Evaluating E-OBS forcing data for large-sample hydrology using model performance diagnostics

Franziska Clerc-Schwarzenbach[*1], Thiago V. M. do Nascimento[*1,2]

[1]Department of Geography, University of Zurich, Zurich, 8057, Switzerland
[2]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, 8600, Switzerland

[*]These authors contributed equally to this work.

Correspondence to: Thiago V. M. do Nascimento (thiago.nascimento@eawag.ch)

Abstract. For large-sample hydrological studies over large spatial domains, large-scale meteorological forcing data are often desired. For large-sample studies across Europe, the EStreams dataset and catalogue satisfies this demand. In EStreams, the meteorological time series are obtained from the Ensemble Observation (E-OBS) product which is available for all of Europe. Due to the large spatial extent of this dataset, limitations and regional variations of data quality have to be expected when the dataset is compared to smaller-scale datasets, e.g., national level. In this study, we compare the meteorological time series included for 2682 catchments in EStreams to eight smaller datasets (mostly CAMELS datasets). We assess how the different meteorological data impact the performance of a bucket-type hydrological model. For most catchments, the precipitation amounts derived from E-OBS are lower than the ones from the CAMELS datasets, while the temperature and the potential evapotranspiration values are higher. Model performances tend to be lower when the E-OBS data are used than when the CAMELS datasets are used for calibration. Exceptions arise when the station density in the E-OBS data is high. This study provides the first assessment of the E-OBS data at a continental scale for hydrological applications and shows that, despite some limitations, the dataset offers a reasonable basis for large-sample hydrological modelling across Europe.

1 Introduction

Driven by their enormous value for hydrological modelling studies, large-sample hydrology (LSH) datasets have developed at a rapid pace in the past decades, and the development continues to gain momentum: Since 2017, more than a dozen "CAMELS" datasets were released or are being developed (Addor et al., 2017; Alvarez-Garreton et al., 2018; Bushra et al., 2025; Chagas et al., 2020; Coxon et al., 2020a; Delaigue et al., 2025a; Fowler et al., 2021; Höge et al., 2023; Jimenez et al., 2025; Liu et al., 2025; Loritz et al., 2024; Mangukiya et al., 2025; Nijzink et al., 2025; Teutschbein, 2024a). Other animals entered the LSH stage as well: Llamas (Helgason and Nijssen, 2024; Klingler et al., 2021), a goat (*cabra* in Portuguese; Almagro et al., 2021), and a bull (Senent-Aparicio et al., 2024b).

In the past years, efforts also went into the creation of more overarching products, i.e., datasets covering not only one country or region. The Caravan dataset (Kratzert et al., 2023) combined the streamflow data from thousands of catchments in already published open source LSH datasets with meteorological time series and catchment attributes from the global ERA5-Land reanalysis (Muñoz-Sabater et al., 2021). Caravan is growing further and has become a quasi-global dataset (Färber et al., 2024). For Europe, a dynamic dataset and a catalogue that provides detailed guidance for retrieving streamflow data from national providers were introduced in EStreams (https://www.estreams.eawag.ch) by do Nascimento et al. (2024).

Although these collections of large-sample datasets are valuable resources, the combination of catchments distributed across different regions and especially across different countries in one dataset almost always goes hand in hand with difficulties in providing high-quality streamflow and forcing data, due to the lower availability of high-quality data for larger spatial extents, while smaller datasets typically benefit from more thorough quality control. Furthermore, data processing choices (e.g., gap filling, interpolation) are more often required at large scales and might introduce added uncertainty in the outcomes (McMillan et al., 2018).

In an earlier study, Clerc-Schwarzenbach et al. (2024) showed that the globally available meteorological data obtained from ERA5-Land (Muñoz-Sabater et al., 2021) in the Caravan dataset (Kratzert et al., 2023) led to a consistently lower hydrological model performance for catchments in the US, Brazil, and Great Britain, compared to when the meteorological forcing data from the corresponding CAMELS datasets (Addor et al., 2017; Chagas et al., 2020; Coxon et al., 2020a) were used. This demonstrates the importance of promoting awareness of potential data quality losses when it comes to large-scale meteorological datasets.

Similar to the ERA5-Land data in Caravan, the meteorological data were also obtained from a large-scale dataset in EStreams (do Nascimento et al., 2024). For EStreams, the data were obtained from the European Ensemble Observation (E-OBS) product (Cornes et al., 2018). After the publication of EStreams, questions on the quality of the meteorological forcing data from E-OBS arose in the LSH community. Recent studies have evaluated the accuracy of the E-OBS precipitation product against reference datasets and meteorological stations in some parts of Europe, including Greece (Mavromatis and Voulanas, 2021), the central Alps, eastern Europe and Scandinavia (Bandhauer et al., 2022). These evaluations indicated that the quality of the E-OBS precipitation data, when compared to data from high-resolution datasets focusing on a smaller area, is higher in regions with a high density of E-OBS stations, such as in central Europe, while the reanalysis product ERA5 (Hersbach et al., 2020) partly outperformed E-OBS in regions with a sparse station network (Bandhauer et al., 2022). Yet, evaluations of the E-OBS data over a larger extent, and specifically for hydrological modelling, remain unexplored.

To be able to inform the users of EStreams (and of the E-OBS data in general) about the effects of the harmonized meteorological data on hydrological applications, a comparison to the meteorological data contained in different national and regional datasets (i.e., CAMELS datasets and similar products) is required.

For this study, we used 2682 catchments from eight European countries and assessed the hydrological efficacy of the meteorological data provided in EStreams (obtained from E-OBS). We did so by comparing the meteorological forcing data from E-OBS to the analogous data contained for the same catchments in national or regional datasets, namely in CAMELS-DK (Liu et al., 2025) for Denmark, CAMELS-FR (Delaigue et al., 2025a) for France, CAMELS-DE (Loritz et al., 2024) for Germany, CAMELS-GB (Coxon et al., 2020a) for Great Britain, the BULL Database (Senent-Aparicio et al., 2024b) for Spain, CAMELS-SE (Teutschbein, 2024a) for Sweden, and CAMELS-CH (Höge et al., 2023) for Switzerland. In addition, we also included catchments from Czechia, with data from the not yet published CAMELS-CZ dataset (Jenicek et al., 2024). The meteorological data in the smaller datasets stem from sources that were created specifically for the respective country.

The methodology used in this work is based on the one presented by Clerc-Schwarzenbach et al. (2024). Following their approach, we did not only compare the meteorological data itself, but also the model performances that were achieved with the different meteorological forcings (but same streamflow data) when calibrating the bucket-type HBV (Hydrologiska Byråns Vattenbalansavdelning) model (Bergström, 1992, 1995; Seibert and Vis, 2012). This allowed us to assess the overall hydrological efficacy of the forcing data.

The reasoning behind this approach is that hydrological models are not only useful for simulating streamflow but can also be used as diagnostic tools to evaluate the efficacy of the input data (Beck et al., 2017; Tarek et al., 2020). The rationale is that, although hydrological models are inherently imperfect representations of reality, systematic differences in their performance when driven by different datasets are unlikely to be random. As noted by Linsley (1982, p. 13), "if the data are too poor for the use of a good simulation model, they are also inadequate for any other model." Building on this idea, our study uses the HBV model as a means to assess the hydrological reliability of different meteorological forcing data across Europe. Therefore, we assume that if a model consistently performs better with one dataset than with the other, this difference likely reflects a closer alignment of the corresponding meteorological data with the actual processes in the catchment.

2 Data and Methods

2.1 Subset of catchments

We conducted this study for 2682 catchments that are available in the EStreams dataset and catalogue and in one of the following datasets: CAMELS-CZ, CAMELS-DK, CAMELS-FR, CAMELS-DE, CAMELS-GB, BULL, CAMELS-SE, CAMELS-CH, for simplicity's sake referred to as the "CAMELS datasets" throughout the remainder of the paper (Table 1). These catchments fulfilled the following cascade of criteria (with the number of catchments still included after each step in brackets, see also Supporting Information, Figure S1):

- Located in a country with access to a CAMELS dataset at the time of data preparation, i.e., Austria, Czechia, Denmark, France, Germany, Great Britain, Iceland, Spain, Sweden, or Switzerland [12 019]
- High-quality catchment delineation in EStreams, as described by do Nascimento et al. (2024) [10 434]
- Catchment area (obtained from EStreams) below 2000 km^2 [9115]
- No redundancy among the EStreams catchments (the catchment with the longer streamflow time series was kept) [8909]
- Availability of at least 90 % of the E_{pot} data between October 1990 and September 2015 [8846]

- Availability of the catchment in one of the CAMELS datasets [3557]
- CAMELS forcing data coming from a smaller-scale dataset (catchments from LamaH-CE (Austria) excluded as meteorological forcings are from ERA5-Land) [3097]
- Availability of at least 90 % of the streamflow data (in the CAMELS dataset) between October 1995 and September 2015 [3097]
- Maximum five lakes upstream (obtained from EStreams) [2841]
- Normalized upstream capacity of reservoirs, calculated using Eq. 9 from Salwey et al. (2023), and derived from the EStreams dataset, smaller or equal than 0.2 [2741]
- Runoff ratio (based on the precipitation data from the CAMELS dataset) between October 1995 and September 2015 below 1.1 [2741]
- Runoff ratio (based on the precipitation data in EStreams) between October 1995 and September 2015 below 1.1 [2682]

We excluded catchments with an area of more than 2000 km^2 as a bucket-type hydrological model is not the most suitable choice for larger catchments.

Unlike the other national datasets, the LamaH-CE dataset for Austria uses ERA5-Land as its meteorological forcing. As the comparison of E-OBS data to globally available data is a different question than the comparison to smaller-scale (national) products, we excluded the Austrian dataset from the analyses. Moreover, a previous study already highlighted several limitations of ERA5-Land as forcing for hydrological models (Clerc-Schwarzenbach et al., 2024).

To minimize the inclusion of catchments potentially affected by human regulation (e.g., reservoirs or diversions), we applied two attribute-based criteria. The maximal number of lakes was chosen arbitrarily, with the goal of excluding highly regulated catchments. The second criterion, the normalized upstream capacity, was calculated following Salwey et al. (2023), and a threshold of 0.2 was adopted based on their findings. While these filters may exclude some basins that are only weakly influenced by regulation, we preferred a conservative approach. Moreover, we selected these two criteria because the relevant information (number of lakes and upstream capacity) is consistently available across Europe in the EStreams dataset, allowing for a uniform filtering procedure even where metadata on human influence are not provided in the national CAMELS datasets.

Following a similar reasoning, we excluded catchments with runoff ratios greater than 1.1, as natural streamflow rarely exceeds precipitation by such large margins in unregulated basins. Such cases likely reflect data inconsistencies or strong anthropogenic influence (e.g., diversions or regulation).

Finally, to make sure that the streamflow data (obtained from the CAMELS datasets) were reasonable, we checked that the average streamflow was not unrealistically high (i.e., not exceeding 10 mm d^{-1} as this may indicate issues with the data) which was the case for all 2682 catchments.

Table 1: Overview of catchments and data sources used in this study.

Country	Number of catchments included in this study	CAMELS dataset	Publication	Dataset
Czechia	244	CAMELS-CZ	*unpublished*	*unpublished*
Denmark	120	CAMELS-DK	Liu et al. (2025)	Koch et al. (2025)
France	496	CAMELS-FR	Delaigue et al. (2025a)	Delaigue et al. (2025b)
Germany	986	CAMELS-DE	Loritz et al. (2024)	Dolich et al. (2024)
Great Britain	489	CAMELS-GB	Coxon et al. (2020a)	Coxon et al. (2020b)
Spain	211	BULL	Senent-Aparicio et al. (2024b)	Senent-Aparicio et al., (2024a)
Sweden	3	CAMELS-SE	Teutschbein (2024a)	Teutschbein (2024b)
Switzerland	133	CAMELS-CH	Höge et al. (2023)	Höge et al. (2025)

2.2 Meteorological data

For the data comparison and the modelling experiments, we investigated and used daily precipitation, E_{pot}, and temperature data from the EStreams dataset and from the different CAMELS datasets (Table 1). We used the latest released version of EStreams (version 1.4), for which precipitation and temperature data were obtained from the E-OBS ensemble mean product with a spatial resolution of 0.1° in both latitude and longitude (do Nascimento et al., 2025). E-OBS provides a pan-European observational dataset of surface climate variables that is derived by statistical interpolation of in-situ measurements, collected from national data providers (Cornes et al., 2018). Potential evapotranspiration (E_{pot}) time series in EStreams were calculated with the Hargreaves formula (Hargreaves and Samani, 1982), using the E-OBS temperature data and catchment elevation as input. Note that there is also a version of E-OBS at a resolution of 0.25° available and originally represented in EStreams. Users should be aware that different resolutions of a forcing dataset can lead to slightly different performances. Similarly, there are different E_{pot} products available from E-OBS as derived indices, but here, we used the E_{pot} product provided in EStreams. The CAMELS meteorological data are usually based on in-situ observations. When more than one option was available, we chose the data with the highest spatial and (original) temporal resolution to represent the CAMELS data for this study (Table 2). While E-OBS was developed specifically for Europe, one can still expect a lower data quality than for datasets created for a smaller region (e.g., national datasets) due to the lower spatial resolution and interpolation choices used to achieve the larger spatial extent of the dataset.

