



# Evaluating the quality of the E-OBS meteorological forcing data in EStreams for large-sample hydrology studies in Europe

Franziska Clerc-Schwarzenbach<sup>\*1</sup>, Thiago V. M. do Nascimento<sup>\*1,2</sup>

<sup>1</sup>Department of Geography, University of Zurich, Zurich, 8057, Switzerland

5 <sup>2</sup>Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, 8600, Switzerland

<sup>\*</sup>These authors contributed equally to this work.

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

**Abstract.** To conduct large-sample hydrological studies over large spatial domains, standardized 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 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 3423 catchments in EStreams to nine 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 (slightly) lower when the E-OBS data are used than when the CAMELS datasets are used for calibration. Exceptions arise when the CAMELS data were derived from global datasets rather than national products, as well as 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 too: Llamas (Helgason and Nijssen, 2024; Klingler et al., 2021a), 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 standardized 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 standardized 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 inclusion of an increasing number of catchments in one dataset almost always goes hand in hand with difficulties in providing high-quality forcing data, due to the lower availability of high-quality data for larger spatial extents. 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 standardized 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, than 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 the potential data quality losses when it comes to standardized meteorological data.

Similar to the standardization in Caravan, the meteorological data were also standardized 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 arised 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 for 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 standardized 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 3423 catchments from nine European countries to assess the quality 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 LamaH-CE (Klingler et al., 2021a) for Austria, CAMELS-DK (Liu et al., 2025) for Denmark, CAMELS-FR (Delaique 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).

Except for LamaH-CE (which includes globally available ERA5-Land data), the meteorological data in the smaller datasets stem from sources that were created specifically for the respective country. Following the approach of Clerc-Schwarzenbach et al. (2024), 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 meteorological forcing data. We hypothesized that (i) the smaller datasets would provide a higher model performance than the standardized one; and that (ii) the difference in model performance introduced by the different forcing data would be smaller where E-OBS gauge densities are higher.

## 2 Data and Methods

### 2.1 Subset of catchments

We conducted this study for 3423 catchments that are available in the EStreams dataset and catalogue and in one of the following datasets: LamaH-CE, 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 that were still included after each step in brackets):

- 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<sup>2</sup> [9115]
- E-OBS meteorological time series available between October 1990 and September 2015 [8525]
- No redundancy among the catchments (the catchment with the longer streamflow time series was kept) [8367]
- Availability of at least 90 % of the  $E_{\text{pot}}$  data between October 1990 and September 2015 [8297]
- Availability of the catchment in one of the CAMELS datasets [4246]
- Availability of at least 90 % of the streamflow data (in the CAMELS dataset) between October 1995 and September 2015 [3614]
- Average streamflow (in the CAMELS dataset) between October 1995 and September 2015 below 10 mm d<sup>-1</sup> [3608]
- Runoff ratio (in the CAMELS dataset) between October 1995 and September 2015 below 1.1 [3563]
- Runoff ratio (in EStreams) between October 1995 and September 2015 below 1.1 [3423]



Note that no catchments were excluded due to a high human influence. While identifying such influences for a large number of catchments is possible, it remains a challenging task (Klotz et al., 2025; Senent-Aparicio et al., 2024b). Given these difficulties and the fact that such influences will affect model performances with all types of forcing data, we chose not to apply this filter.

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

| Country       | Number of catchments included in this study | CAMELS dataset | Publication                    | Dataset                         |
|---------------|---|----------------|--------------------------------|---------------------------------|
| Austria       | 428   | LamaH-CE       | Klingler et al. (2021a)        | Klingler et al. (2021b)         |
| Czechia       | 297   | CAMELS-CZ      | <i>unpublished</i>             | <i>unpublished</i>              |
| Denmark       | 139   | CAMELS-DK      | Liu et al. (2025)              | Koch et al. (2025)              |
| France        | 515   | CAMELS-FR      | Delaigue et al. (2025a)        | Delaigue et al. (2025b)         |
| Germany       | 1054  | CAMELS-DE      | Loritz et al. (2024)           | Dolich et al. (2024)            |
| Great Britain | 560   | CAMELS-GB      | Coxon et al. (2020a)           | Coxon et al. (2020b)            |
| Spain         | 245   | BULL           | Senent-Aparicio et al. (2024b) | Senent-Aparicio et al., (2024a) |
| Sweden        | 23  | CAMELS-SE      | Teutschbein (2024a)            | Teutschbein (2024b)             |
| Switzerland   | 162   | 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,  $E_{\text{pot}}$ , and temperature data from the EStreams dataset and from the different CAMELS datasets (Table 1). In EStreams, precipitation and temperature data were obtained from the E-OBS ensemble mean product with a spatial resolution of  $0.25^\circ$  in both latitude and longitude (do Nascimento et al., 2024). 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 ( $E_{\text{pot}}$ ) 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. EStreams is a ready-to-use product derived from E-OBS and is likely to be increasingly used by the LSH community for studies on European catchments. Therefore, we used the meteorological data of precipitation,  $E_{\text{pot}}$ , and temperature as provided in EStreams for the evaluation of the E-OBS meteorological data. Note that there is also a version of E-OBS at a resolution of  $0.1^\circ$  available, but not represented in EStreams. Similarly, there are different  $E_{\text{pot}}$  products available from E-OBS, but here, we used the  $E_{\text{pot}}$  product provided in EStreams. The CAMELS meteorological data are usually based on in-situ observations, but some datasets also incorporate satellite or reanalysis information at several resolutions. 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.

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

| Country       | Variable(s)            | Source / equation  | Resolution | Reference(s)  |
|---------------|------------------------|--|------------|---|
| Austria       | $P, T$                 | ERA5-Land  | 9 km       | Muñoz-Sabater et al. (2021)   |
|               | $E_{\text{pot}}$       | Thornthwaite   | *          | Thornthwaite and Mather (1957)  |
| Czechia       | $P, T$                 | <i>unpublished</i> data from Czech Hydrometeorological Institute (M. Jeníček / O. Ledvinka, pers. comm.) | 500 m      | Štěpánek et al. (2011, 2013)  |
|               | $E_{\text{pot}}$       | <i>unpublished</i> data based on Oudin equation (M. Jeníček / O. Ledvinka, pers. comm.)                  | *          | Oudin et al. (2005)   |
| Denmark       | $P$                    | Danish Meteorological Inst.  | 10 km      | Scharling (1999b)   |
|               | $T$                    | Danish Meteorological Inst.  | 20 km      | Scharling (1999a)   |
|               | $E_{\text{pot}}$       | Makkink  | 40 km      | van Kraalingen and Stol (1997)  |
| France        | $P, T$                 | SAFRAN by Météo-France   | 8 km       | Quintana-Seguí et al. (2008); Vidal et al. (2010)                           |
|               | $E_{\text{pot}}$       | Oudin  | *          | Oudin et al. (2005)   |
| Germany       | $P$                    | HYRAS by Deutscher Wetterdienst (DWD)  | 1 km       | Rauthe et al. (2013)  |
|               | $T$                    | HYRAS by DWD   | 5 km       | Razafimaharo et al. (2020)  |
|               | $E_{\text{pot}}$       | Modified Hargreaves  | *          | Adam et al. (2006); Droogers and Allen (2002); Hargreaves and Samani (1982) |
| Great Britain | $P$                    | CEH-GEAR   | 1 km       | Keller et al. (2015); Tanguy et al. (2016)                                  |
|               | $T$                    | CHESS-met  | 1 km       | Robinson et al. (2017a)   |
|               | $E_{\text{pot}}$       | CHESS-PE (based on Penman-Monteith)  | 1 km       | Robinson et al. (2016, 2017b)   |
| Spain         | $P, T, E_{\text{pot}}$ | Spanish Meteorological Agency (AEMET)  | 0.05 °     | Peral García et al. (2017)  |
| Sweden        | $P, T$                 | PTHBV database by Swedish Meteorological and Hydrological Institute (SMHI)                               | 4 km       | SMHI (2025)   |
|               | $E_{\text{pot}}$       | <i>unpublished</i> data based on Hamon equation (C. Teutschbein, pers. comm.)                            | *          | Hamon (1963)  |
| Switzerland   | $P$                    | RhiresD by MeteoSwiss  | 2 km       | MeteoSwiss (2021b)  |
|               | $T$                    | TabsD by MeteoSwiss  | 2 km       | MeteoSwiss (2021a)  |
|               | $E_{\text{pot}}$       | 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,  $E_{\text{pot}}$ , 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  $E_{\text{pot}}$ , 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  $E_{\text{pot}}$ ), 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

We defined different combinations of forcing data (“scenarios”) to calibrate the hydrological model (Table 3). This allowed us to determine how the forcing data from E-OBS and from CAMELS individually impacted hydrological model performance. Note that due to the dependency of the E-OBS  $E_{\text{pot}}$  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 (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) 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    | $E_{\text{pot}}$ | 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 $E_{\text{pot}}$ | E-OBS (EStreams) | CAMELS           | E-OBS (EStreams) |
| V   | E-OBS with CAMELS temperature      | E-OBS (EStreams) | E-OBS (EStreams) | CAMELS           |

### 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 3042 catchments (89 %). For 1506 catchments (44 %), the deviation of the mean annual precipitation sums in E-OBS from the ones in CAMELS exceeded -10 %. Conversely, there were only 57 catchments (2 %) 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).

