CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking

**Olivier Delaigue[1], Guilherme Mendoza Guimarães[1], Pierre Brigode[1,2,3], Benoît Génot[1,a],
Charles Perrin[1], Jean-Michel Soubeyroux[4], Bruno Janet[5], Nans Addor[6,7], and Vazken Andréassian[1]**

[1]Université Paris-Saclay, INRAE, HYCAR, Antony, France
[2]Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Sophia-Antipolis, France
[3]Université de Rennes, CNRS, Géosciences Rennes, Rennes, France
[4]Météo-France, DCSC, Toulouse, France
[5]Service central d'hydrométéorologie et d'appui á la prévision des inondations (MTES/DGPR/SCHAPI),
Toulouse, France
[6]Fathom, Bristol, UK
[7]Department of Geography, University of Exeter, Exeter, UK
[a]now at: U.R.B.S., Saint-Étienne, France

Correspondence: Olivier Delaigue (olivier.delaigue@inrae.fr)

Abstract. Over the last decade, large-sample approaches, i.e., based on large catchment sets, have become increasingly popular in hydrological studies. Efforts were made to assemble and disseminate national catchment datasets. This article aims to make a contribution to the construction of a large international database of catchments by proposing the CAMELS-FR dataset, a contribution to the CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) initiative. The first version presented here gathers hydroclimatic data and physical attributes for a set of 654 catchments in France. These catchments cover a wide spectrum of hydroclimatic conditions (from oceanic to continental, mountainous, or Mediterranean conditions) and are considered to have limited human influence. Data include time series of daily streamflow (with at least 30 years over the 1970–2021 period, also aggregated to monthly and yearly time steps) and of 11 catchment-scale daily climate variables (including precipitation, potential evaporation, and air temperature), as well as a total of 255 catchment attributes organized into 10 classes (e.g., geology, soil, land cover). River flow time series were quality-checked. Along with the database itself, two graphical tools are proposed, namely dynamic graphs to visualize time series and graphical fact sheets to summarize the main catchment characteristics. Care was taken to provide as many metadata as possible to help users interpret their results based on this dataset. We intend to update the database regularly to include new available data and account for end users' feedback. CAMELS-FR (Delaigue et al., 2024c) is available at https://doi.org/10.57745/WH7FJR.

1   Introduction

In the early days of hydrological modeling, access to data, followed by computing resources, was a major limiting factor of progress, and many of the scientific articles of that time reported results obtained for a single catchment (and sometimes using only a few flood events). The precursors of modern hydrological models were conscious that this was a major shortcoming. Linsley (1982, p. 15) for example recommended that "because almost any model with sufficient free parameters can yield good results when applied to a short sample from a single catchment, effective testing requires that models be tried on many catchments of widely differing characteristics, and that each trial cover a period of many years." Roughly 20 years later, under the tireless leadership of John Schaake, the so-called *MOPEX* dataset started to be used worldwide (Schaake et al., 2006), breathing new life into the hard drives of data-thirsty modelers. The paper published by Gupta et al. (2014) 10 years later presented an already impressive list of 94 "large-sample" studies focusing on rainfall-runoff modeling, which had used a dataset of more than 30 catchments. Today, the number of such studies would probably be multiplied by 10, and it has become almost impossible to keep track of all of them. Obviously, computing time difficulties and data sharing possibilities have completely changed.

In France, there has been a long tradition of working with large datasets, following the recommendations of two French leading hydrologists of the 20th century: Maurice Pardé and Marcel Roche. Later, however, large datasets were sometimes mocked as "hydro-bulimia" (Andréassian et al., 2009), and during a certain time, the academic world favored the production of individual basin monographs rather than modeling studies involving large datasets. Over the last decade, progress has been made towards automation of the production (and the updating) of large catchment datasets. In collaboration with the main data producers of hydrometric and meteorological data, we worked to produce the reference hydrological dataset that we present in this paper. For the selection of this dataset, we used what we considered to be "high" quality standards, and because we do acknowledge that there is some subjectivity in this, let us paraphrase Ghislain de Marsily's comment on model validation and state that we strove to do "our level best" (de Marsily et al., 1992) using a variety of automatic data verification procedures, which were complemented by manual checks (including a time-consuming – but clearly necessary – visual data inspection).

With this work, we wish to contribute to the general effort to provide large hydroclimatic datasets as was done in the United States of America (Newman et al., 2015; Addor et al., 2017), Canada (Arsenault et al., 2016), Chile (Alvarez-Garreton et al., 2018), Great Britain (Coxon et al., 2020), Brazil (Chagas et al., 2020; Almagro et al., 2021), Australia (Fowler et al., 2021), central Europe (Klingler et al., 2021), Africa (Tramblay et al., 2021), Denmark (Koch, 2021; Liu et al., 2024), Switzerland (Höge et al., 2023), Saxony (Hauffe et al., 2023), and Germany (Loritz et al., 2024), with many others that will follow. As a national dataset, CAMELS-FR should also be seen as complementary to other datasets built at larger scales that include France, e.g., the EStreams dataset at the European scale (do Nascimento et al., 2024a, b). CAMELS-FR differs from such datasets in the criteria used in the catchment selection process, the data analysis methods, and the choices in data sources that may be available at national but not larger scales.