Note that the shapefiles that were used in EStreams and in CAMELS to calculate the areal averages for the meteorological forcings potentially differed. In addition to the different data sources, this can affect the forcing data.

Table 2: Overview of the data sources for the meteorological data (precipitation *P*, temperature *T*, and potential evapotranspiration E_{pot}) for the different CAMELS datasets in this study.

Country	Variable(s)	Source / equation	Resolution	Dataset type	Reference(s)
Czechia	*P*, *T*	*unpublished* data from Czech Hydrometeorological Institute (M. Jeníček / O. Ledvinka, pers. comm.)	500 m	station-based	Štěpánek et al. (2011, 2013)
	E_{pot}	*unpublished* data based on Oudin equation (M. Jeníček / O. Ledvinka, pers. comm.)	*	-	Oudin et al. (2005)
Denmark	*P*	Danish Meteorological Inst.	10 km	station-based	Scharling (1999b)
	T	Danish Meteorological Inst.	20 km	station-based	Scharling (1999a)
	E_{pot}	Makkink	40 km	-	van Kraalingen and Stol (1997)
France	*P*, *T*	SAFRAN by Météo-France	8 km	reanalysis	Quintana-Seguí et al. (2008); Vidal et al. (2010)
	E_{pot}	Oudin	*	-	Oudin et al. (2005)
Germany	*P*	HYRAS by Deutscher Wetter-dienst (DWD)	1 km	station-based	Rauthe et al. (2013)
	T	HYRAS by DWD	5 km	station-based	Razafimaharo et al. (2020)
	E_{pot}	Modified Hargreaves	*	-	Adam et al. (2006); Droogers and Allen (2002); Hargreaves and Samani (1982)
Great Britain	*P*	CEH-GEAR	1 km	station-based	Keller et al. (2015); Tanguy et al. (2016)
	T	CHESS-met	1 km	station-based	Robinson et al. (2017a)
	E_{pot}	CHESS-PE (based on Penman-Monteith)	1 km	-	Robinson et al. (2016, 2017b)
Spain	*P*, *T*, E_{pot}	Spanish Meteorological Agency (AEMET*)*	5 km	station-based	Peral García et al. (2017)
Sweden	*P*, *T*	PTHBV database by Swedish Meteorological and Hydrological Institute (SMHI)	4 km	station-based	SMHI (2025)
	E_{pot}	*unpublished* data based on Hamon equation (C. Teutschbein, pers. comm.)	*	-	Hamon (1963)

Country	Variable(s)	Source / equation	Resolution	Dataset type	Reference(s)
Switzerland	P	RhiresD by MeteoSwiss	2 km	station-based	MeteoSwiss (2021b)
	T	TabsD by MeteoSwiss	2 km	station-based	MeteoSwiss (2021a)
	E_{pot}	Penman–Monteith without interception correction	*	-	Viviroli et al. (2007)

* calculation for each catchment based on its meteorological data

2.3 Calculations of the differences in the CAMELS and E-OBS data

We compared the precipitation, E_{pot}, and temperature data from EStreams (i.e., the E-OBS data) to the corresponding data from the different CAMELS datasets (Table 2) for the twenty years between October 1995 and September 2015 to get an overview of the differences in the data. For precipitation and E_{pot}, we determined the relative difference in the mean annual sums. For temperature, we determined the mean absolute difference for the daily data. When comparing the two datasets, we used the E-OBS data obtained from EStreams as minuend and the analogous data obtained from the CAMELS datasets as subtrahend, i.e., positive differences indicate higher values in the E-OBS data, while negative differences indicate lower values in the E-OBS data than the CAMELS data. To calculate relative differences (for precipitation and E_{pot}), we divided by the mean annual sum determined from the CAMELS dataset. Thus, for example, a value of -20 % indicates that the mean annual sum obtained from E-OBS is 80 % of the mean annual sum obtained from the CAMELS dataset, and a value of 40 % indicates that the mean annual sum obtained from E-OBS equals 140 % of the mean annual sum obtained from the CAMELS dataset.

2.4 Modelling experiments

Following the methodology of Clerc-Schwarzenbach et al. (2024), we defined different combinations of forcing data (“scenarios”) to calibrate the hydrological model (Table 3). This allowed us to determine how the forcing data and their individual variables impacted hydrological model performance. Since EStreams does not provide daily streamflow data, but where to find them, we used the observed streamflow data contained in the CAMELS datasets for all modelling experiments. Thus, we made sure that the hydrological model performance was not affected by different streamflow data.

In the first two scenarios, we used the CAMELS forcing data (scenario I) or the E-OBS data obtained from EStreams (scenario II). To isolate the impact of the forcing variables, we additionally defined three mixed scenarios. Scenario III used precipitation from CAMELS and E_{pot} and temperature from E-OBS. Similarly, scenario IV evaluated the effect of using only E_{pot} from CAMELS, and consequently if using different E_{pot} formulations would change our results, while scenario V assessed the effect of using only temperature from CAMELS. Note that due to the dependency of the E-OBS E_{pot} data on the E-OBS temperature data, model calibration was influenced by the E-OBS temperature data even when replacing the temperature data from E-OBS with those from CAMELS (scenario V).

Analogously to Clerc-Schwarzenbach et al. (2024), we calibrated the HBV model (Bergström, 1992, 1995) in the version HBV-light (Seibert and Vis, 2012) with a genetic algorithm (Seibert, 2000). Each catchment was divided into elevation zones of 200 m elevation difference, whereby an elevation zone had to account for at least 5 % of the catchment area. For the determination of the elevation zones, we used the shapefiles provided by EStreams, and the Copernicus DEM at a resolution of 30 m (European Space Agency and Airbus, 2022).

We used the five years from October 1990 to September 1995 as the warming-up period for the model, and the twenty years from October 1995 to September 2015 as the simulation period for which we optimized daily streamflow simulation in terms of the Kling-Gupta efficiency KGE (Gupta et al., 2009). One calibration consisted of a total of 3500 model runs. We conducted each calibration ten times to account for equifinality. We used equal weights on the ten simulated hydrographs to calculate an ensemble mean hydrograph. We determined the model performance (using again the KGE as well as the PBIAS, i.e., the percent bias of the simulated streamflow compared to the observed one) for each catchment and each scenario by comparing this ensemble mean hydrograph to the observed hydrograph.

Table 3: Combinations of forcing data ("scenarios") for the modelling experiments.

Scenario	Description	Precipitation	E_{pot}	Temperature
I	CAMELS	CAMELS	CAMELS	CAMELS
II	E-OBS (EStreams)	E-OBS (EStreams)	E-OBS (EStreams)	E-OBS (EStreams)
III	E-OBS with CAMELS precipitation	CAMELS	E-OBS (EStreams)	E-OBS (EStreams)
IV	E-OBS with CAMELS E_{pot}	E-OBS (EStreams)	CAMELS	E-OBS (EStreams)
V	E-OBS with CAMELS temperature	E-OBS (EStreams)	E-OBS (EStreams)	CAMELS

2.5 Statistical tests

We used the Spearman rank correlation coefficient (r) and corresponding p-values to assess the relationships between the model performance differences in terms of the KGE and different catchment attributes. We used the locally-weighted polynomial regression (lowess; Cleveland, 1979) to visually represent the relationships. In addition, we used the Wilcoxon signed-rank test (Wilcoxon, 1945) on paired median KGE values to evaluate whether the differences in model performance between the scenarios were statistically significant.

3 Results

3.1 Comparison of the meteorological data

The mean annual precipitation sums in the E-OBS data were lower than the mean annual precipitation sums in the CAMELS data for 2362 catchments (88 %). For 758 catchments (28 %), the deviation of the mean annual precipitation sums in E-OBS from the ones in CAMELS exceeded -10 %. Conversely, there were only 33 catchments (1 %) for which the mean annual precipitation sums in E-OBS were overestimated by +10 % or more from the ones in CAMELS. Differences between the two data sources were largest for the catchments in Spain and smallest for the catchments in Germany (Fig. 1).

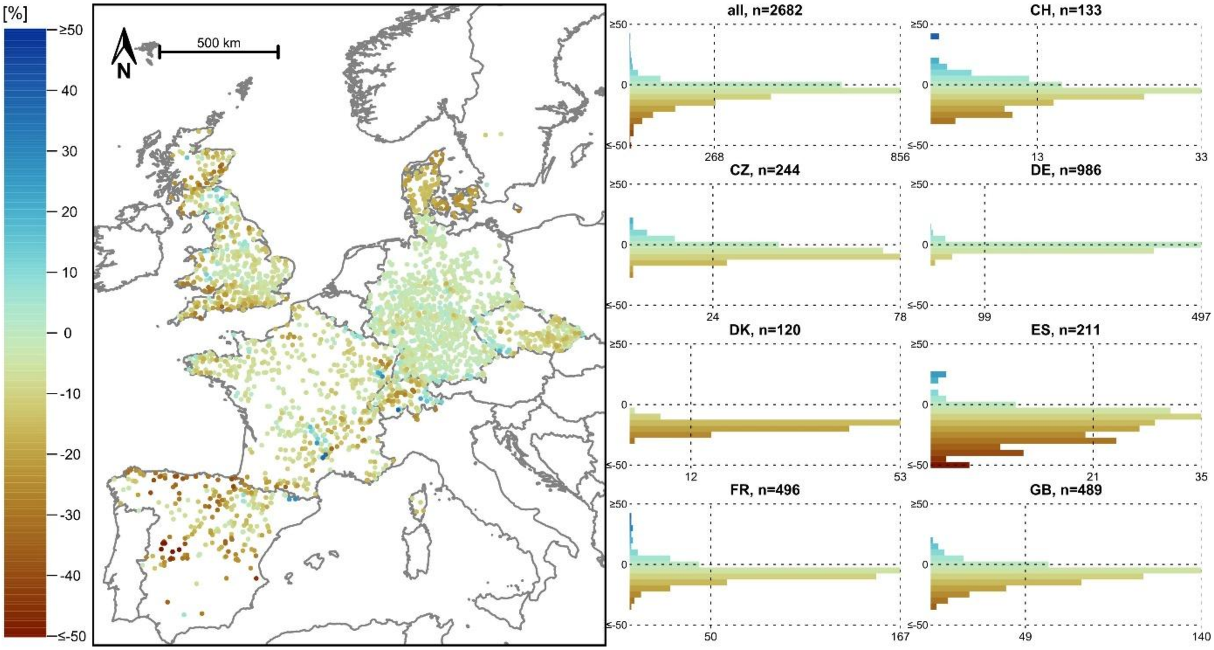

Figure 1: Relative difference in the mean annual precipitation (for a 20-year period: 1995-2015) between the E-OBS data obtained from EStreams and the different CAMELS datasets. Negative values and brown colours indicate less precipitation in E-OBS than in CAMELS, positive values and blue colours more precipitation in E-OBS. On the map, the catchments with the largest deviations were plotted last to increase their visibility. Note that there is no separate histogram for the three catchments in Sweden and that the number of catchments per histogram differ. This is illustrated by the vertical lines indicating 10 % (rounded) of the number of catchments per histogram. The colour scale was cut at ±50 %. The scale bar refers to the map center. The base map was obtained from Natural Earth (naturalearthdata.com). The colour palette used in this and all other maps are scientific colour palettes from Crameri (2023).

The opposite was found for the annual sums of E_{pot}: For 2508 catchments (94 %), the mean annual E_{pot} calculated from the E-OBS data was higher than for the CAMELS data. For 1353 catchments (50 %), the deviation of the E-OBS E_{pot} sums from the CAMELS E_{pot} sums were at least 10 %. Clearly lower E_{pot} sums derived from E-OBS than from CAMELS could only be observed for catchments in Sweden, on the Danish islands, in southern Switzerland, and for some catchments in northern Spain (Fig. 2). As different equations or data sources were used in the different CAMELS datasets (see Table 2) to obtain the E_{pot} data, the order of magnitude of the deviations changed abruptly along the national borders in some cases (e.g., along the border between Czechia and Germany). It is noteworthy that for E_{pot}, there tend to be small differences between the two datasets for Spain (while this was not the case for precipitation).