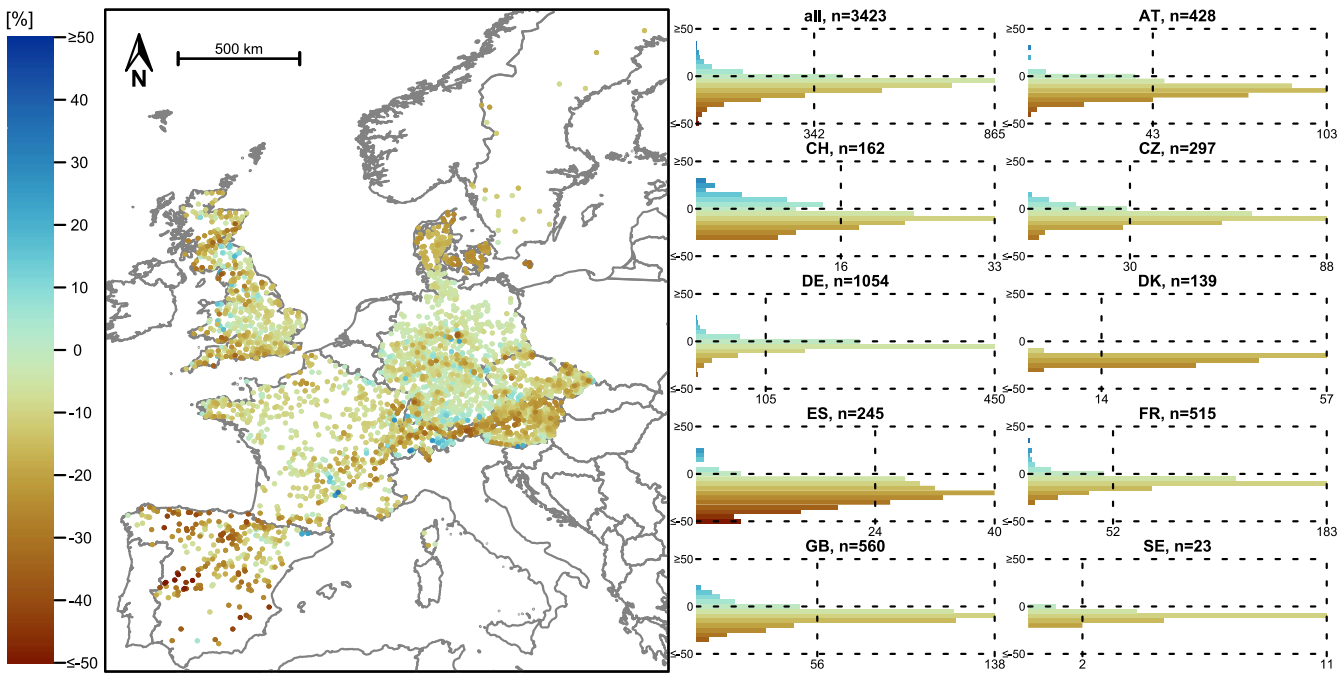
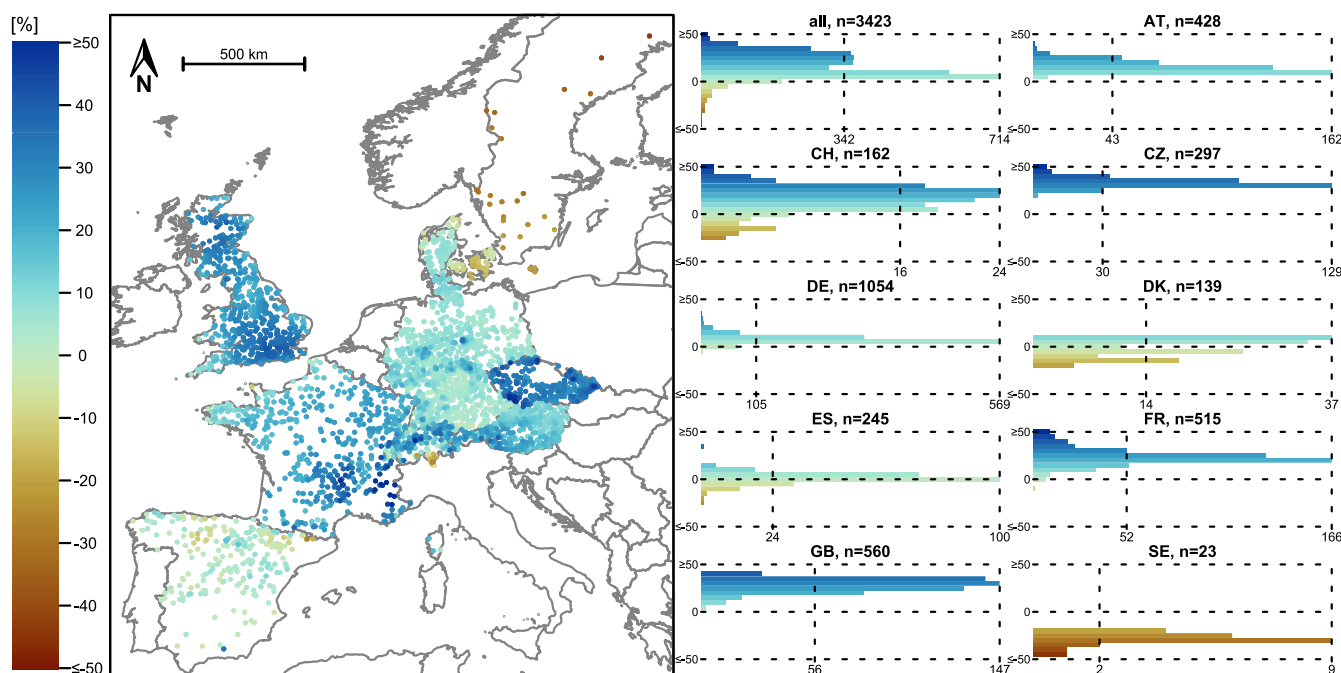


Figure 1: Relative difference in the mean annual precipitation (for a 20-year period: 1995-2015) between the E-OBS data obtained from EStreams and the different CAMELS datasets. Negative values and brown colours indicate less precipitation in E-OBS than in CAMELS, positive values and blue colours more precipitation in E-OBS. On the map, the catchments with the largest deviations were plotted last to increase their visibility. Note that 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  $\pm 50$  %. The scale bar refers to the map centre. 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  $E_{\text{pot}}$ : For 3215 catchments (94 %), the mean annual  $E_{\text{pot}}$  calculated from the E-OBS data was higher than for the CAMELS data. For 2012 catchments (59 %), the deviation of the E-OBS  $E_{\text{pot}}$  sums from the CAMELS  $E_{\text{pot}}$  sums were at least 10 %. Lower  $E_{\text{pot}}$  sums derived from E-OBS than from CAMELS could only be observed for catchments in Sweden, on the Danish islands, in the canton of Ticino in 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  $E_{\text{pot}}$  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 or Austria). It is noteworthy that for  $E_{\text{pot}}$ , there tend to be small differences between the two datasets for Spain (while this was not the case for precipitation).

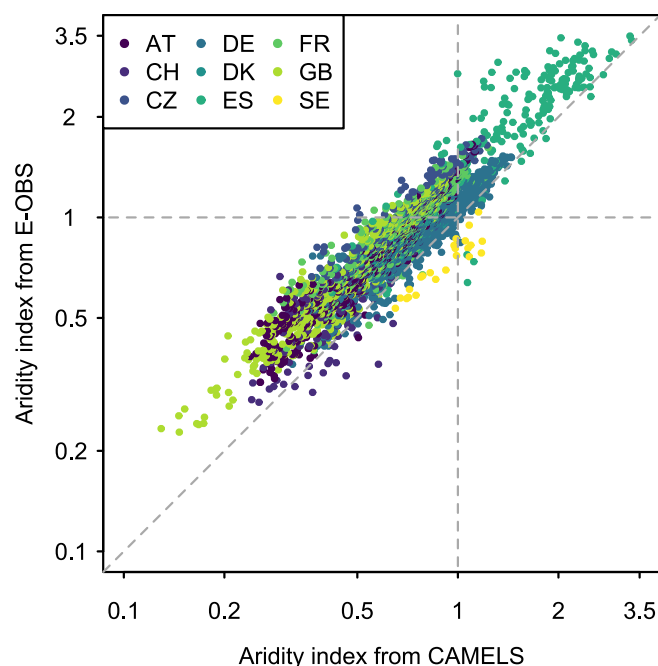


**Figure 2: Relative difference in the mean annual  $E_{\text{pot}}$  (for a 20-year period: 1995–2015) obtained from EStreams and calculated from the E-OBS data compared to the mean annual  $E_{\text{pot}}$  calculated from the different CAMELS datasets. Negative values and brown colours indicate a lower  $E_{\text{pot}}$  in E-OBS than in the CAMELS datasets, positive values and blue colours a higher  $E_{\text{pot}}$ . 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  $E_{\text{pot}}$  data, the aridity indices ( $E_{\text{pot}}/P$ ) calculated from the two data sources differed, although they were still highly correlated (Pearson's correlation coefficient of 0.93) (Fig. 3). Given the lower precipitation and higher  $E_{\text{pot}}$  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  $E_{\text{pot}}$  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 ( $E_{\text{pot}}/P$ ) 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:  $0.3^{\circ}\text{C}$ ) for the vast majority of catchments than the average temperature in CAMELS (Fig. A3). There were 537 catchments (16 %) 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 actually the differences in  $E_{\text{pot}}$  which are highly affected by the temperature data used as input to the calculations (Fig. 2Figure 2).