Before moving to the presentation of the dataset, let us warn the reader that for reasons of scientific ethics, we refused to exclude hydrological "outliers" (i.e., catchments that exhibit unusual hydrological behavior, such as karstic catchments, groundwater-dominated catchments fed by the chalk aquifer of the Parisian Basin, and those with very high baseflow indices) from the CAMELS-FR dataset. We did so because we believe that the outliers are part of the "natural hydrological diversity" (Andréassian et al., 2010) and also because, from a scientific point of view, even good-looking catchments can turn into a "modeler's nightmare" (Refsgaard and Hansen, 2010). This is why we kept, for example, the karstic catchments which have been excluded from other datasets. Note that other catchment sets at the national scale already exist in France for more specific purposes, for example the Reference Low Flow Network to study the long-term evolution of low flows (see Giuntoli et al., 2013; Hodgkins et al., 2024) and elsewhere (see a review of large-sample studies in Gupta et al., 2014).

2   French physical and hydroclimatic context

2.1   Physical characteristics

Metropolitan France (see Fig. 1) covers an area of $550\,000\,km^2$, including mainland France (spanning $1000\,km$ from north to south and $1000\,km$ from east to west) and Corsica, the fourth-largest Mediterranean island ($8700\,km^2$). It is mainly bounded by coastlines (with the North Sea, the English Channel, the Atlantic Ocean, and the Mediterranean Sea) and shares terrestrial borders with eight neighboring countries (Belgium, Luxembourg, Germany, Switzerland, Italy, Monaco, Spain, Andorra). Metropolitan France offers a wide variety of natural landscapes inherited from several geological phases, giving rise to ancient (e.g., Armorican Massif, Massif des Vosges, and Massif Central) and younger mountain ranges such as the Jura, Alps, and Pyrenees (see Fig. 2a). The average altitude is $344\,m$, and the elevation reaches a maximum of $4806\,m$ at Mont Blanc (in the Alps). These mountain ranges form the boundaries of several sedimentary basins, including the Aquitaine Basin in the southwest and the Paris Basin to the north.

The French geological diversity is illustrated with the wide range of colors found on the lithological map of France (see Fig. 2c). We give here a short description of the (complex)

Figure 1. Map of the main geographical features of France: mountain ranges (text with white outline), regions (text without outline), and main basins (pink: Seine; yellow: Loire; orange: Garonne; purple: Rhône) (river network: Lehner and Grill, 2013; digital terrain model, DTM: Lehner et al., 2008; shoreline and political boundaries: Wessel and Smith, 1996; NOAA, 2017).

geology of France, focusing on its implications for understanding the hydrogeology of French catchments. A more detailed account is provided by Pelletier (2021).

- A quarter of France is covered by basement rocks that are metamorphic or igneous in origin. Basement formations host small local aquifers, but no large-scale aquifers. The two main areas are the Armorican Massif in the northwest, almost exclusively composed of ancient formations, and the Massif Central in the center of the country (a more complex area where Hercynian basement, sedimentary formations and recent volcanic formations coexist). Basement rocks also outcrop in smaller massifs of lesser extent (e.g., the Massif du Morvan in the upper Yonne basin, the Massif des Albères at the border of France and Spain in Catalonia, the Massif ardennais around the Meuse River at the border with Belgium, the Massif de la Serre in the Jura Mountains near the Swiss border, and the Massif des Maures on the Mediterranean coast near Saint-Tropez). Note that the Massif des Vosges is divided between basement formations to the south and sedimentary formations to the north.

- Intensely folded areas are hydrogeologically very complex because in these regions, aquifers, if they exist, are of limited spatial extent and therefore difficult to identify for mapping on a regional or national scale. These

areas correspond to the most recent massifs: the central and inner part of the Alpine arc, the Pyrenees, and the Languedoc massifs.

- The rest of the territory is occupied by two main types of formations: sedimentary formations and alluvial plains. The sedimentary formations form two major basins: the Paris Basin, which covers a third of metropolitan France, and the Aquitaine Basin in the southwest. Formed by sedimentary deposits left by marine intrusions, they appear as a succession of geological layers with diverse aquifer properties, whose outcrop areas form rings, known as the "pile d'assiettes" (plate stack) in the Paris Basin.

- Alluvial plains and the accompanying aquifers are formed by the scouring and degradation by rivers of various materials, which are then deposited along the river bed. In mainland France, they are generally of limited geographical extent, unlike in other parts of the world. Two alluvial aquifers, however, are of greater importance: they are located in the upper Rhine plain in Alsace and in Bresse (around the Saône River, north of Lyon).

- The variability in surface aquifer properties is clearly apparent when one looks at the density of the observed river network in Fig. 2c, where for example the large areas underlaid by chalk formations in the Paris Basin stand out with their much lower drainage density, while the Champagne Humide region, characterized by clayey and loamy soils, has a much denser river network. The Beauce region, underlaid by the Beauce limestone aquifer complex, is clearly visible as a white spot southwest of Paris, at the border of the Seine and the Loire basins.