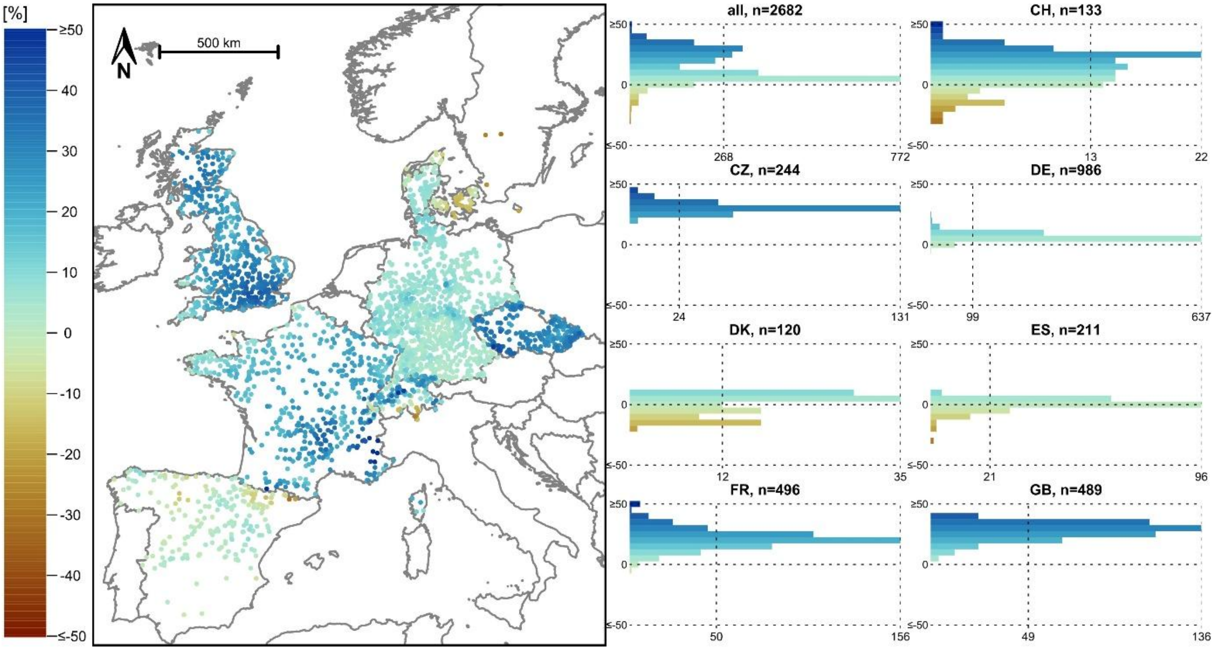

Figure 2: Relative difference in the mean annual E_{pot} (for a 20-year period: 1995-2015) obtained from EStreams and calculated from the E-OBS data compared to the mean annual E_{pot} calculated from the different CAMELS datasets. Negative values and brown colours indicate a lower E_{pot} in E-OBS than in the CAMELS datasets, positive values and blue colours a higher E_{pot}. On the map, the catchments with the largest deviations were plotted last to increase their visibility.

Due to the differences in the precipitation and the E_{pot} data, the aridity indices (E_{pot}/P) calculated from the two data sources differed, although they were still highly correlated (Pearson's correlation coefficient of 0.94) (Fig. 3). Given the lower precipitation and higher E_{pot} sums for most catchments, the aridity indices were generally higher when the E-OBS data obtained from EStreams were used than when the CAMELS data were used. This did not apply for Sweden, as the E_{pot} sums based on E-OBS were lower than the ones from CAMELS for this country. The two calculated aridity indices aligned best for Germany and worst for Spain, Great Britain, and Czechia. Spatially, the aridity indices derived from both datasets followed the expected pattern, with more arid catchments in southern Europe and north-eastern Germany and more humid catchments in the other regions (see Fig. A1 for the CAMELS data and Fig. A2 for the E-OBS data).

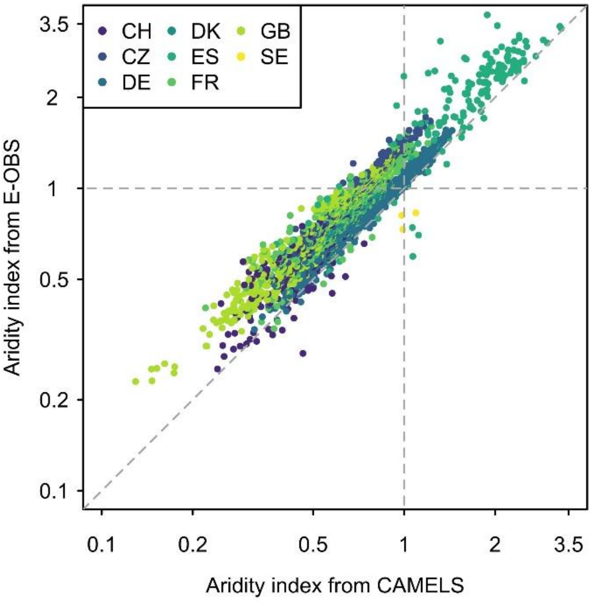

Figure 3: Comparison of the aridity indices (E_{pot}/P) based on the CAMELS and the E-OBS data (for a 20-year period: 1995-2015), colour-coded by country. Note the logarithmic axes.

Comparison of the temperature data in the two datasets revealed that the average temperature in E-OBS was higher (median difference: 0.3° C) for most catchments than the average temperature in CAMELS (see Fig. S2 in the Supporting Information). There were 636 catchments (24 %) for which the average temperature was lower in E-OBS than in the CAMELS datasets. Note that in the HBV model, temperature has an effect on the snow routine, with higher temperatures resulting in a larger fraction of precipitation falling as rain (and thus faster streamflow generation). However, as the threshold temperature for the differentiation between rain and snow is adapted during calibration, it is expected that the model can compensate comparably well for biased temperature time series. Thus, the main effect of the differences in temperature are the differences in E_{pot} which are highly affected by the temperature data used as input to the calculations (Fig. 22Figure 2).

3.2 Model performances

3.2.1 Model performances with the CAMELS and the E-OBS forcing data

Overall, high model performances were achieved for most catchments when the CAMELS data (scenario I) were used for model calibration (Fig. A3). For 2507 of the 2682 catchments (93 %) the KGE was higher than 0.70, with a median performance of 0.89.

The model performances were also high for most catchments when the E-OBS forcing data (scenario II) were used for model calibration (Fig. 4). For 2434 of the catchments (91 %) the KGE was higher than 0.70, which is comparable to the 2507 catchments that fulfilled this criterion for the CAMELS data (scenario I). Furthermore, the median performance achieved with the E-OBS data from EStreams (scenario II) of 0.87 was very similar to the 0.88 achieved with the CAMELS data (scenario I). However, over all catchments and according to the Wilcoxon signed rank test, performances were significantly higher for scenario I than for scenario II.

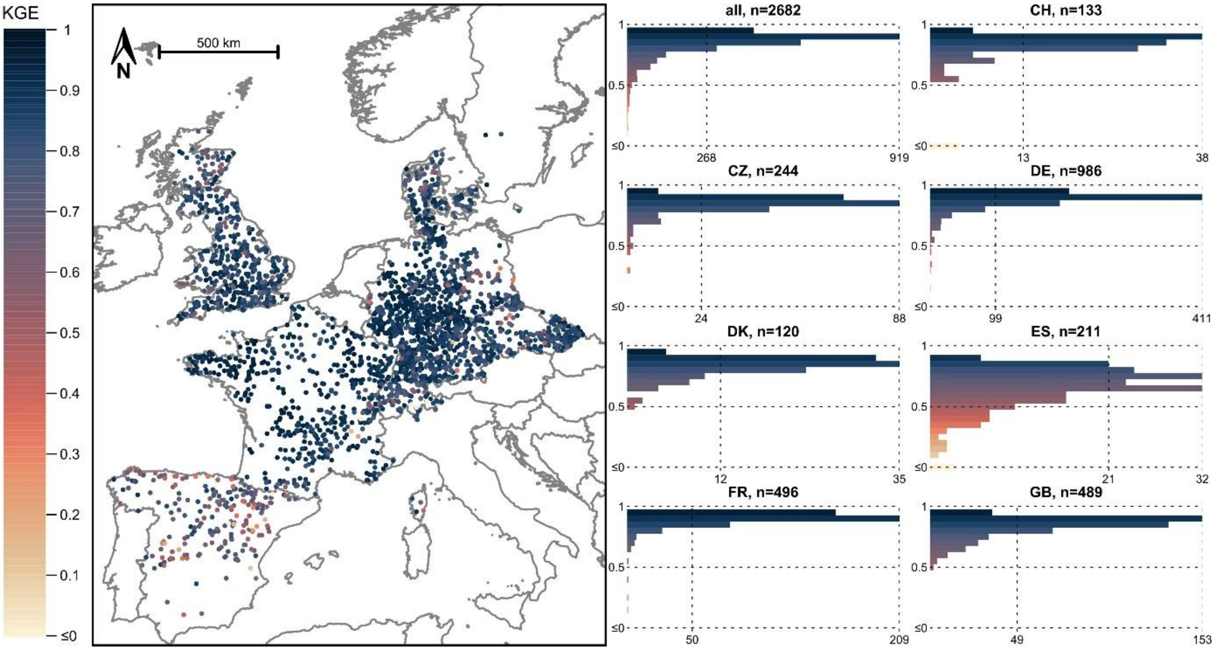

Figure 4: Model performance (Kling-Gupta efficiency, KGE) achieved for the 20-year period between October 1995 and September 2015 when the E-OBS data obtained from EStreams were used for model calibration (scenario II). Note that the lower limit of the colour scale was cut at zero. Lower performances were plotted last to improve their visibility.

Considering the PBIAS as an additional measure for model performance (Fig. A4 for scenario I and Fig. 5 for scenario II), we found a small PBIAS (between –10 % and 10 %) for 2477 catchments for scenario I and 2468 catchments for scenario II (both 92 %). There were more occurrences of streamflow overestimations (i.e., positive PBIAS) when the CAMELS forcing data were used (scenario I) than when the E-OBS data were used: For 176 catchments (7 %), the PBIAS was larger than 10 %, and for 26 catchments (1 %), it was larger than 100 %. Meanwhile, a streamflow overestimation of at least 10 % only happened in 31 catchments (1 %) with the E-OBS forcing data (scenario II), but there was a considerable number of catchments for which the streamflow was underestimated (181 catchments (7 %) with a PBIAS smaller than –10 %).

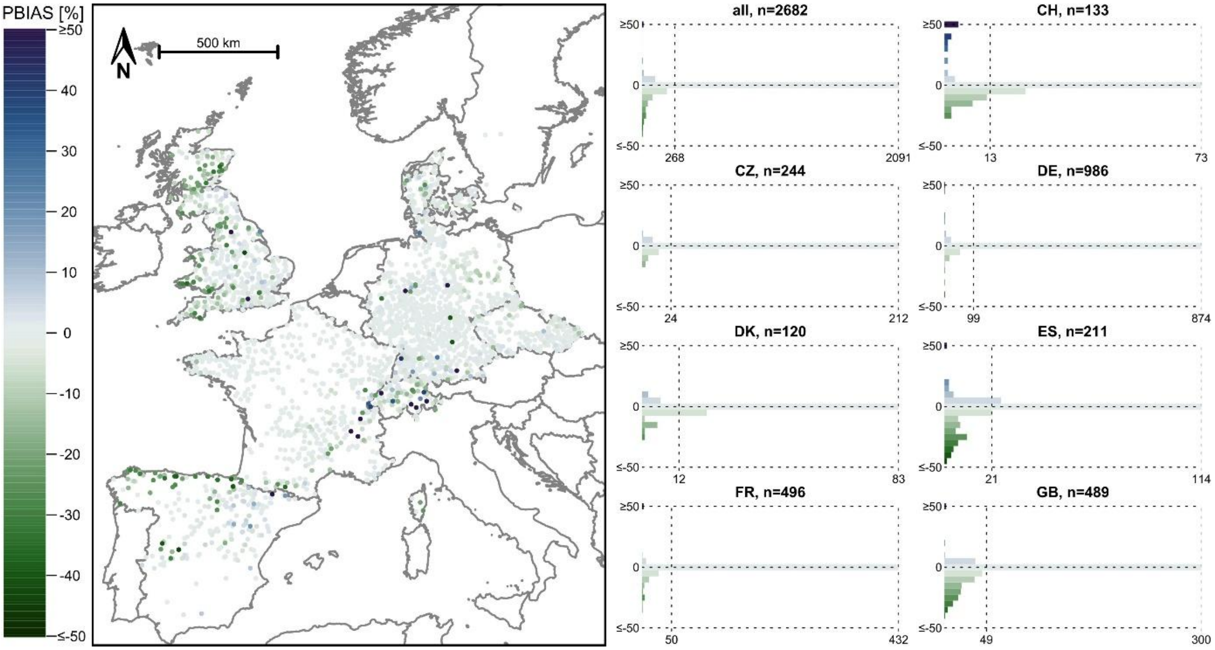

Figure 5: PBIAS (relative deviation of the simulated streamflow from the observed streamflow) for the 20-year period between October 1995 and September 2015 when the E-OBS data obtained from EStreams were used for model calibration (scenario II). Note that the limits of the colour scale were cut at ±50 %. Largest deviations were plotted last to improve their visibility.

3.2.2 Differences in model performance between scenario II and I

To directly assess the differences in model performances between scenario II and I, we looked at the differences in model performance (Fig. 6). For 1669 catchments (62 %), model performances were (at least slightly) higher when the CAMELS data were used (scenario I), while for the other 38 % of all catchments, the use of E-OBS data resulted in better model performances. The strongest regional differences were found for the catchments in Spain and Great Britain.