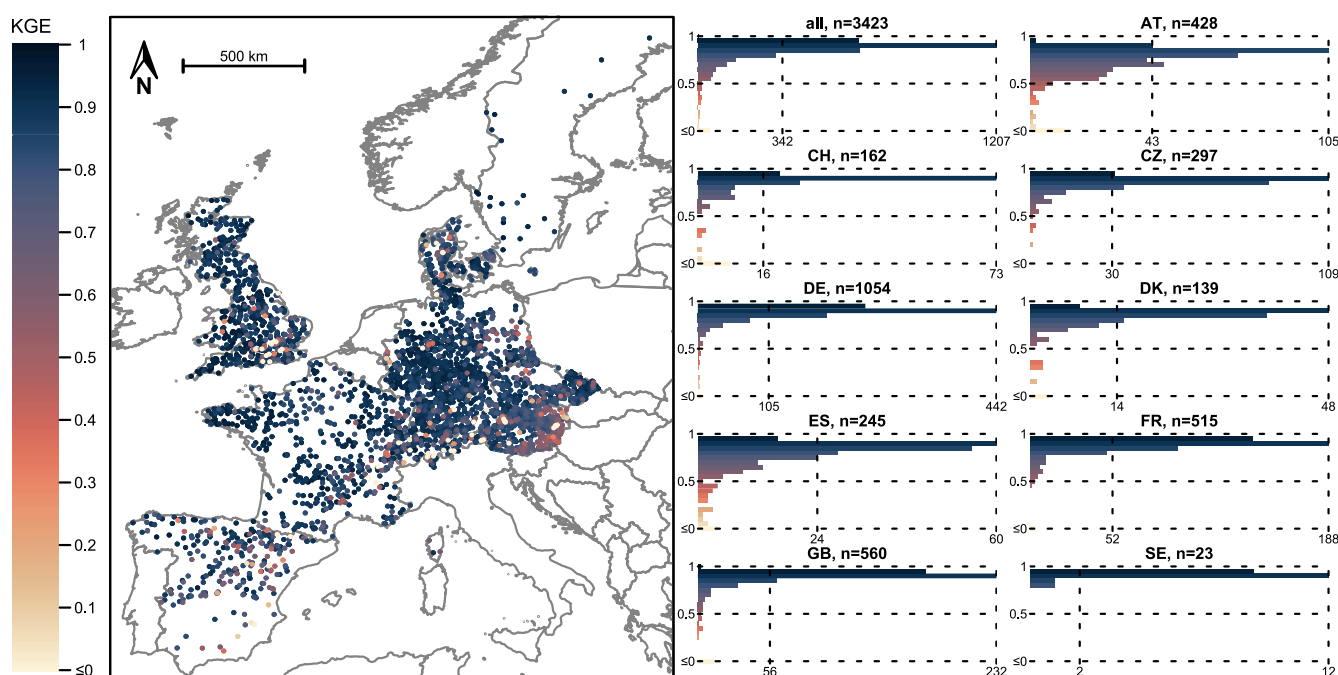
## 3.2 Model performances

### 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. 4). For 3066 of the 3423 catchments (90 %) the KGE was higher than 0.70, with a median performance



of 0.88. Among the nine countries, the median KGE was the highest for the catchments in Sweden (0.92, 100 % of the catchments had a KGE above 0.70), followed by Great Britain (median 0.91, 93 % above 0.70) and France (median 0.90, 94 % above 0.70).  
 Conversely, the model performances were the lowest for catchments in Austria and Spain. The median KGE for Austria was 0.78; only 69 % of the catchments achieved a KGE above 0.70. The pattern was similar for Spain, with a median KGE of 0.83 and 78 % of the catchments having a KGE above 0.70. However, there were also other regions in which clusters of catchments with low model performances could be observed, such as the karstic area around London and some parts of south-eastern Switzerland.



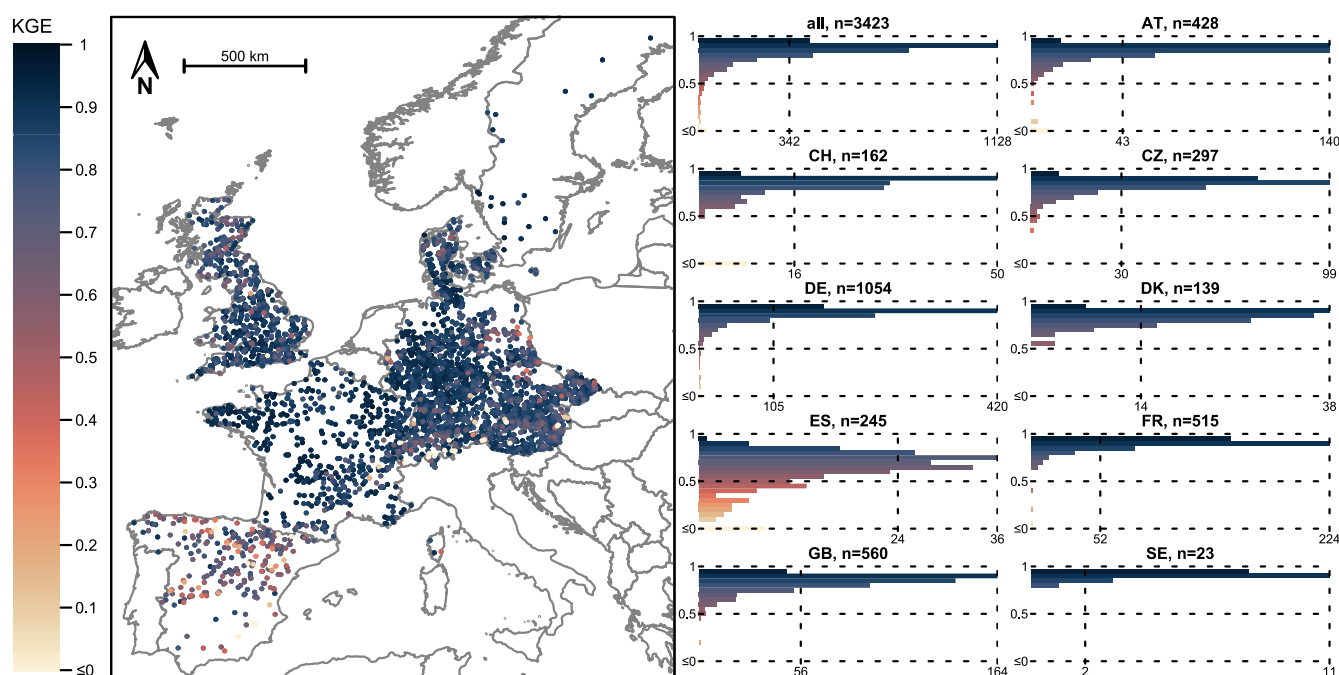
**Figure 4: Model performance (Kling-Gupta efficiency, KGE) achieved for the 20-year period between October 1995 and September 2015 when the input data from the CAMELS datasets were used for model calibration (scenario I). Note that the lower limit of the colour scale was cut at zero. Lower performances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

The model performances were also high for most catchments when the E-OBS forcing data (scenario II) were used for model calibration (Fig. 5). For 3059 of the catchments (90 %) the KGE was higher than 0.70, which is comparable to the 3066 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 one achieved with the CAMELS data (scenario I) (0.88). The median performance among all countries was still the highest for the catchments in Sweden (0.91).  
 Despite these overall similarities, there were also some notable differences in scenario II (E-OBS data) compared to scenario I (CAMELS data). The second-highest median KGE (0.91) was recorded for the catchments in France, where 97 % of the





catchments had a KGE above 0.70. After France, Germany followed with a median KGE of 0.88 and 95 % of the catchments  
 230 having a KGE above 0.70. For the catchments in Austria, where the precipitation data were obtained from ERA5-Land, clearly  
 higher model performances were achieved with the E-OBS data (scenario II) than with the CAMELS data (scenario I). The  
 median KGE in scenario II was 0.86 (scenario I: 0.78), and for 90 % of the Austrian catchments, the KGE was above 0.70  
 (scenario I: 69 %). However, for the catchments in Spain, the model performances were even lower than in scenario I: The  
 median KGE was 0.66 (scenario I: 0.83), and for only 40 % of the catchments, the KGE was above 0.70 (scenario I: 78 %).  
 235 Furthermore, for the catchments in the karstic area around London, the performances were higher with the E-OBS data (sce-  
 nario II) than with the CAMELS data (scenario I).



