J. C., Robertson, A. W., Smith, S., Soret, A., Sun, T., Todd, M. C., Tozer, C. R., Vasconcelos, F. C., Vigo, I., Waliser, D. E., Wetterhall, F., and Wilson, R. G.: Advances in the Application and Utility of Subseasonal-to-Seasonal Predictions, *B. Am. Meteorol. Soc.*, 103, E1448–E1472, <https://doi.org/10.1175/BAMS-D-20-0224.1>, 2022.
- WWH: HYPEweb - Seasonal Forecasts Global, <https://hypeweb.smhi.se/seasonal-forecasts/>, accessed on 10-03-2024, 2019.