For France, there were notable improvements in model performance when using the E-OBS dataset: 316 catchments (64 %) had higher performances with the E-OBS data (median ΔKGE=0.01). Note that the French catchments that benefitted most from the E-OBS forcing data are the catchments for which the PBIAS was strongly positive when the CAMELS forcing data were used (scenario I; Fig. A4). For Sweden, all three catchments performed better with the E-OBS data, but the differences were very small (median ΔKGE=0.006; Fig. 6). For the catchments in Spain, it was the opposite: 200 catchments (95 %) performed better with the CAMELS dataset, reaching a median ΔKGE of -0.10. Higher KGE for scenario I was also observed for the catchments in Great Britain (80 %), Czechia (66 %), Switzerland (65 %), Denmark (61 %) and Germany (59 %).

The results also indicated some interesting intercountry patterns (Fig. 6). In France, the most considerable positive differences occurred for the catchments in the eastern, more mountainous part of the country, while for Great Britain, the CAMELS data resulted in clearly higher performances in most regions but not in the area around London.

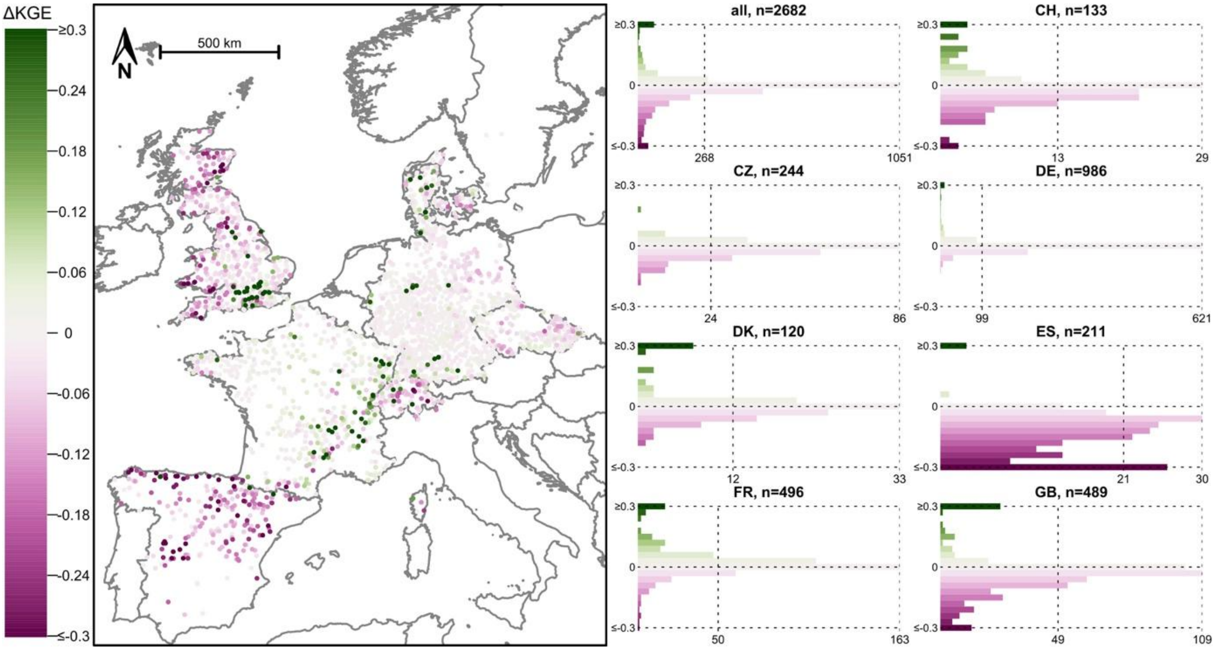

Figure 6: Difference in model performance (Kling-Gupta efficiency, KGE) between scenario II and scenario I. Positive values and green colours indicate higher performances when the E-OBS data obtained from EStreams were used, negative values and pink colours indicate higher performances when the CAMELS data were used. For the model performances, see Figs. 4 and A4. The catchments with the largest differences (in absolute terms) were plotted last to improve their visibility. Note that the colour scale was cut at ±0.3.

3.2.3 Differences in model performance between scenario II and scenarios III, IV, and V

Replacing precipitation from E-OBS with data from CAMELS had by far the strongest impact on model performance (scenario III, Fig. A5). For most catchments, the performance differences between scenarios II and III closely mirrored the performance differences between scenarios II and I, indicating that precipitation accounted for a large share of the overall differences in performance. For only a few catchments (mostly in Great Britain), the performance gap between scenarios II and I was notably larger than between scenarios II and III. The opposite occurred for very few catchments (see Fig. S3 in the Supporting Information).

The effect of replacing the E_{pot} data (scenario IV) was quite limited. The higher E_{pot} data based on E-OBS were beneficial for a handful of catchments (ΔKGE>0.30 for 19 catchments), but the median difference was 0.00 (Fig. A6). Replacing only the temperature time series with the CAMELS data (scenario V) had virtually no effect on model performance for most catchments. There were no catchments for which the replacement of the temperature data increased the KGE by more than 0.10 and no catchments for which it decreased the KGE by more than -0.10 (see Fig. S4 in the Supporting Information). Note that only the temperature time series provided as input data to the HBV model were replaced, and not the data that were used to calculate E_{pot}.

3.3 Model performance linked to catchment attributes

We calculated the Spearman rank correlation between model performance and several catchment attributes available in EStreams (see Table S1 and Fig. S5 in the Supporting Information). The number of E-OBS precipitation stations and the aridity index emerged as particularly interesting variables given their apparent relationships with model performance.

3.3.1 Number of E-OBS precipitation stations

To assess the impact of the variable coverage of meteorological stations used to produce the gridded E-OBS dataset, we examined the relationships between the number of E-OBS stations within or near each catchment and the model performance for scenario II (i.e., using the E-OBS data contained in EStreams for all meteorological variables; see Fig. 44). Here, we present these relationship assessments per country (Fig. 7). The number of E-OBS precipitation stations was obtained from the EStreams dataset, defined as the count of stations located within a 10 km buffer of the catchment boundary (do Nascimento et al., 2024).

Model performances for scenario II tended to be higher when more E-OBS stations were located in or around a catchment. Except Sweden (for which we only considered three catchments), there was a significant positive correlation (*p*-value<0.05) between the density of stations and the model performances achieved with the E-OBS forcing data for all countries when all catchments were considered. To avoid that correlations are only due to a tendency for higher model performances in large catchments and more stations in large catchments, we also analysed the relationship for only the catchments smaller than 100 km^2. For Spain, the threshold was set to 500 km^2 due to the small number of catchments smaller than 100 km^2. There were still significant positive correlations for most countries, but not for Denmark and France. The correlations even increased for Switzerland (r=0.53 to r=0.71) and Great Britain (r=0.32 to r=0.36).

The performances were generally the lowest for areas with sparse station coverage, such as Spain and Great Britain (Fig. 7). However, low E-OBS station density did not always result in poor model performance. For catchments in countries like France, Great Britain, Denmark, and Sweden, KGE values remained mostly above 0.5 despite a comparatively lower station density, suggesting that factors other than station density (such as the spatial variability of the rainfall due to topography or convective rainfall) also influence model accuracy. For additional insights, we also provide the scatterplots of the differences in model performance between scenarios II and I compared to the number of E-OBS stations per catchment (see Fig. S6 in the Supporting Information), which further supports this discussion.

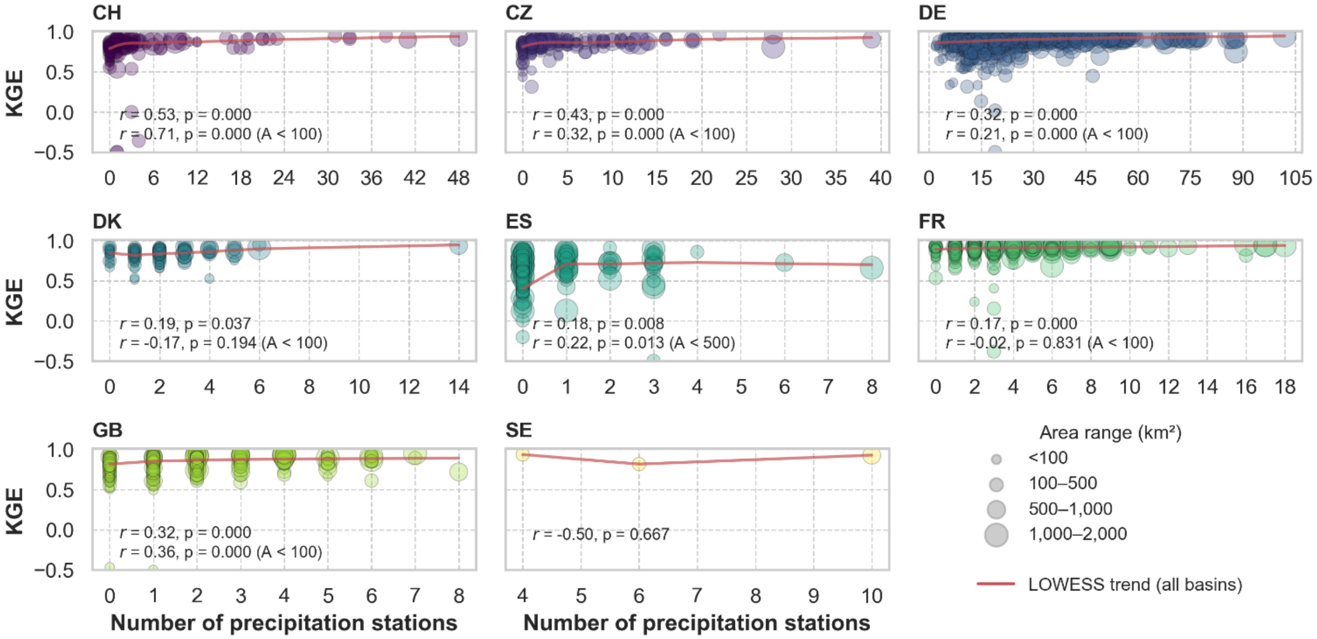

Figure 7: Scatterplots showing the model performance (Kling-Gupta efficiency, KGE) for scenario II (y-axes) versus the number of E-OBS precipitation stations in and around each catchment, per country. Each circle represents one catchment, the size is based on catchment area. Each subplot contains the Spearman rank correlation (*r*) and corresponding *p*-value, and the lowess trend line. The *r* and the *p*-value are computed for all catchments per country and also for the ones with an area below 100 km^2 (500 km^2 for Spain). Note that the y-axes were cut at -0.5, in accordance to Fig. 4, and the x-axes differ for the different subplots.

3.3.2 Aridity index

We also evaluated the model performances for scenarios I and II in relation to the aridity indices derived from the respective forcing data (Fig. 8). Despite some atypical cases (Denmark in scenario I and Great Britain in scenario II), the model performances tended to be significantly lower in catchments with higher aridity indices (drier catchments). This trend was particularly evident for the catchments in Czechia, Germany, Spain, and Great Britain. Although the pattern appeared with both forcing datasets, it was more pronounced for the CAMELS data (scenario I), especially for the catchments in Czechia, and Spain (both with *r*=-0.48). For Switzerland and France, the Spearman rank correlations were non-significant and close to zero for both forcing data types.

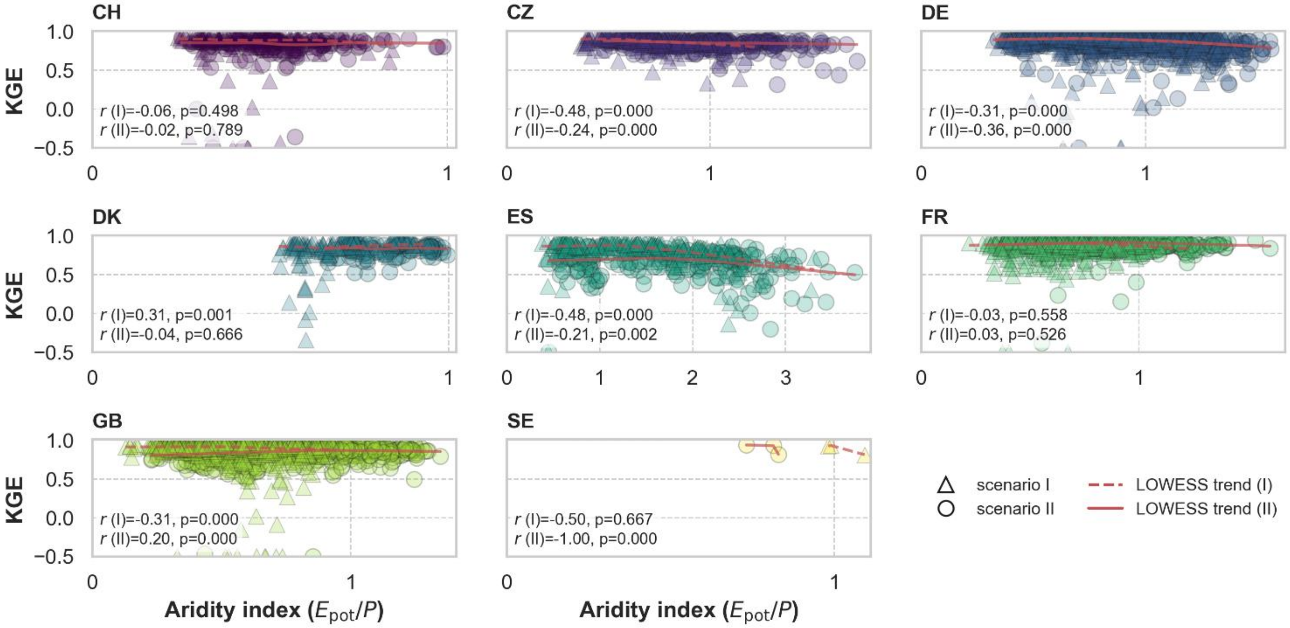

Figure 8: Scatterplots showing the model performance (Kling-Gupta efficiency, KGE) for scenario I and II (y-axes) versus the aridity indices derived from the corresponding forcing data (CAMELS for scenario I, E-OBS from EStreams for scenario II), per country. Each subplot contains the Spearman rank correlation (*r*) and corresponding *p*-value, and the lowess trend line. Note that the y-axes were cut at -0.5, in accordance to Fig. 4. Note that the x-axes differ for the different subplots.