240 **Figure 5: 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.**

### 3.2.2 Differences in model performance between scenario II and I

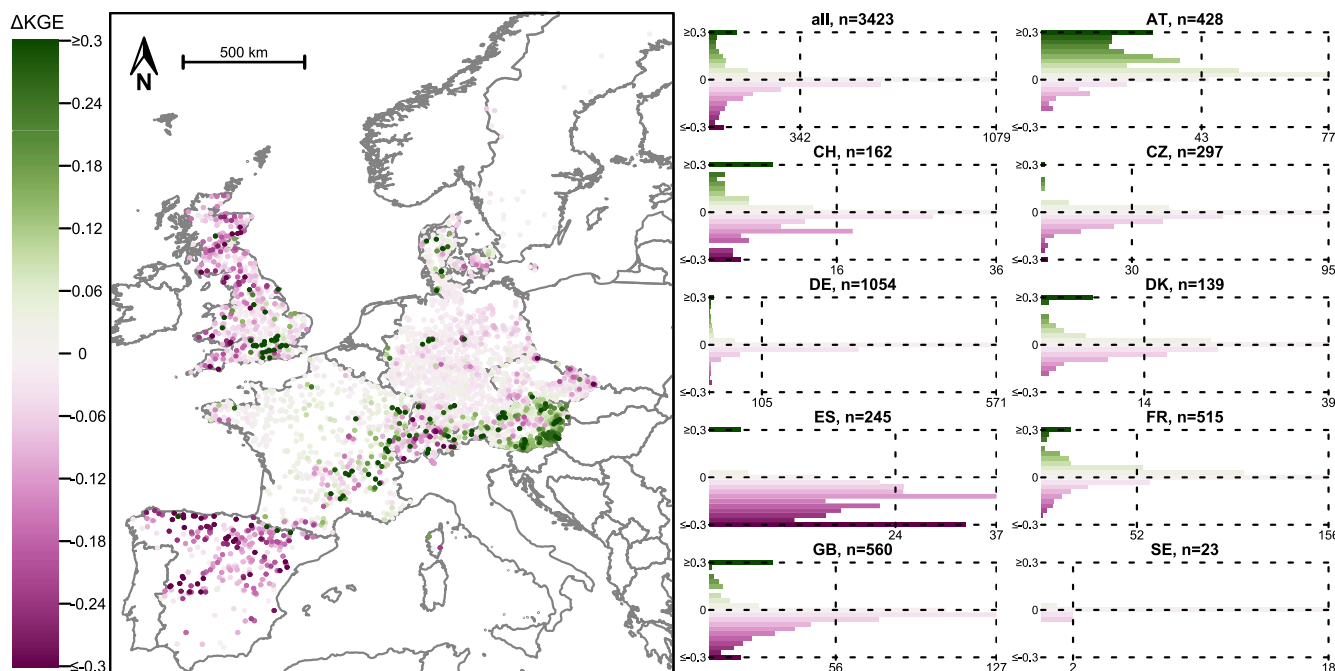
To more easily assess the differences in model performances between scenario II and I, we looked at the differences in model  
 245 performance directly (Fig. 6). For 61 % of the catchments, model performances were (at least slightly) higher when the CAM-  
 ELS data were used (scenario I), while for the other 39 % of all catchments, the use of E-OBS data resulted in better model  
 performances. Not surprisingly, based on the results described above, the strongest regional differences were found for the



catchments in Austria, Spain, and Great Britain. When the catchments in Austria were excluded from the analysis (i.e., to only compare the E-OBS data with forcing data from smaller-scale datasets) the results were more pronounced: Then, 33 % of the catchments performed better with the E-OBS forcing data, while for 67 % of the catchments, model performances were higher with the CAMELS forcing data.

For two countries (Austria and France) there were notable improvements in model performance when using the E-OBS dataset. In Austria, 353 catchments (82 %) performed better with E-OBS forcing data obtained from EStreams ( $\Delta KGE > 0$ ), and the negative differences were not more pronounced than -0.15 (i.e., for the catchments with better performances in scenario I than in scenario II, the benefit was limited). In France, 331 catchments (62%) had higher performances with the E-OBS data, but the differences in performance were variable. For Sweden, 16 catchments (72 %) performed better with the E-OBS data, but the differences were very small (Fig. 6). For the catchments in Spain, it was the opposite: 220 catchments (92 %) performed better with the CAMELS dataset, and for some catchments, the differences were large. Similar trends were observed for the catchments in Great Britain (74 %), Germany (72 %), Switzerland (67 %), Czechia (62 %), and Denmark (57 %).

The results also indicated some interesting intracountry patterns (Fig. 6). In France, the most striking positive differences occurred for the catchments in the eastern 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. For the catchments in the center of Austria, the CAMELS data sometimes led to better model performances than the E-OBS data, while the opposite was the case in most other catchments (see above).



**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 5. The catchments with the largest differences (in absolute terms) were plotted last to improve their visibility. Note that the colour scale was cut at  $\pm 0.3$ .**

### 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. A4). 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), was the performance gap between scenarios II and I notably larger than between scenarios II and III. The opposite occurred for very few catchments (Fig. A5).

The effect of replacing the  $E_{\text{pot}}$  data was quite limited (scenario IV). The higher  $E_{\text{pot}}$  data based on E-OBS were beneficial for a handful of catchments ( $\Delta\text{KGE} > 0.30$  for 21 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 only five catchments for which it decreased the KGE by more than -0.10 (Fig. A7). 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  $E_{\text{pot}}$ .

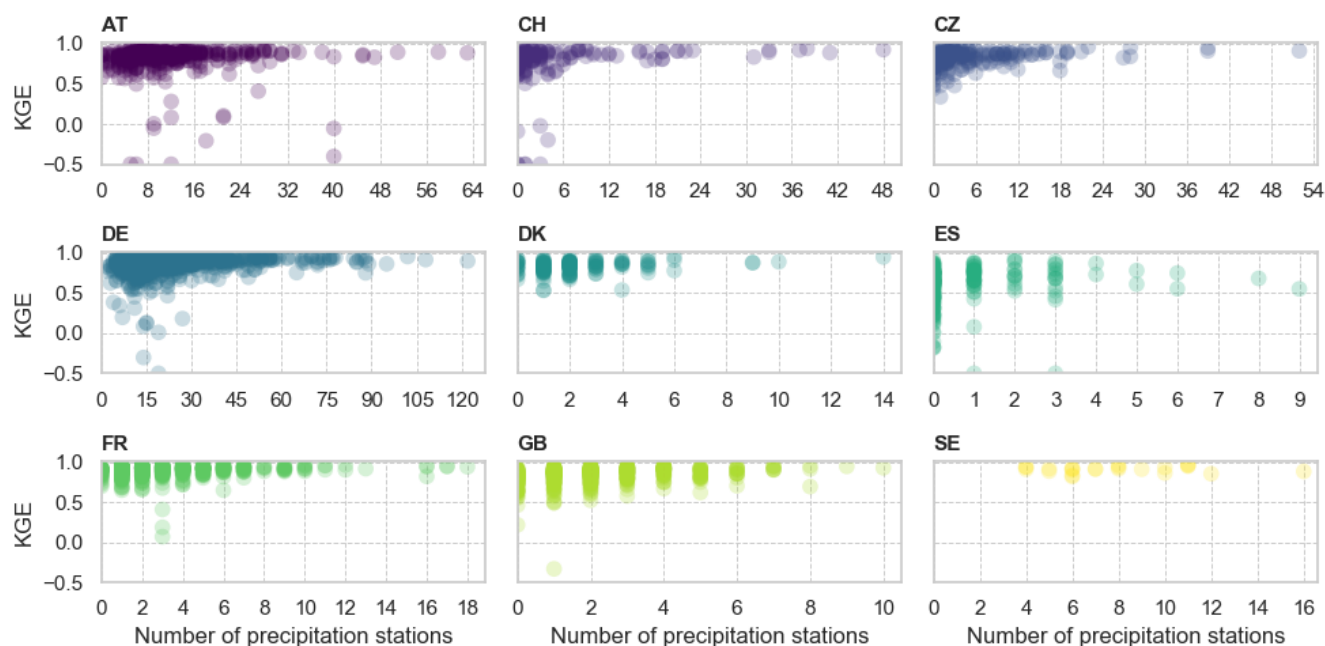


### 3.3 Model performance linked to catchment attributes

#### 3.3.1 Number of E-OBS precipitation stations

It is well documented that the coverage of precipitation stations used to produce the gridded E-OBS dataset is highly variable (Bandhauer et al., 2022; do Nascimento et al., 2024). To assess the impact of this variability, 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. 5). 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. The performances were generally the lowest for areas with sparse station coverage, such as Spain and Great Britain. 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 further 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 (Fig. A8).