4 Discussion

4.1 Differences in the meteorological data

Our results show that the mean annual precipitation sums in E-OBS are systematically lower than those in the CAMELS datasets for most catchments (Section 3.1), with the largest differences occurring in Spain, whereas the smallest deviations were found in Germany (Fig. 1). This pattern is consistent with the findings of Bandhauer et al. (2022), who showed that E-OBS tends to underestimate precipitation to smooth spatial contrasts in comparison to a reference dataset, particularly in mountainous regions and in areas with sparse station coverage. In contrast, where E-OBS is supported by dense observation networks, such as in Germany, precipitation contrasts were better represented and the agreement with reference datasets was substantially improved (Bandhauer et al., 2022). Taken together, this suggests that the lower precipitation estimates in E-OBS relative to CAMELS are largely driven by the combination of coarser grid-resolution and lower underlying station density, especially in complex terrain. By contrast, mean temperatures from E-OBS and mean annual E_{pot} were generally higher than those from the CAMELS forcings (Fig. 2). We do not have a clear explanation for this behaviour, but potential reasonings are homogenization procedures, elevation corrections, and interpolation methods used in the different national datasets, which is beyond the scope of this study. The difference in E_{pot} is likely driven by the differences in temperature and may be further amplified by differences in the E_{pot} formulations and parameterizations used in the different products. The differences between the two forcing sources are the smallest in Germany also for E_{pot}, where there is both higher E-OBS station density, and a similar equation used for E_{pot} derivation in CAMELS to the one in E-OBS (Table 2).

4.2 Model performances

The HBV model is known for its capabilities in simulating streamflow, particularly in humid catchments, where water flow is related to varying soil saturation and hydrological connectivity (Knapp et al., 2022, 2024). This, in part, seems to explain the consistently high model performances achieved using either the CAMELS or the E-OBS forcing data (Figs. 4 and A3 as well as Figs. 5 and A4) for the more humid catchments, such as those in Sweden and Denmark. In contrast, in Spain, where the most arid catchments are located, the KGE values were the lowest and most variable for both scenarios. These findings are reinforced by the observed relationship between model performance and the aridity index shown in Fig. 8. The trend of decreasing performance with increasing aridity further supports the assertion that arid catchments pose significant challenges for hydrological modelling. Several other studies have suggested that for dry catchments more complex model structures may be needed for streamflow simulation, and even then, they still tend to yield lower model performance (Atkinson et al., 2002; David et al., 2022; Massmann, 2020).

Yet, the lower model performance for the catchments in Spain may be attributed not only to the inherent complexities of streamflow generation in arid environments, but also to the higher variability and limited availability of observational hydrometeorological data in these regions, which complicates model calibration and validation, as noted in previous studies (do Nascimento et al., 2023; Yu et al., 2011). Additionally, previous studies have pointed out that many Spanish catchments, including the ones available in the currently used BULL dataset, are highly regulated, with dams and diversions (Klotz et al., 2025; Senent-Aparicio et al., 2024b). Although we purposefully adopted the criteria based on the number of lakes, and normalized upstream capacity area discussed in section 2.1, some of these heavily modified catchments may still not be adequately filtered out, thereby further impairing overall model performance.

4.3 Influence of forcing data characteristics on model performance

Our findings indicate that model performance in scenario II is strongly influenced by the density of stations used to obtain the E-OBS data (see section 3.3.1). As a result, the reliability of model outputs varies considerably across regions—an observation that is consistent with previous research (Klotz et al., 2025). This spatial dependency is visually supported by Figure 6 in the EStreams paper by do Nascimento et al. (2024), which shows the density of E-OBS stations across Europe. Notably, for regions with a high density of stations, such as Germany, the model achieved the highest KGE values with E-OBS data, underscoring the critical role of data availability and quality in hydrological modelling accuracy. Importantly, the significant correlations also found on smaller catchments (Fig. 7) confirmed that this relationship is not just an artifact of catchment size. To further understand the regional variations in model performance, we also examined how the type and characteristics of the CAMELS forcing data varied across countries, extending the comparison to both scenario I and II. This comparison provides insights into the role of data resolution, origin, and processing methodology in shaping the current model outcomes.

Germany presents a particularly consistent case: E-OBS shows high station density over the country (Fig. 6 in do Nascimento et al., 2025), likely overlapping with the ground observations used in the national HYRAS dataset (Rauthe et al., 2013; Razafimaharo et al., 2020), which explains the high agreement between the precipitation forcings (Fig. 1) and the similar model accuracy obtained with both input datasets (Fig. 6). Minor discrepancies are expected, given HYRAS's finer spatial resolution (1 km for precipitation, 5 km for temperature) compared to E-OBS (0.1°).

In contrast, the national products for Spain, Switzerland, and Great Britain are based on substantially denser station networks than E-OBS (Table 2), which likely contributes to their higher model accuracy in scenario I relative to scenario II. As shown in Figure 6 by do Nascimento et al. (2025), E-OBS displays sparse station coverage in these countries. At the same time, it is worth noting that their respective national datasets—AEMET (5 km), RhiresD/TabsD (2 km), and CEH-GEAR/CHESS (1 km)—offer much finer spatial detail, which likely offer a better local representation of forcing patterns.

The meteorological data in the CAMELS of Denmark and France stand out for their coarser spatial resolution (10 km and 8 km) compared to the other CAMELS datasets. In Denmark, this likely reduced the performance advantage of the CAMELS forcings, resulting in similar outcomes between scenarios I and II. In France, however, the situation differs: unlike the other CAMELS datasets, the SAFRAN product combines reanalysis and station-based data (Quintana-Seguí et al., 2008; Vidal et al., 2010). The KGE performances in scenario II being slightly better in 64 % of the French basins might suggest that SAFRAN's reanalysis nature, alongside coarser gridded-resolution may explain the marginally lower accuracy relative to the purely station-based E-OBS forcing. However, note that the lower precipitation data and higher E_{pot} data in E-OBS compared to CAMELS-FR can also just be advantageous for model calibration, e.g., if there are additional outflows from the catchments that are not represented in the model structure (see section 4.4 for further discussion).

For Czechia, the CAMELS forcings resulted in clearly better model performances than the E-OBS forcings, consistent with its high-resolution station-based dataset, while Sweden, represented by only three catchments, provides insufficient evidence for interpretation.

Overall, these results indicate that differences in model performance seem to be mainly driven by the station density used to derive the forcing, spatial resolution, and type of the product (station-based or reanalysis), leading to variability in the data quality even within the same source.

4.4 Influence of different E_{pot} data in model performance

In this work, the differences in model performance across the various sources of forcing data were mainly attributed to the differences in precipitation inputs (scenario III), whereas discrepancies in E_{pot} and temperature data hardly affected model performance (scenarios IV and V). These findings are consistent with those of Clerc-Schwarzenbach et al. (2024), who compared the Caravan and CAMELS datasets. Although their analysis revealed larger differences in E_{pot} than in precipitation data, it was still the precipitation inputs that exerted the greatest influence on model performance.

In this study, the differences between the E_{pot} data derived from E-OBS and the E_{pot} data derived from the CAMELS datasets were much smaller, but still obvious (which makes sense, considering the differences in the temperature data that were generally used as input to the E_{pot} calculations). This demonstrates once again the large uncertainties that we face when using different approaches to estimate E_{pot}. Several studies have already identified this issue as a persistent “blind spot” in hydrological modelling (Bai et al., 2016; Federer et al., 1996; Hanselmann et al., 2024). E_{pot} calculated with the Hargreaves equation, as in EStreams, has been found to be a reliable method in various hydrological modelling applications, including in Central Europe (Pimentel et al., 2023), Germany (Loritz et al., 2024) and other regions (Bangi and Soraganvi, 2023; Sperna Weiland et al., 2012). Furthermore, as shown in Fig. A6, the differences in E_{pot} data did not affect model performance results strongly, so it can be expected that the use of a different equation would not notably change the findings of this study.

While the different E_{pot} data had a very limited effect on model performances in general, the higher E_{pot} data derived from E-OBS were beneficial for the model performance in some cases, as they likely allowed the model more flexibility to adjust the water balance. For example, the low performances with the CAMELS-GB data for the catchments in the karstic area around London can (partially) be explained by the inability of the HBV model to simulate groundwater losses (Lane et al., 2019; Oldham et al., 2023; Seibert et al., 2018). In such catchments, the higher E_{pot} values from E-OBS effectively helped to improve the water balance by allowing for more evaporation, thereby compensating for the unmodelled groundwater losses. This is supported by the fact that the strong streamflow overestimation in this and other regions (e.g., southeastern France) when the CAMELS forcing data were used (Fig. A4) could be avoided when the E-OBS forcing data were used (Fig. 5). However, it is important to note that while this adjustment led to improved model performance, such compensatory effects are not desirable when the objective is to accurately represent internal catchment processes. Achieving realistic process representation should, generally, remain a central goal in hydrological modelling (Kirchner, 2006).

4.5 Limitations

As mentioned with the example of the compensation effects due to higher E_{pot} data, a higher model performance does not necessarily mean a better representation of the hydrological processes. Still, we used the performance as an indicator for the hydrological efficacy of different forcing data. Although model performances are often used as an aggregated measure of data quality (Beck et al., 2017; Clerc-Schwarzenbach et al., 2024; Tarek et al., 2020), model performance can also be heavily influenced by the chosen model structure, particularly if it does not align well with the physical characteristics of a given catchment. This structural sensitivity means that performance differences may reflect model limitations as much as data quality (Beven, 2018).

Beyond model structure, it is worth noting that here we performed single-basin calibrations. While this allows for localized optimization, it does not reflect how models are typically regionalized for prediction in ungauged basins. Future research should explore how the identified performance patterns translate to a regionalization framework, which would provide more practical insights for prediction in data-scarce environments, and therefore, where model calibration is not possible.

Finally, all simulations in this study were conducted at a daily time step. For smaller catchments, a finer temporal resolution, such as hourly, could provide more meaningful insights. With the increasing availability of temporally high-resolution datasets (Coxon et al., 2025; Dolich et al., 2025; Nijzink et al., 2025), future studies may benefit from repeating similar analyses at the sub-daily timescale.

5 Conclusions

In this study, we compared the meteorological time series for 2682 European catchments in the EStreams dataset with time series from eight smaller-scale datasets (mostly CAMELS datasets). Moreover, we evaluated how the different types of meteorological forcing data influence the performance of a bucket-type hydrological model.

Our results showed that for most catchments, mean annual precipitation values obtained from the E-OBS dataset were lower than those from the corresponding CAMELS datasets. The opposite was true for the average temperature and thus the annual sums of E_{pot} (higher values in E-OBS than in CAMELS). These discrepancies led to consistently higher aridity indices computed with the E-OBS data in comparison to the CAMELS data for most catchments, although the spatial pattern remained similar. Such systematic differences highlight important inconsistencies across the two data sources that can affect the outcomes of hydrological synthesis studies across large areas.

Despite these differences, model calibration using either set of forcing data achieved good model performances for most catchments (KGE of at least 0.70 in more than 90 % of the catchments). However, performances were generally slightly lower when using E-OBS data than when using CAMELS data: For approximately 60 % of the catchments, the model performance was higher when using CAMELS forcing data. Considering the national curation and higher resolution of the CAMELS datasets, this makes sense.

Our findings indicate that cross-country differences in model performance are primarily driven by variations in station density, spatial resolution, and the inclusion of reanalysis components, rather than by substantial inconsistencies in data quality between E-OBS and national products. We observed that model performances using E-OBS forcing data were lower in regions with a lower E-OBS station density. This highlights the critical role of station coverage on hydrological model performance, an issue that becomes even more pronounced in mountainous regions, where steep climatic gradients in precipitation and temperature make dense and spatially representative data essential for reliable simulations, as well as in arid regions and regions where convective rainfall is relevant.