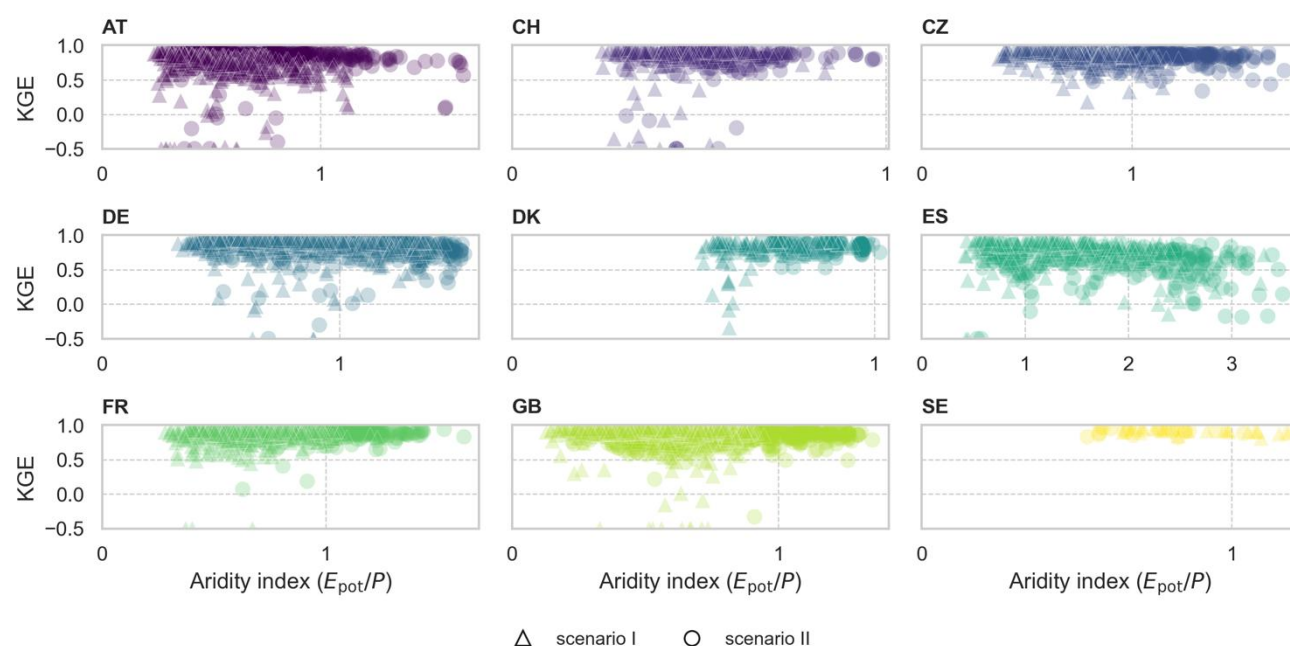
**Figure 7:** Scatterplots showing the model performance (Kling-Gupta efficiency, KGE) for scenario II (y-axes) versus the number of precipitation stations used to derive the E-OBS precipitation data, per country. Each circle represents one catchment. Note that the y-axes were cut at -0.5, in accordance to Fig. 5, and the x-axes differ for the different subplots.



305

### 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, the model performances tended to be lower in catchments with higher aridity indices (drier catchments). This trend was particularly evident for the catchments in Austria, Czechia, Germany, Spain, and Sweden. Although the pattern appeared with both forcing datasets, it was more pronounced for the CAMELS data (scenario I), especially for the catchments in Austria, Czechia, and Spain.



**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. Note that the y-axes were cut at -0.5, in accordance to Fig. 5. Note that the x-axes differ for the different subplots.

## 4 Discussion

### 4.1 Model performances

The HBV model is known for its capabilities for 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 data (Figs. 4 and 5) for the wet



catchments, such as those in Sweden and Denmark. In contrast, in Spain, which has the most arid catchments, the KGE values were the lowest and most variable for both data scenarios, except for Austria under scenario I, where performance was also notably low. These findings are reinforced by the observed relationship between model performance and the aridity index, as 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 hydro-meteorological 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. These heavily modified catchments may not be adequately filtered out in studies focused on natural hydrological simulation, thereby further impairing overall model performance (Klotz et al., 2025; Senent-Aparicio et al., 2024b).

## 4.2 Evaluation of the E-OBS data in comparison to the E-OBS station density

Our findings indicate that the model performance is strongly influenced by the density of stations used to obtain the E-OBS data. As a result, the reliability of model outputs varies considerably across regions—an observation that is consistent with previous research (Bandhauer et al., 2022; 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 and Austria, the model reached the highest KGE values with E-OBS data, underscoring the critical role of data availability and quality in hydrological modelling accuracy.

As mentioned earlier, the results for the catchments in Austria are an interesting case, because the simulations with the E-OBS data (scenario II) were clearly better than those with the CAMELS data (scenario I). Unlike other CAMELS datasets, which rely on national meteorological data products, the CAMELS dataset for Austria (LamaH-CE) uses ERA5-Land as its meteorological forcing. Previous studies have already highlighted several limitations of ERA5-Land in simulating streamflow (Clerc-Schwarzenbach et al., 2024).

The fundamental differences between E-OBS and ERA5-Land stem from the data assimilation. The gridded product of E-OBS is based on a spatial interpolation of ground-based station networks, while ERA5-Land is a reanalysis dataset and derives precipitation from a model without direct incorporation of gauge data (Muñoz-Sabater et al., 2021). In densely monitored regions like Austria, this difference becomes especially relevant. This fundamental discrepancy, alongside the high E-OBS station density in Austria, seems to explain the superior performance with E-OBS data for the Austrian catchments.





### 4.3 Effect of different $E_{\text{pot}}$ data

355 In this work, the differences in model performance across the various sources of forcing data were attributed to the differences in precipitation inputs (scenario III), whereas discrepancies in  $E_{\text{pot}}$  and temperature data hardly affected model performances (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  $E_{\text{pot}}$  than in precipitation data, it was still the precipitation inputs that exerted the greatest influence on model performance.

360 In this study, the differences between the  $E_{\text{pot}}$  data derived from E-OBS and the  $E_{\text{pot}}$  data derived from the CAMELS datasets were much smaller, but still striking (which makes sense, considering the differences in the temperature data that were generally used as input to the  $E_{\text{pot}}$  calculations). This demonstrates once again the large uncertainties that we face when using different approaches to estimate  $E_{\text{pot}}$ . Several studies have already identified this issue as a persistent “blind spot” in hydrological modelling (e.g., Bai et al., 2016; Federer et al., 1996; Hanselmann et al., 2024). The  $E_{\text{pot}}$  calculations for each catchment

365 in EStreams with the Hargreaves equation (do Nascimento et al., 2024) thus also affected the resulting  $E_{\text{pot}}$  data that we used to represent the E-OBS  $E_{\text{pot}}$ . However, the Hargreaves equation was found to be reliable, among other regions especially in Central Europe (Pimentel et al., 2023) and this choice can therefore be supported. Furthermore, the differences in  $E_{\text{pot}}$  data did not affect model performance results strongly, so it can be expected that the use of a different equation would not change the findings of this study.

370 While the different  $E_{\text{pot}}$  data had a very limited effect on model performances in general, the higher  $E_{\text{pot}}$  data derived from E-OBS were beneficial for the model performance in some cases, as they 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  $E_{\text{pot}}$  values from E-OBS effectively helped to improve the water

375 balance by allowing for more evaporation, thereby compensating for the unmodelled groundwater losses. 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).

### 4.4 Limitations

380 As mentioned with the example of the compensation effects due to higher  $E_{\text{pot}}$  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. Apart the study by Clerc-Schwarzenbach et al. (2024), on which the methodology of this study was based, model performances are often used as an aggregated measure of data quality (Beck et al., 2017; Tarek et al., 2020). However, model performance can also be heavily influenced by the chosen model structure, particularly if



385 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).

Another limitation is that we did not exclude any catchments due to human influences. Potentially, this affected model performances. However, this is the case then independent from the forcing data used, so we assumed that our conclusions would not change if these catchments were excluded.

390 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.

In this study, we evaluated the E-OBS data at a resolution of  $0.25^\circ$ , as these are the data on which EStreams is based on (from  
 395 where we obtained the data for this study). Being included in EStreams, it can be expected that this E-OBS product will increasingly be used for LSH studies in Europe in the future. However, the E-OBS product with a higher resolution of  $0.1^\circ$  may lead to different results. Testing this is out of the scope of this study.

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  
 400 (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 3423 European catchments in the EStreams dataset with time series from nine smaller-scale datasets (mostly CAMELS datasets). Moreover, we evaluated how the different meteorological  
 405 forcing data influence the performance of a bucket-type hydrological model.

Our results showed that, for the majority of catchments (89 %), annual precipitation values obtained from the E-OBS dataset were lower than those from the corresponding CAMELS datasets. The opposite was true for the annual sums of  $E_{\text{pot}}$  and the average temperature (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  
 410 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 model performances above a KGE of 0.70 in more than 87 % of the catchments. However, performances were generally slightly lower when using E-OBS data than when using CAMELS data: For approximately 60 % of catchments, the model performance was higher when using CAMELS  
 415 forcing data. Notably, exceptions occurred in cases where the CAMELS data were derived from global rather than national



products (ERA5-Land dataset in Austria). This aligns with our first hypothesis that smaller-scale, nationally curated datasets would lead to higher model performances than the standardized dataset.

Furthermore, our findings confirmed our second hypothesis that model performances using E-OBS forcing data would be 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 regions where convective rainfall is relevant.

Overall, this study presents the first assessment of the meteorological data from E-OBS across a continental domain and offers valuable insights for future large-sample hydrology studies.

## 6 Data availability

Except for CAMELS-CZ and the  $E_{\text{pot}}$  data for CAMELS-SE, all the CAMELS data used are available from their respective repositories (see Table 1). The  $E_{\text{pot}}$  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 1.2 of the EStreams dataset is available at a Zenodo repository (<https://doi.org/10.5281/zenodo.14778580>). 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 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

440 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  $E_{\text{pot}}$  data for CAMELS-SE available and Michal Jeníček and Ondřej  
 445 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).