Overall, while local or national datasets often yield the best model performances, our results suggest that the meteorological forcing data from E-OBS that is included in EStreams represents a valuable and harmonized alternative for pan-European studies. The advantage of E-OBS lies in its observational basis, consistent methodology, and coverage across all of Europe, making it especially useful when national datasets are unavailable or inconsistent. As such, E-OBS and EStreams provide a practical foundation for expanding large-sample hydrology beyond national boundaries while maintaining sufficient data quality for robust model applications.

6 Data availability

Except for CAMELS-CZ and the E_{pot} data for CAMELS-SE, all the CAMELS data used are available from their respective repositories (see Table 1). The E_{pot} data for Sweden were provided by Claudia Teutschbein and the CAMELS-CZ data by Michal Jeníček and Ondřej Ledvinka. These unpublished data are available upon reasonable request. The current version 1.4 of the EStreams dataset is available at a Zenodo repository (https://doi.org/10.5281/zenodo.17598150). The files containing the model performances used in this study are stored at a GitHub repository (https://github.com/thiagovmdon/EOBS-quality).

7 Code availability

All code used to analyse the resolute and derive the figures for this manuscript are available at a GitHub repository (https://github.com/thiagovmdon/EOBS-quality).

8 Author contribution

TN designed the research question, TN and FCS prepared the data, FCS did the model runs, TN and FCS evaluated the results and plotted them, TN and FCS wrote the manuscript.

9 Competing interests

The authors declare that they have no conflict of interest.

10 Acknowledgements

We acknowledge all teams that made the CAMELS datasets and the E-OBS data available. We acknowledge the co-authors with whom TN had developed the EStreams dataset. We thank the Science IT team of the University of Zurich for providing the infrastructure for cloud computing. We thank Ross Woods for awakening our curiosity about the research questions for this manuscript in the first place. We thank Marc Vis for providing us with data ready to be used in HBV for some CAMELS datasets. We thank Claudia Teutschbein for making the E_{pot} data for CAMELS-SE available and Michal Jeníček and Ondřej Ledvinka for providing us with the CAMELS-CZ data. We thank our supervisors Fabrizio Fenicia, Jan Seibert, and Ilja van Meerveld for fruitful discussions and their support, and Ilja van Meerveld also for her comments that helped to clarify the manuscript. TN was funded by a “Money Follows Cooperation” project (Project No. OCENW.M.21.230) between the Netherlands Organization for Scientific Research (NWO) and the Swiss National Science Foundation (SNSF). We thank Alexander Dolich and three anonymous reviewers as well as the editor Albrecht Weerts for their valuable comments.

11 References

Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of global precipitation products for orographic effects, J. Clim., 19, 15–38, https://doi.org/10.1175/JCLI3604.1, 2006.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: Catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Almagro, A., Oliveira, P. T. S., Meira Neto, A. A., Roy, T., and Troch, P.: CABra: A novel large-sample dataset for Brazilian catchments, Hydrol. Earth Syst. Sci., 25, 3105–3135, https://doi.org/10.5194/hess-25-3105-2021, 2021.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies - Chile dataset, Hydrol. Earth Syst. Sci., 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resour. Res., 38, 50-1-50–17, https://doi.org/10.1029/2002WR001487, 2002.

Bai, P., Liu, X., Yang, T., Li, F., Liang, K., Hu, S., and Liu, C.: Assessment of the influences of different potential evapotranspiration inputs on the performance of monthly hydrological models under different climatic conditions, J. Hydrometeorol., 17, 2259–2274, https://doi.org/10.1175/JHM-D-15-0202.1, 2016.

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B., Szentes, O., Tveito, O. E., and Frei, C.: Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions, Int. J. Climatol., 42, 727–747, https://doi.org/10.1002/joc.7269, 2022.

Bangi, S. C. and Soraganvi, V. S.: A modified temperature based model for estimation of potential evapotranspiration over Ghataprabha river basin, south India, Spat. Inf. Res., 31, 583–595, https://doi.org/10.1007/s41324-023-00517-1, 2023.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

Bergström, S.: The HBV Model — its structure and applications, SMHI Reports RH, Norrköping, Sweden, 1992.

Bergström, S.: The HBV model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, Colorado, USA, 443–476, 1995.

Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, WIREs Water, 5, e1278, https://doi.org/10.1002/wat2.1278, 2018.

Bushra, S., Shakya, J., Cattoën, C., Fischer, S., and Pahlow, M.: CAMELS-NZ: Hydrometeorological time series and landscape attributes for Aotearoa New Zealand, https://doi.org/10.5194/essd-2025-244, 26 May 2025.

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., and Siqueira, V. A.: CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth Syst. Sci. Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, 2020.

Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, Hydrol. Earth Syst. Sci., 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, 2024.

Cleveland, W. S.: Robust Locally Weighted Regression and Smoothing Scatterplots, J. Am. Stat. Assoc., 74, 829–836, https://doi.org/10.1080/01621459.1979.10481038, 1979.

Cornes, R. C., Van Der Schrier, G., Van Den Besselaar, E. J. M., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, J. Geophys. Res. Atmospheres, 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020a.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Hannaford, J., Howden, N. J. K., Lane, R., Robinson, E. L., Wagener, T., and Woods, R.: Catchment attributes and hydro-meteorological timeseries for 671 catchments across Great Britain (CAMELS-GB) [Dataset], https://doi.org/10.5285/8344e4f3-d2ea-44f5-8afa-86d2987543a9, 2020b.

Coxon, G., Zheng, Y., Barbedo, R., Fileni, F., Fowler, H., Fry, M., Green, A., Harfoot, H., Lewis, E., Qiu, X., Salwey, S., and Wendt, D.: CAMELS-GB v2: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Copernicus Meetings, https://doi.org/10.5194/egusphere-egu25-4371, 2025.

Crameri, F.: Scientific colour maps, Zenodo [dataset], https://doi.org/10.5281/zenodo.8409685, 2023.

David, P. C., Chaffe, P. L. B., Chagas, V. B. P., Dal Molin, M., Oliveira, D. Y., Klein, A. H. F., and Fenicia, F.: Correspondence Between Model Structures and Hydrological Signatures: A Large-Sample Case Study Using 508 Brazilian Catchments, Water Resour. Res., 58, https://doi.org/10.1029/2021WR030619, 2022.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, Earth Syst. Sci. Data, 17, 1461–1479, https://doi.org/10.5194/essd-17-1461-2025, 2025a.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., and Andréassian, V.: CAMELS-FR dataset [Dataset] (3), https://doi.org/10.57745/WH7FJR, 2025b.

Dolich, A., Espinoza, E. A., Ebeling, P., Guse, B., Götte, J., Hassler, S., Hauffe, C., Kiesel, J., Heidbüchel, I., Mälicke, M., Müller-Thomy, H., Stölzle, M., Tarasova, L., and Loritz, R.: CAMELS-DE: hydrometeorological time series and attributes for 1582 catchments in Germany [Dataset] (1.0.0), https://doi.org/10.5281/zenodo.13837553, 2024.

Dolich, A., Acuña Espinoza, E., and Loritz, R.: Towards Accurate Flood Predictions in Small, Fast-Responding Catchments Using Hourly CAMELS-DE Data, , https://doi.org/10.5194/egusphere-egu25-11922, 2025.

Droogers, P. and Allen, R. G.: Estimating reference evapotranspiration under inaccurate data conditions, Irrig. Drain. Syst., 16, 33–45, https://doi.org/10.1023/A:1015508322413, 2002.

European Space Agency and Airbus: Copernicus DEM, https://doi.org/10.5270/ESA-c5d3d65, 2022.

Färber, C., Plessow, H., Mischel, S., Kratzert, F., Addor, N., Shalev, G., and Looser, U.: GRDC-Caravan: extending Caravan with data from the Global Runoff Data Centre, https://doi.org/10.5194/essd-2024-427, 13 November 2024.

Federer, C. A., Vörösmarty, C., and Fekete, B.: Intercomparison of Methods for Calculating Potential Evaporation in Regional and Global Water Balance Models, Water Resour. Res., 32, 2315–2321, https://doi.org/10.1029/96WR00801, 1996.

Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: Hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth Syst. Sci. Data, 13, 3847–3867, https://doi.org/10.5194/essd-13-3847-2021, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: 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, 2009.

Hamon, W. R.: Estimating Potential Evapotranspiration, Trans. Am. Soc. Civ. Eng., 128, 324–338, https://doi.org/10.1061/TACEAT.0008673, 1963.

Hanselmann, N., Osuch, M., Wawrzyniak, T., and Alphonse, A. B.: Evaluating potential evapotranspiration methods in a rapidly warming Arctic region, SW Spitsbergen (1983–2023), J. Hydrol. Reg. Stud., 56, 101979, https://doi.org/10.1016/j.ejrh.2024.101979, 2024.

Hargreaves, G. H. and Samani, Z. A.: Estimating potential evapotranspiration, J. Irrig. Drain. Div., 108, 225–230, https://doi.org/10.1061/JRCEA4.0001390, 1982.

Helgason, H. B. and Nijssen, B.: LamaH-Ice: LArge-SaMple DAta for Hydrology and Environmental Sciences for Iceland, Earth Syst. Sci. Data, 16, 2741–2771, https://doi.org/10.5194/essd-16-2741-2024, 2024.

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., De Chiara, G., 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, J. N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth Syst. Sci. Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH) [Dataset] (0.9), https://doi.org/10.5281/zenodo.15025258, 2025.

Jenicek, M., Tyl, R., Nedelcev, O., Ledvinka, O., Šercl, P., Bernsteinová, J., and Langhammer, J.: CAMELS-CZ: A catchment attribute database for hydrological and climatological studies using a large sample of catchments, , https://doi.org/10.5194/egusphere-egu24-3872, 2024.

Jimenez, D. A., Meneses, J. E., Solha, P. H. B., Avila-Diaz, A., Quesada, B., Melo Brentan, B., and Ferreira Rodrigues, A.: CAMELS-COL: A Large-Sample Hydrometeorological Dataset for Colombia, https://doi.org/10.5194/essd-2025-200, 23 June 2025.

Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G., and Dixon, H.: CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications, Earth Syst. Sci. Data, 7, 143–155, https://doi.org/10.5194/essd-7-143-2015, 2015.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, https://doi.org/10.1029/2005WR004362, 2006.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-saMple dAta for Hydrology and environmental sciences for Central Europe, Earth Syst. Sci. Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

Klotz, D., Miersch, P., Do Nascimento, T. V. M., Fenicia, F., Gauch, M., and Zscheischler, J.: EARLS: A runoff reconstruction dataset for Europe [preprint], https://doi.org/10.5194/essd-2024-450, 3 January 2025.

Knapp, J. L. A., Li, L., and Musolff, A.: Hydrologic connectivity and source heterogeneity control concentration–discharge relationships, Hydrol. Process., 36, e14683, https://doi.org/10.1002/hyp.14683, 2022.

Knapp, J. L. A., Berghuijs, W. R., Floriancic, M. G., and Kirchner, J. W.: Catchment hydrological response and transport are affected differently by precipitation intensity and antecedent wetness, Hydrol. Earth Syst. Sci. Discuss., 1–19, https://doi.org/10.5194/hess-2024-371, 2024.

Koch, J., Liu, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., Hansen, M. F. T., and Schneider, R. J. M.: CAMELS-DK: Hydrometeorological Time Series and Landscape Attributes for 3330 Catchments in Denmark [Dataset] (6), https://doi.org/10.22008/FK2/AZXSYP, 2025.

van Kraalingen, D. W. G. and Stol, W.: Evapotranspiration modules for crop growth simulation. Implementation of the algorithms from Penman, Makkink and Priestley-Taylor, AB-DLO, Wageningen, 1997.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Sci. Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. A., and Reaney, S. M.: Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain, Hydrol. Earth Syst. Sci., 23, 4011–4032, https://doi.org/10.5194/hess-23-4011-2019, 2019.

Linsley, R. K.: Rainfall-Runoff Models—An Overview, in: Rainfall-runoff Relationship, Book-Crafters Inc., Kansas City, 3–22, 1982.

Liu, J., Koch, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., Hansen, M. F. T., and Schneider, R. J. M.: CAMELS-DK: hydrometeorological time series and landscape attributes for 3330 Danish catchments with streamflow observations from 304 gauged stations, Earth Syst. Sci. Data, 17, 1551–1572, https://doi.org/10.5194/essd-17-1551-2025, 2025.

Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S. K., Hauffe, C., Heidbüchel, I., Kiesel, J., Mälicke, M., Müller-Thomy, H., Stölzle, M., and Tarasova, L.: CAMELS-DE: hydro-meteorological time series and attributes for 1582 catchments in Germany, Earth Syst. Sci. Data, 16, 5625–5642, https://doi.org/10.5194/essd-16-5625-2024, 2024.

Mangukiya, N. K., Kumar, K. B., Dey, P., Sharma, S., Bejagam, V., Mujumdar, P. P., and Sharma, A.: CAMELS-IND: hydrometeorological time series and catchment attributes for 228 catchments in Peninsular India, Earth Syst. Sci. Data, 17, 461–491, https://doi.org/10.5194/essd-17-461-2025, 2025.