## 11 References

- 450 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.
- 455 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-Garretón, 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.
- 460 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.
- 465 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.
- 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 hydrolog-  
 470 ical 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.
- 475 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.
- 480 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.
- 485 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.
- 490 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.
- 495 Crameri, F.: Scientific colour maps, Zenodo [software], <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.
- 500 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.
- 505 Dolich, A., Espinoza, E. A., Ebeling, P., Guse, B., Götze, J., Hassler, S., Hauße, C., Kiesel, J., Heidebüchel, 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, 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.
- 510 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.
- 515 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.
- 520 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.
- 525 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.
- 530 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.
- 535 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.
- 540 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.

- 545 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.:  
 550 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  
 555 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>, 2021a.

Klingler, C., Kratzert, F., Schulz, K., and Herrnegger, M.: LamaH-CE: LARge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe – files [Dataset] (1.0), <https://doi.org/10.5281/zenodo.5153305>, 2021b.

- 560 Klotz, D., Miersch, P., Do Nascimento, T. V. M., Fenicia, F., Gauch, M., and Zscheischler, J.: EARLS: A runoff reconstruction dataset for Europe, <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., Florianic, M. G., and Kirchner, J. W.: Catchment hydrological response and transport are  
 565 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.

- 570 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.

- 575 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.



- 580 Liu, J., Koch, J., Stisen, S., Trolldborg, 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., Götze, 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.
- 585 Mangunkiya, 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.
- 590 Mavromatis, T. and Voulanas, D.: Evaluating ERA-Interim, Agri4Cast, and E-OBS gridded products in reproducing spatio-temporal 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.
- 595 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.
- 600 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.
- 605 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.
- 610 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.



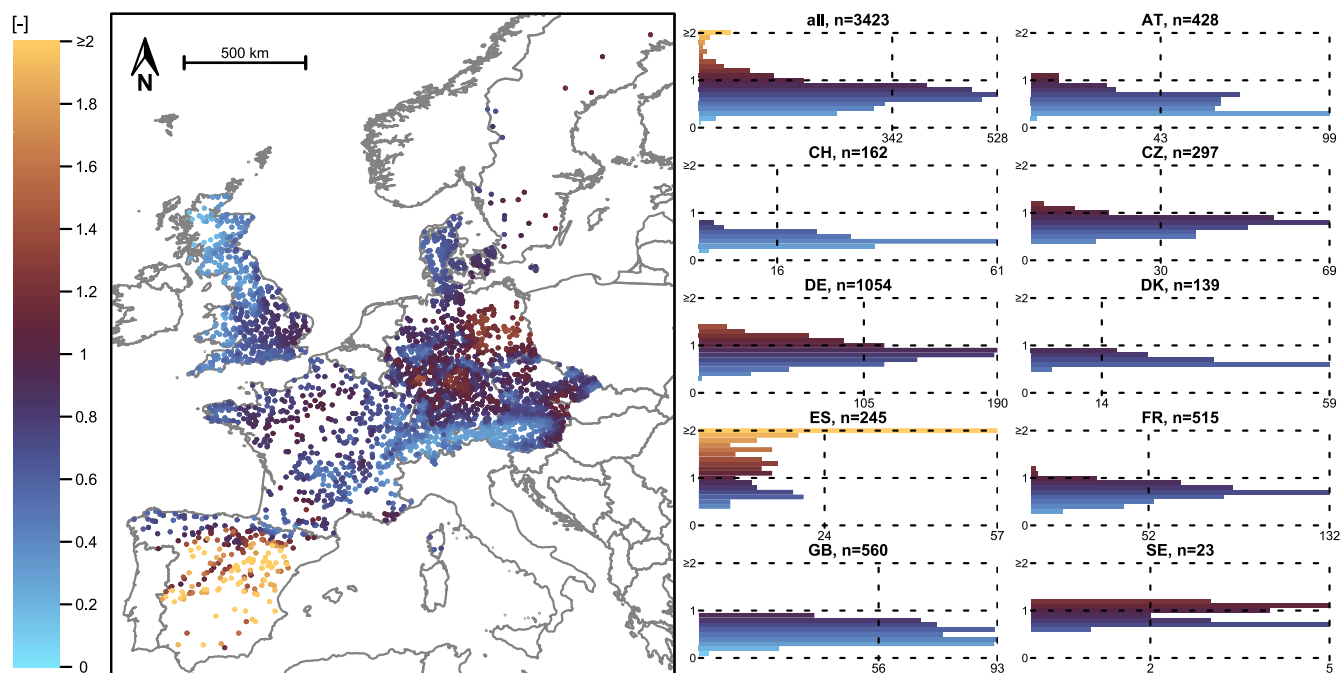
- 615 Peral García, C., Navascués Fernández-Victorio, B., and Ramos Calzado, P.: Serie de precipitación diaria en rejilla con fines climáticos - Nota técnica 24 de AEMET, Agencia Estatal de Meteorología, Madrid, 2017.  
  
Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, 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.
- 620 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.
- 625 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.
- 635 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.  
  
Scharling, M.: Klimagrid Danmark – Nedbør, lufttemperatur og potentiel fordampning 20 × 20 & 40 × 40 km – Metodebeskrivelse, Danish Meteorological Institute, Copenhagen, 1999a.
- 640 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.
- 645 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.
- 650



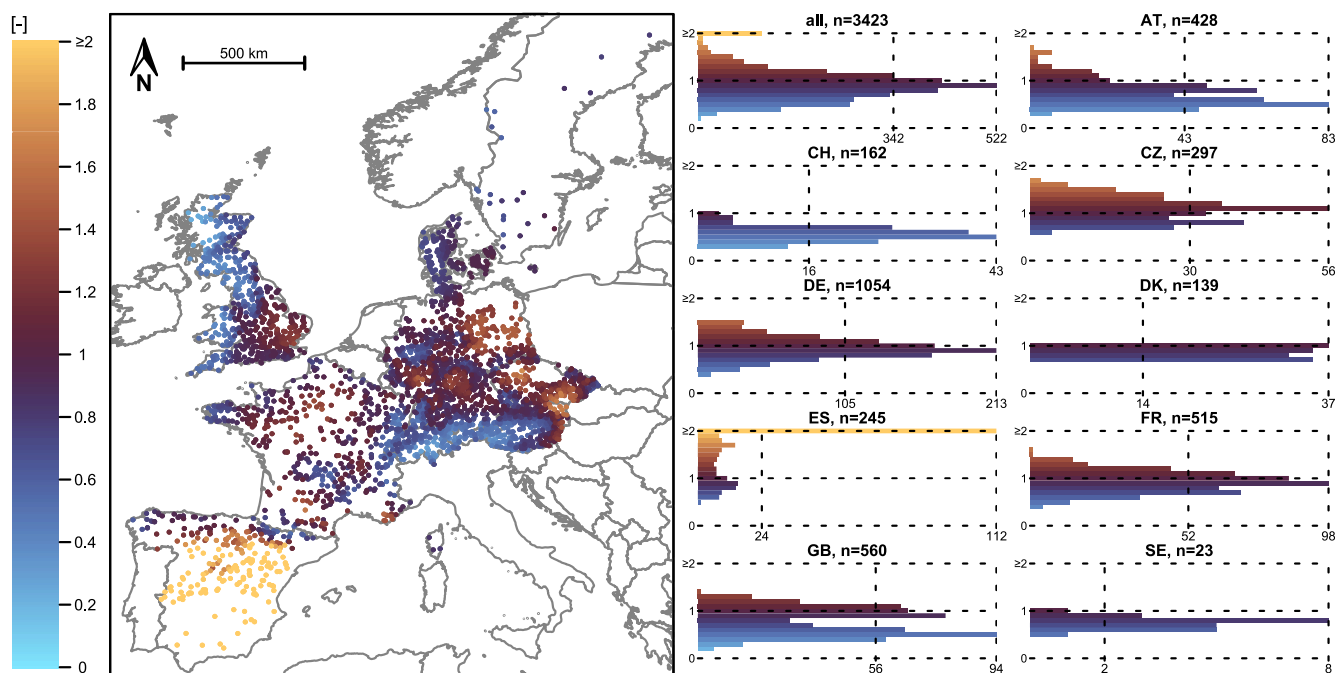
- Senent-Aparicio, J., Castellanos-Osorio, G., Segura-Méndez, F., López-Ballesteros, A., Jimeno-Sáez, P., and Pérez-Sánchez, J.: BULL Database – Spanish Basin attributes for Unravelling Learning in Large-sample hydrology, *Sci. Data*, 11, 737, <https://doi.org/10.1038/s41597-024-03594-5>, 2024b.
- 655 SMHI: Nedladdning av griddad nederbörd- och temperaturdata: <https://www.smhi.se/data/ladda-ner-data/griddade-nederbord-och-temperaturdata-ptbvhv>, last access: 19 June 2025.
- 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.
- 660 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.
- 665 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.
- 670 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.
- Thornthwaite, C. and Mather, J.: Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance, *Lab. Climatol.*, 10 (3), 1957.
- 675 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.
- 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

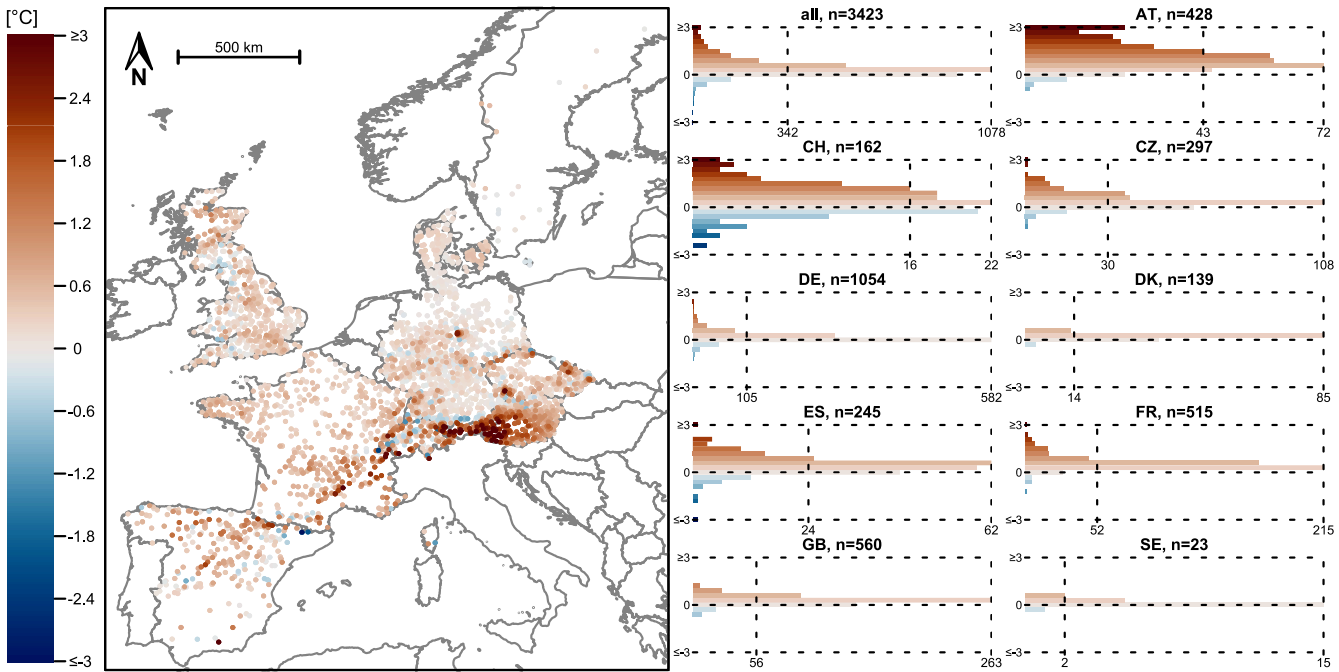


**Figure A1:** Aridity index ( $E_{pot}/P$ ) calculated from the CAMELS data (for a 20-year period: 1995-2015). Note that the colour scale was cut at two.



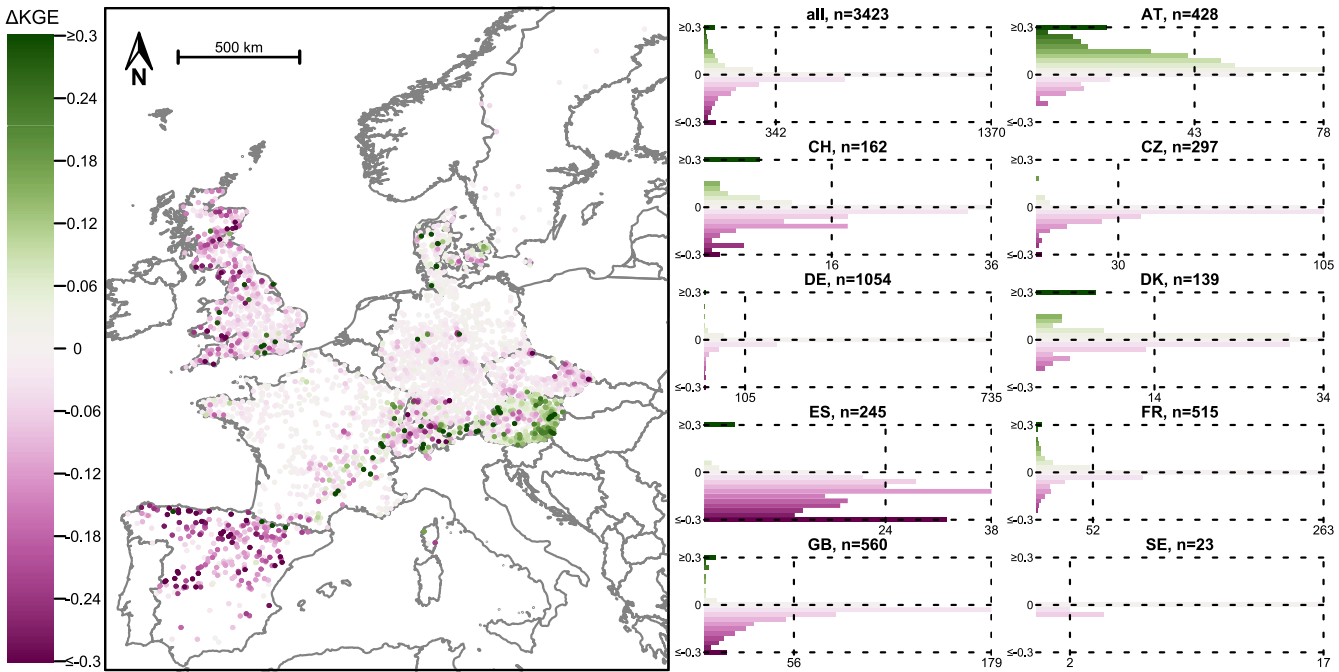
**Figure A2:** Aridity index ( $E_{pot}/P$ ) calculated from the E-OBS data obtained from EStreams (for a 20-year period: 1995-2015). Note that the colour scale was cut at two.





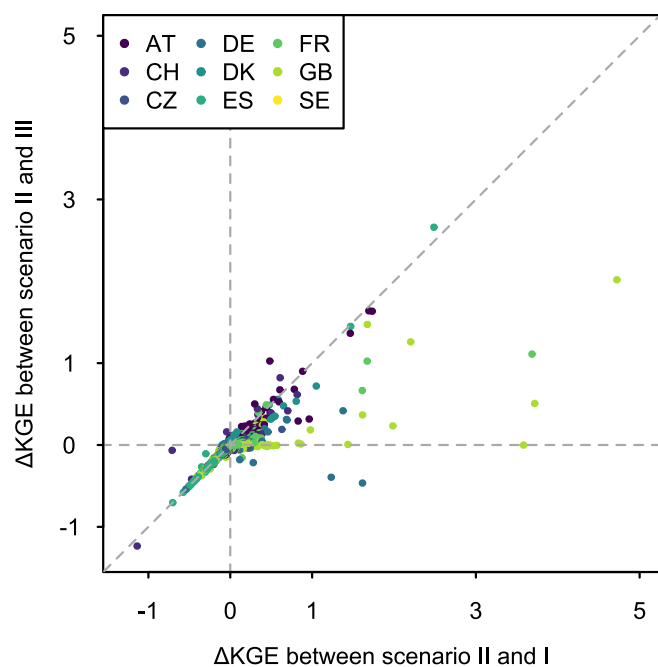
**Figure A3: Difference in mean annual temperature for each catchment when calculated from the E-OBS and the CAMELS datasets for the 20-year period 1995-2015. Positive values and red colours indicate higher temperatures in the E-OBS data obtained from EStreams, negative values and blue colours indicate lower temperatures in the E-OBS data. Note that the colour scale was cut at  $\pm 3$  °C.**