Massmann, C.: Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments, Hydrol. Process., 34, 4–20, https://doi.org/10.1002/hyp.13566, 2020.

Mavromatis, T. and Voulanas, D.: Evaluating ERA-Interim, Agri4Cast, and E-OBS gridded products in reproducing spatiotemporal characteristics of precipitation and drought over a data poor region: The Case of Greece, Int. J. Climatol., 41, 2118–2136, https://doi.org/10.1002/joc.6950, 2021.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, WIREs Water, 5, e1319, https://doi.org/10.1002/wat2.1319, 2018.

MeteoSwiss: Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD, 2021a.

MeteoSwiss: Daily Precipitation (final analysis): RhiresD, 2021b.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J. N.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

do Nascimento, T. V. M., de Oliveira, R. P., and Condesso de Melo, M.: Impacts of Large-Scale Irrigation and Climate Change on the Hydrological Cycle: The Case Study of the Alqueva Irrigation Scheme and the Gabros De Beja Aquifer System, https://doi.org/10.2139/ssrn.4514395, 2023.

do Nascimento, T. V. M., Rudlang, J., Höge, M., van der Ent, R., Chappon, M., Seibert, J., Hrachowitz, M., and Fenicia, F.: EStreams: An integrated dataset and catalogue of streamflow, hydro-climatic and landscape variables for Europe, Sci. Data, 11, 879, https://doi.org/10.1038/s41597-024-03706-1, 2024.

do Nascimento, T. V. M., Rudlang, J., Höge, M., van der Ent, R., Chappon, M., Seibert, J., Hrachowitz, M., and Fenicia, F.: EStreams: An Integrated Dataset and Catalogue of Streamflow, Hydro-Climatic Variables and Landscape Descriptors for Europe (1.4), https://doi.org/10.5281/zenodo.17598150, 2025.

Nijzink, J., Loritz, R., Gourdol, L., Zoccatelli, D., Iffly, J. F., and Pfister, L.: CAMELS-LUX: Highly Resolved Hydro-Meteorological and Atmospheric Data for Physiographically Characterized Catchments around Luxembourg, https://doi.org/10.5194/essd-2024-482, 26 May 2025.

Oldham, L. D., Freer, J., Coxon, G., Howden, N., Bloomfield, J. P., and Jackson, C.: Evidence-based requirements for perceptualising intercatchment groundwater flow in hydrological models, Hydrol. Earth Syst. Sci., 27, 761–781, https://doi.org/10.5194/hess-27-761-2023, 2023.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, J. Hydrol., 303, 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Peral García, C., Navascués Fernández-Victorio, B., and Ramos Calzado, P.: Serie de precipitatión diaria en rejilla con fines climáticos - Nota técnica 24 de AEMET, Agencia Estatal de Meteorología, Madrid, 2017.

Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, I. G., and Gustafsson, D.: Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide?, Water Resour. Res., 59, e2022WR033447, https://doi.org/10.1029/2022WR033447, 2023.

Quintana-Seguí, P., Moigne, P. L., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France, J. Appl. Meteorol. Climatol., 47, 92–107, https://doi.org/10.1175/2007JAMC1636.1, 2008.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteorol. Z., 22, 235–256, https://doi.org/10.1127/0941-2948/2013/0436, 2013.

Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M., and Deutschländer, T.: New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS), Theor. Appl. Climatol., 142, 1531–1553, https://doi.org/10.1007/s00704-020-03388-w, 2020.

Robinson, E. L., Blyth, E., Clark, D. B., Comyn-Platt, E., Finch, J., and Rudd, A. C.: Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015) [CHESS-PE] [dataset]. NERC Environmental Information Data Centre, https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7, 2016.

Robinson, E. L., Blyth, E., Clark, D. B., Comyn-Platt, E., Finch, J., and Rudd, A. C.: Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961-2015) [CHESS-met] v1.2 [dataset]. NERC Environmental Information Data Centre, https://doi.org/10.5285/b745e7b1-626c-4ccc-ac27-56582e77b900, 2017a.

Robinson, E. L., Blyth, E. M., Clark, D. B., Finch, J., and Rudd, A. C.: Trends in atmospheric evaporative demand in Great Britain using high-resolution meteorological data, Hydrol. Earth Syst. Sci., 21, 1189–1224, https://doi.org/10.5194/hess-21-1189-2017, 2017b.

Salwey, S., Coxon, G., Pianosi, F., Singer, M. B., and Hutton, C.: National-Scale Detection of Reservoir Impacts Through Hydrological Signatures, Water Resour. Res., 59, e2022WR033893, https://doi.org/10.1029/2022WR033893, 2023.

Scharling, M.: Klimagrid Danmark – Nedbør, lufttemperatur og potentiel fordampning 20 × 20 & 40 × 40 km – Metodebeskrivelse, Danish Meteorological Institute, Copenhagen, 1999a.

Scharling, M.: Klimagrid Danmark Nedbør 10 × 10 km (ver. 2) – Metodebeskrivelse, Danish Meteorological Institute, Copenhagen, 1999b.

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215–224, https://doi.org/10.5194/hess-4-215-2000, 2000.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, 2012.

Seibert, J., Vis, M. J. P., Lewis, E., and van Meerveld, H. J.: Upper and lower benchmarks in hydrological modelling, Hydrol. Process., 32, 1120–1125, https://doi.org/10.1002/hyp.11476, 2018.

Senent-Aparicio, J., Castellanos Osorio, G., Segura Méndez, F. J., Jimeno Sáez, P., Adrian, L. B., and Julio, P. S.: BULL Database – Spanish Basin attributes for Unraveling Learning in Large-sample hydrology [Dataset] (1), https://doi.org/10.5281/zenodo.10605646, 2024a.

Senent-Aparicio, J., Castellanos-Osorio, G., Segura-Méndez, F., López-Ballesteros, A., Jimeno-Sáez, P., and Pérez-Sánchez, J.: BULL Database – Spanish Basin attributes for Unravelling Learning in Large-sample hydrology, Sci. Data, 11, 737, https://doi.org/10.1038/s41597-024-03594-5, 2024b.

SMHI: Nedladdning av griddad nederbörd- och temperaturdata: https://www.smhi.se/data/ladda-ner-data/griddade-nederbord-och-temperaturdata-pthbv, last access: 19 June 2025.

Sperna Weiland, F. C., Tisseuil, C., Dürr, H. H., Vrac, M., and van Beek, L. P. H.: Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study, Hydrol. Earth Syst. Sci., 16, 983–1000, https://doi.org/10.5194/hess-16-983-2012, 2012.

Štěpánek, P., Zahradníček, P., and Huth, R.: Interpolation techniques used for data quality control and calculation of technical series: an example of a Central European daily time series, Idojaras, 115, 87–98, https://www.met.hu/en/ismeret-tar/kiadvanyok/idojaras/index.php?id=34, 2011.

Štěpánek, P., Zahradníček, P., and Farda, A.: Experiences with data quality control and homogenization of daily records of various meteorological elements in the Czech Republic in the period 1961–2010, Idojaras, 117, 123–141, https://www.met.hu/en/ismeret-tar/kiadvanyok/idojaras/index.php?id=82, 2013.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., and Keller, V. D. J.: Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2015) [CEH-GEAR] [dataset]. NERC Environmental Information Data Centre, https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca, 2016.

Tarek, M., Brissette, F. P., and Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, Hydrol. Earth Syst. Sci., 24, 2527–2544, https://doi.org/10.5194/hess-24-2527-2020, 2020.

Teutschbein, C.: CAMELS-SE: Long-term hydroclimatic observations (1961–2020) across 50 catchments in Sweden as a resource for modelling, education, and collaboration, Geosci. Data J., gdj3.239, https://doi.org/10.1002/gdj3.239, 2024a.

Teutschbein, C.: Hydroklimatiska förhållanden i Sverige 1961-2020 - Nederbörd, temperatur och avrinningsobservationer i 50 avrinningsområden (CAMELS-SE) [Dataset], https://doi.org/10.57804/t3rm-v029, 2024b.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, https://doi.org/10.1002/joc.2003, 2010.

Viviroli, D., Gurtz, J., and Zappa, M.: The Hydrological Modelling System PREVAH, Geographica Bernensia, CH, 2007.

Wilcoxon, F.: Individual comparisons by ranking methods, Biom. Bull, 1, 80–83, 1945.

Yu, M., Chen, X., Li, L., Bao, A., and Paix, M. J. de la: Streamflow Simulation by SWAT Using Different Precipitation Sources in Large Arid Basins with Scarce Raingauges, Water Resour. Manag., 25, 2669–2681, https://doi.org/10.1007/s11269-011-9832-z, 2011.

12 Appendix

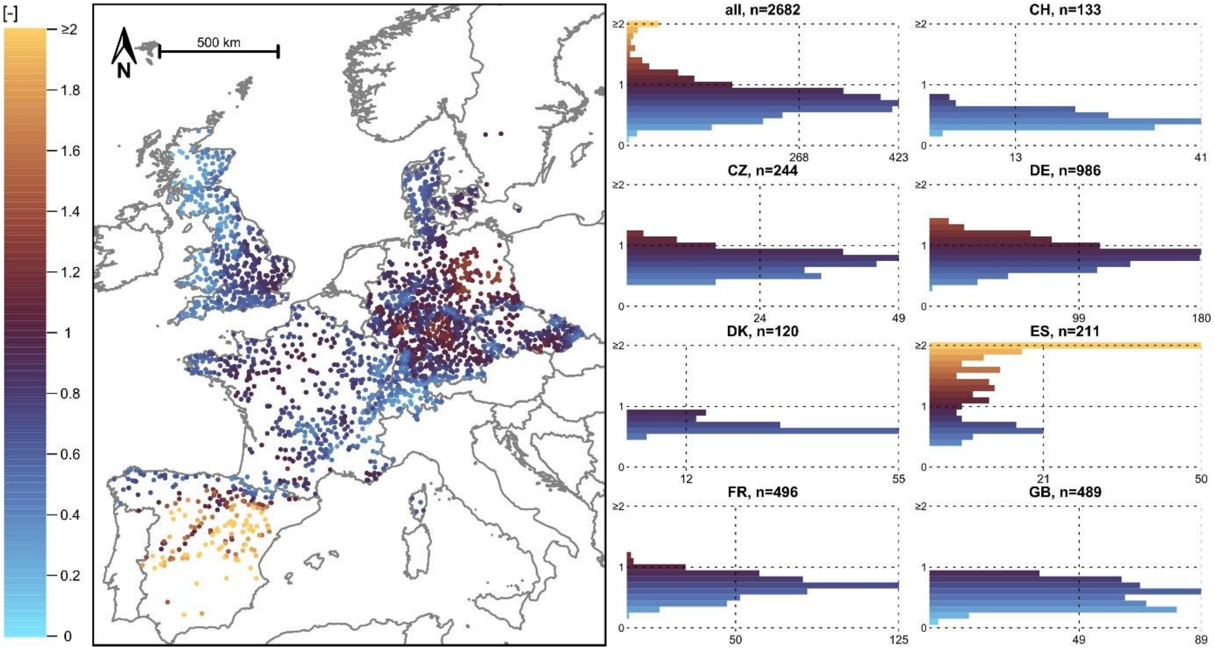

Figure A1: Aridity index (E_{pot}/P) calculated from the CAMELS data (for a 20-year period: 1995-2015). Note that the colour scale was cut at a value of two.

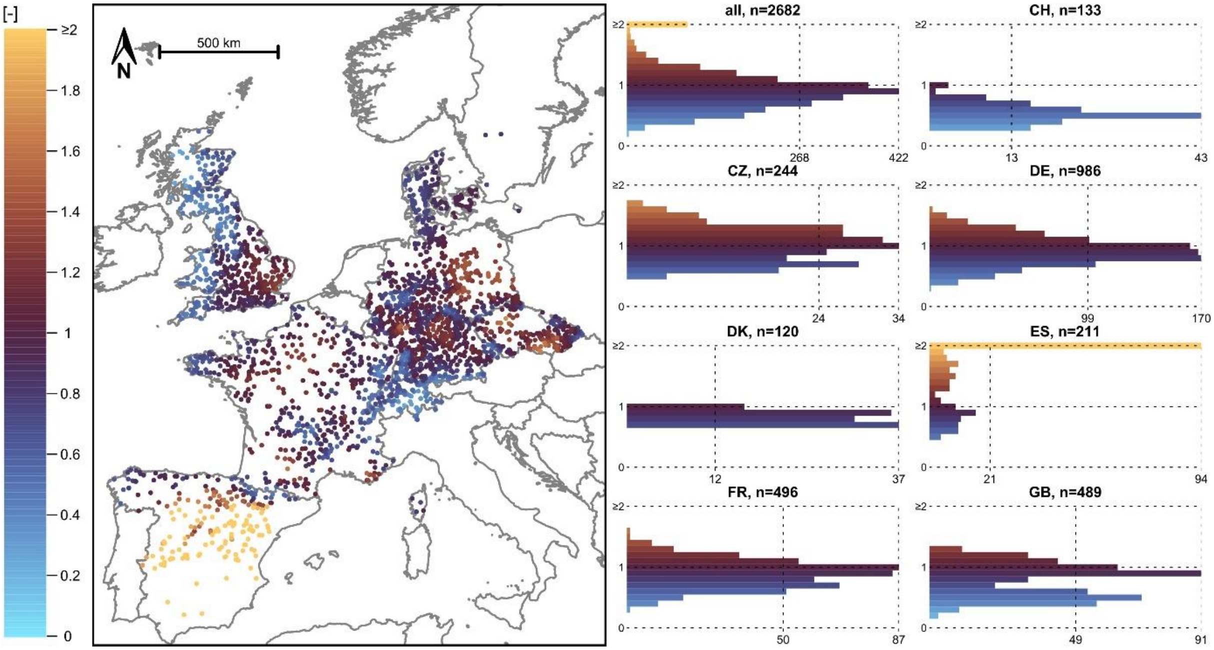

Figure A2: Aridity index (E_{pot}/P) calculated from the E-OBS data obtained from EStreams (for a 20-year period: 1995-2015). Note that the colour scale was cut at a value of two.

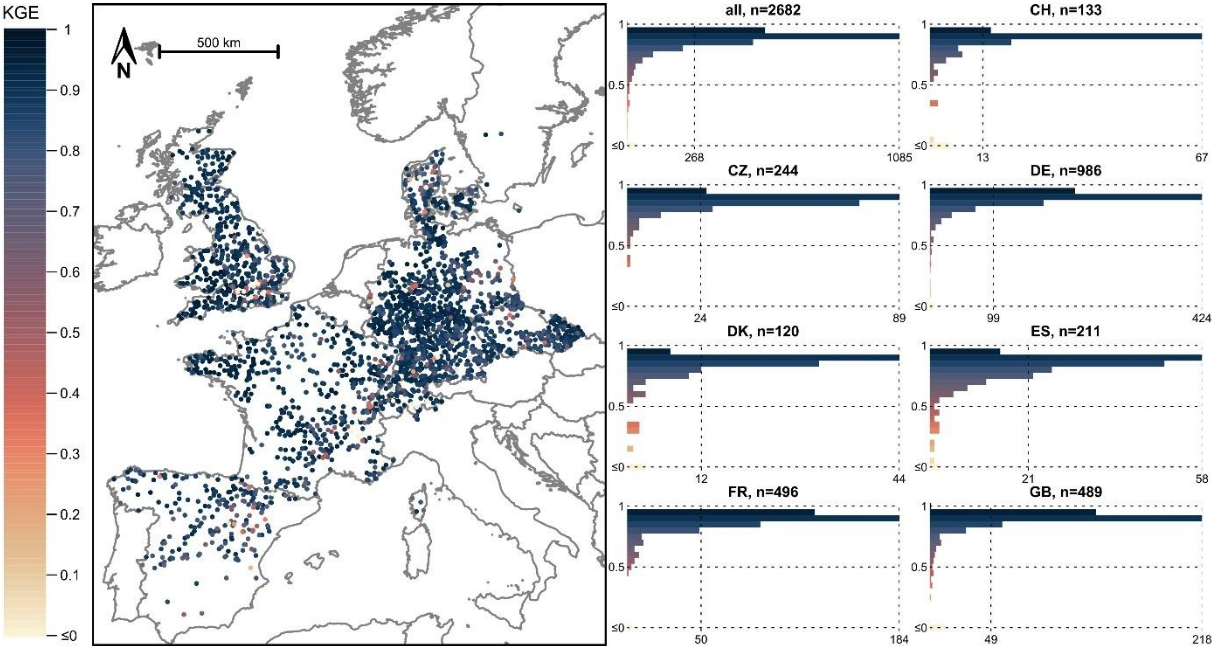

Figure A3: Model performance (Kling-Gupta efficiency, KGE) achieved for the 20-year period between October 1995 and September 2015 when the input data from the CAMELS datasets were used for model calibration (scenario I). Note that the lower limit of the colour scale was cut at zero. Lower performances were plotted last to improve their visibility.

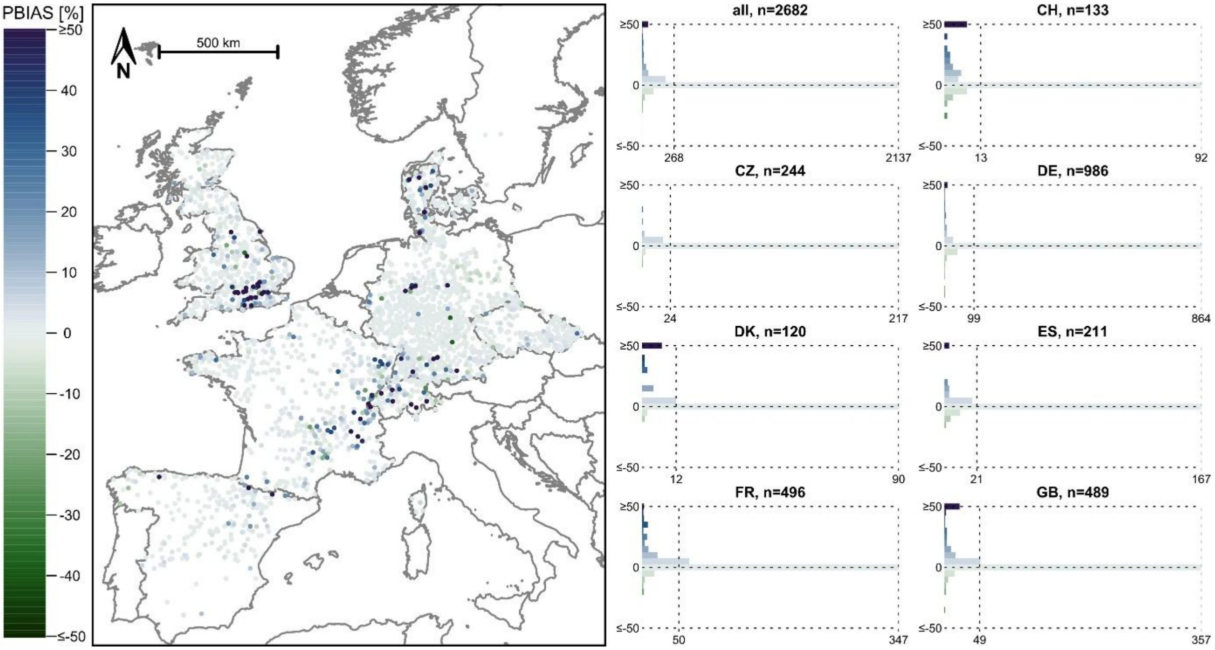

Figure A4: PBIAS (relative deviation of the simulated streamflow from the observed streamflow) for the 20-year period between October 1995 and September 2015 when the E-OBS data obtained from EStreams were used for model calibration (scenario II). Note that the limits of the colour scale were cut at ±50 %. Largest deviations were plotted last to improve their visibility.

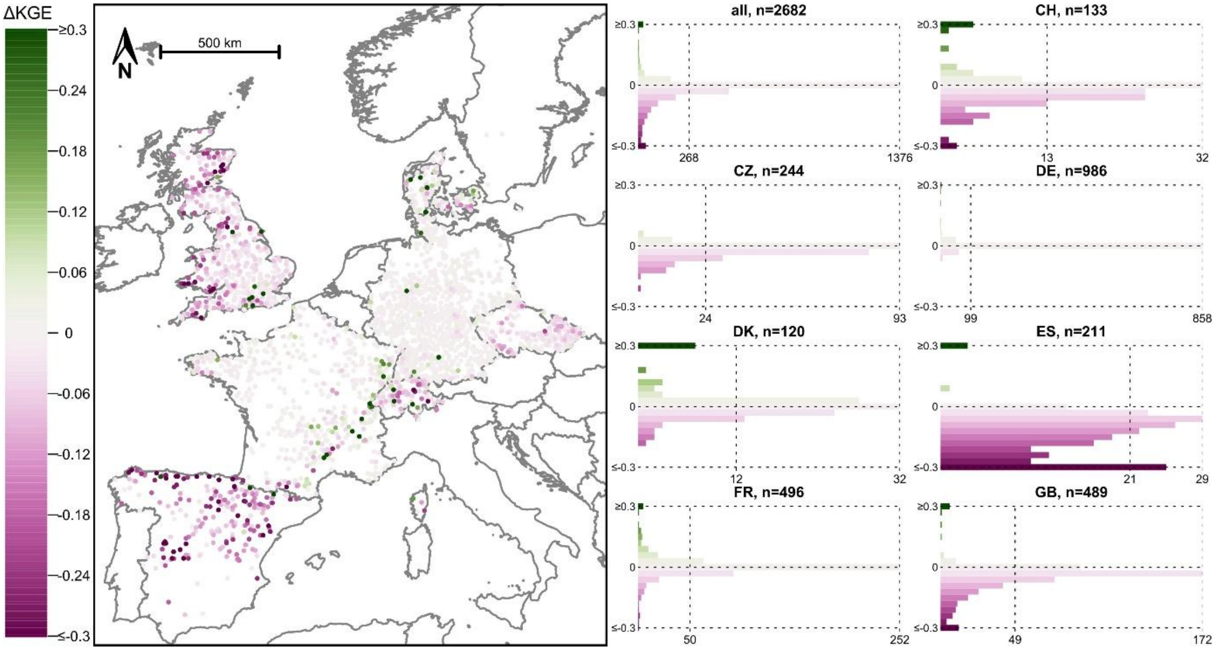

Figure A5: Difference in model performance when all meteorological input data were obtained from E-OBS (i.e., EStreams, scenario II) and when the precipitation data from E-OBS were replaced with those from CAMELS (scenario III). Positive values and green colours indicate higher model performances with the precipitation data from E-OBS, negative values and pink colours indicate higher model performances with the precipitation data from CAMELS. Note that the colour scale was cut at a difference in KGE of ±0.3. The catchments with the largest differences in model performance were plotted last to increase their visibility.

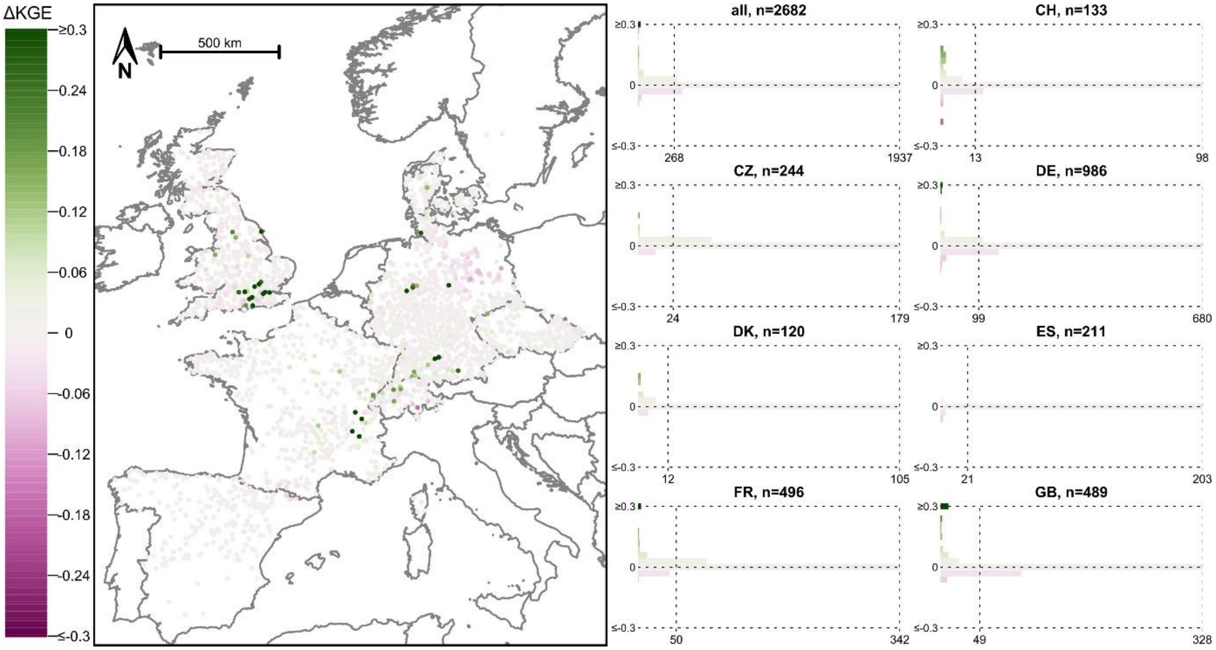

Figure A6: Difference in model performance when all meteorological input data were obtained from E-OBS (i.e., EStreams, scenario II) and when the E_{pot} data from E-OBS were replaced with those from CAMELS (scenario IV). Positive values and green colours indicate higher model performances with the E_{pot} data from E-OBS, negative values and pink colours indicate higher model performances with the E_{pot} data from CAMELS. Note that the colour scale was cut at a difference in KGE of ±0.3. The catchments with the largest differences in model performance were plotted last to increase their visibility.