695

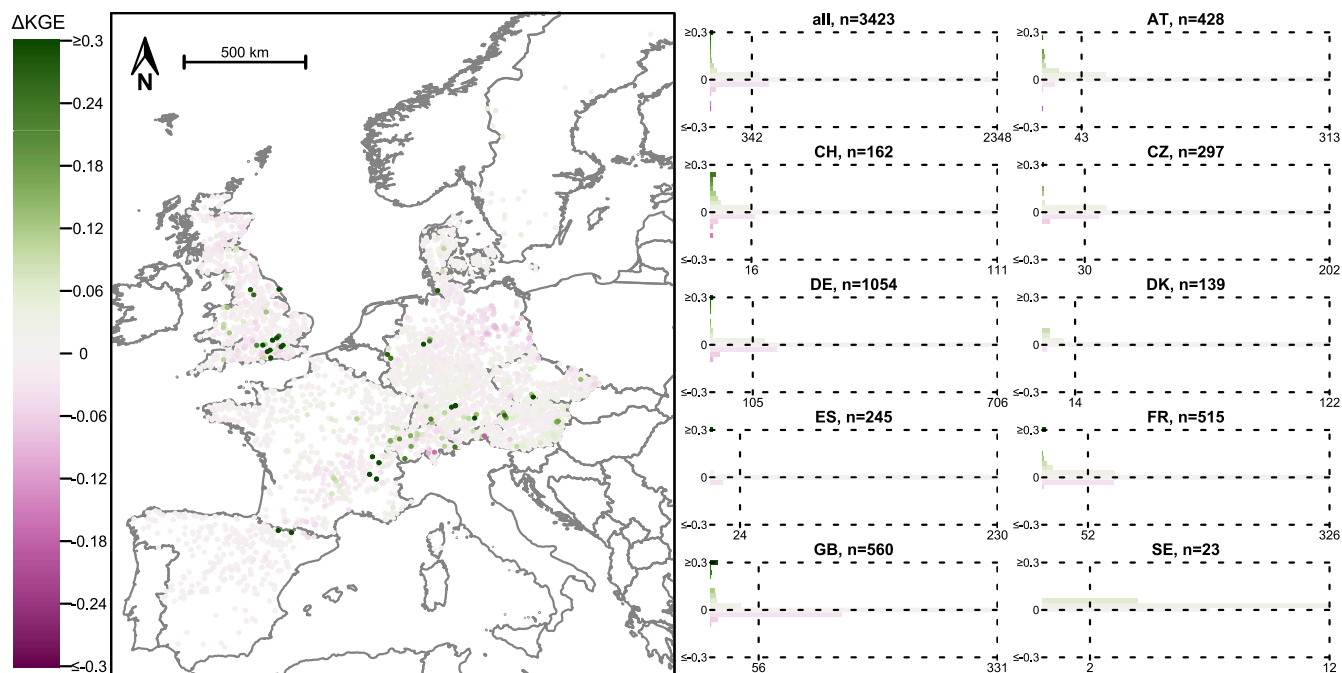


**Figure A4: Difference in model performance when all meteorological input data were obtained from E-OBS (i.e., EStreams, scenario II) and when the precipitation data from E-OBS were replaced with those from CAMELS (scenario III). Positive values and green colours indicate higher model performances with 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  $\pm 0.3$ . The catchments with the largest differences in model performance were plotted last to increase their visibility.**

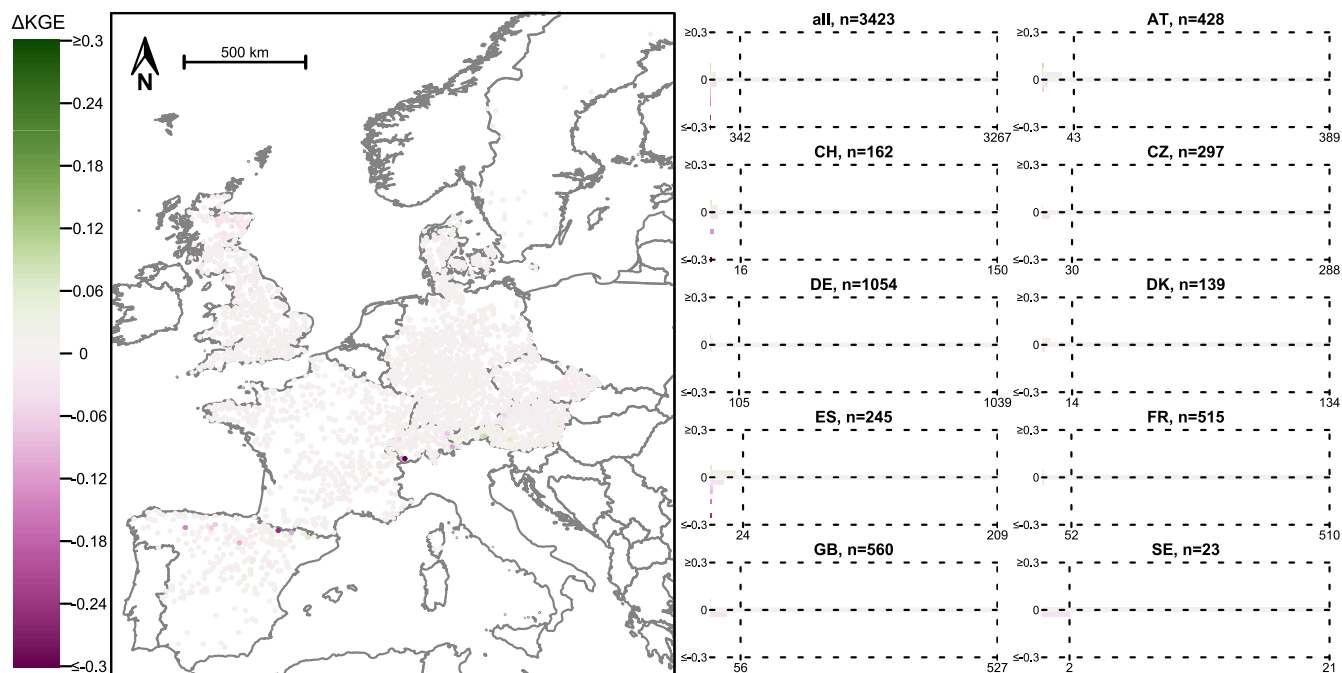
700



**Figure A5:** Differences in model performance between scenario II and I (positive values indicate higher performances with the E-OBS data obtained from EStreams, negative values indicate higher performances with the CAMELS data, see Fig. 6) compared to differences in model performance between scenarios II and III (positive values indicate higher performances with the E-OBS data, negative values indicate higher performances when the precipitation data were replaced with those from CAMELS, see Fig. A4). One catchment (in Great Britain) plotted outside the axis limits (11.1 / 5.5). Pearson's correlation coefficient was 0.86.

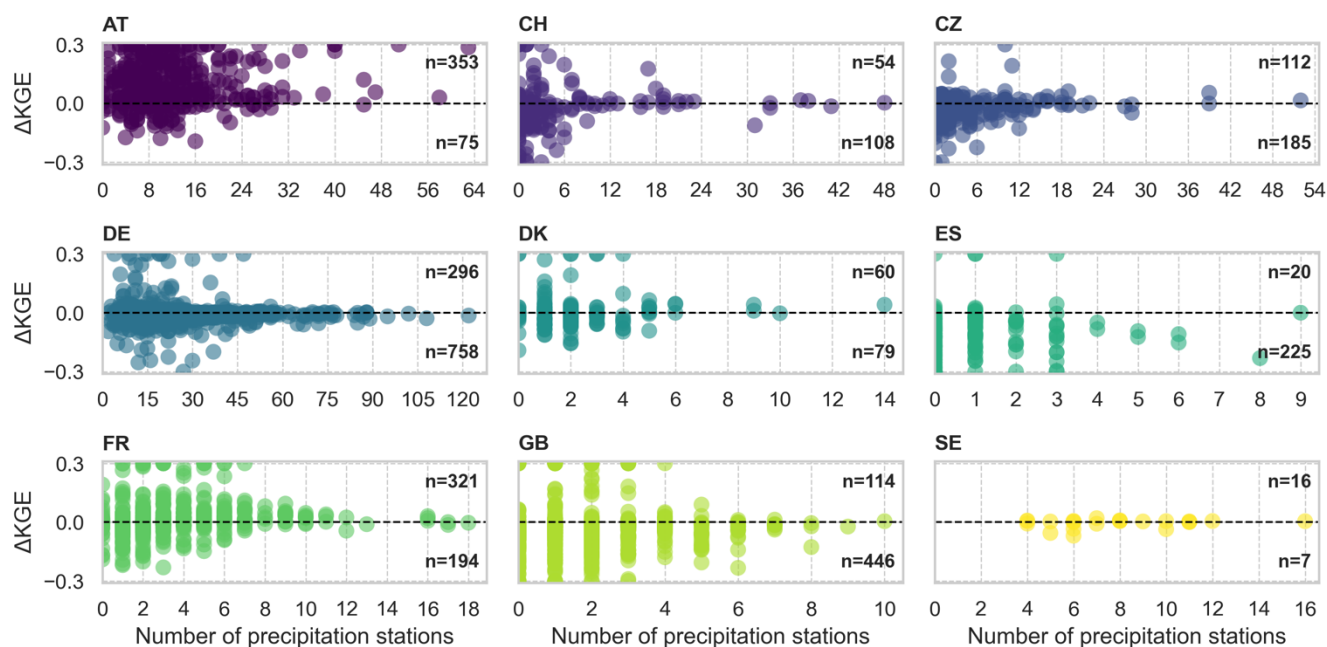


**Figure A6: Difference in model performance when all meteorological input data were obtained from E-OBS (i.e., EStreams, scenario II) and when the  $E_{\text{pot}}$  data from E-OBS were replaced with those from CAMELS (scenario IV). Positive values and green colours indicate higher model performances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



**Figure A7:** Difference in model performance when all meteorological input data were obtained from E-OBS (i.e., EStreams, scenario II) and when the temperature data from E-OBS were replaced with those from CAMELS (scenario V). Positive values and green colours indicate higher model performances with the  $T$  data from E-OBS, negative values and pink colours indicate higher model performances with the  $T$  data from CAMELS. Note that the colour scale was cut at a difference in KGE of  $\pm 0.3$ . The catchments with the largest differences in model performance were plotted last to increase their visibility.

720



**Figure A8:** Scatterplots showing the difference in model performance (Kling-Gupta efficiency, KGE) between scenario II (E-OBS data obtained from EStreams) and scenario I (CAMELS data) (y-axis) versus the number of precipitation stations used to derive the E-OBS precipitation data per country. Each circle represents one catchment. Positive values indicate higher performances when the E-OBS data were used, negative values indicate higher performances when the CAMELS data were used. Note that the y-axes were cut at  $\pm 0.3$ , in accordance to Fig. 6. Note that the x-axes differ for the different subplots.