In 2015, farmland covered around 51 % of metropolitan France, compared with around 40 % for soils not directly subject to anthropogenic pressures (woodlands, wetlands, and water surfaces) and around 9 % for artificialized soils (see Fig. 2d).

2.2 Hydroclimatic characteristics

According to the Köppen–Geiger climate classification (see, e.g., Peel et al., 2007, and Strohmenger et al., 2024, for a recent analysis of France), more than 90 % of the French territory belongs to the *Cfb* class (i.e., temperate without dry season and with warm summers), with Corsica and the Mediterranean shore belonging to the *Csa* and *Csb* classes (temperate with dry and, respectively, hot or warm summers). The mountain ranges belong to class *Dfb* (continental without dry season and warm summers), *Dfc* (continental without dry season and cold summers), and the high-elevation summits belong to the *ET* class (polar climate). Average annual
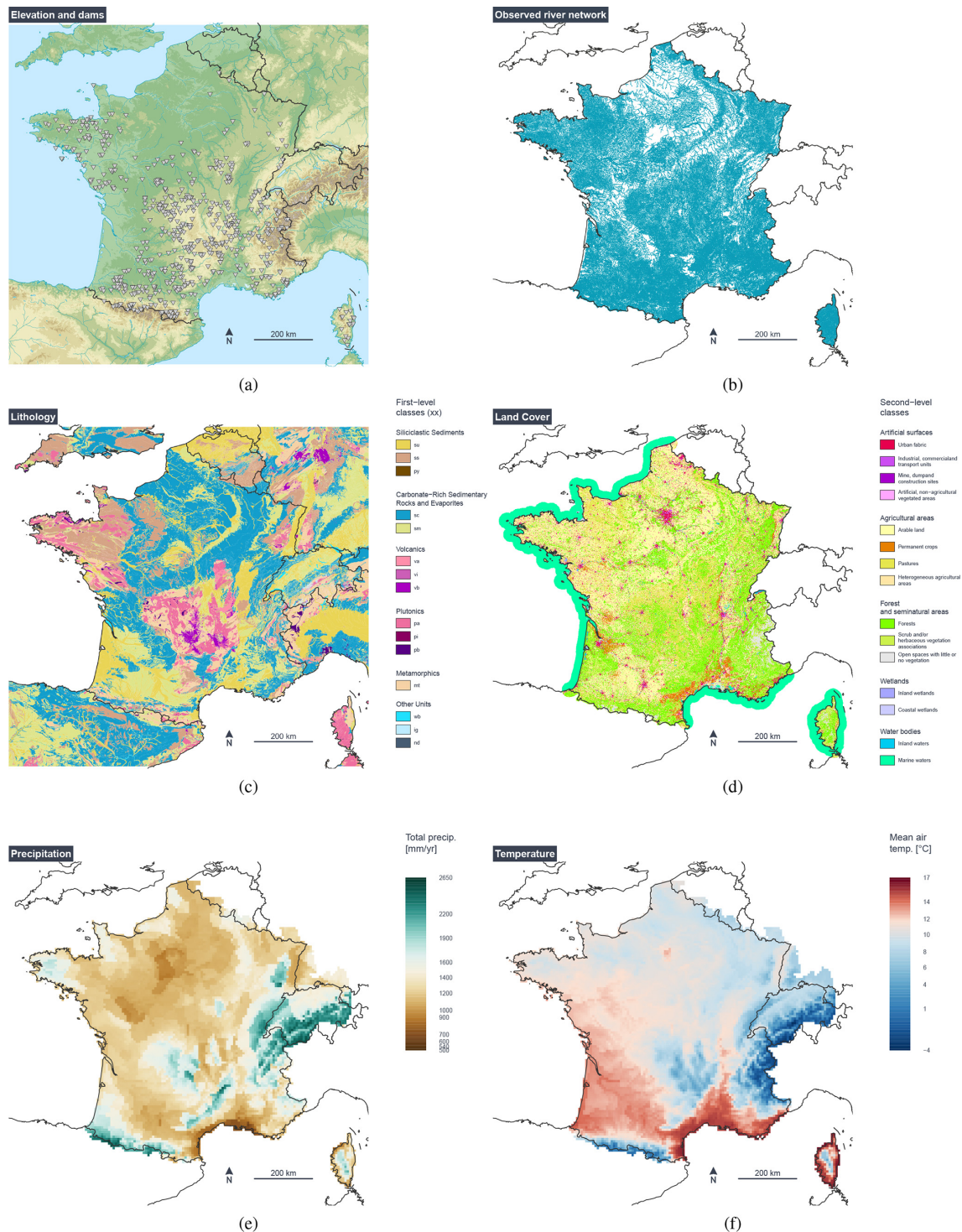
Figure 2. Maps describing metropolitan France territory. **(a)** Elevation, main rivers, and main dam locations (triangles) (river network: Lehner and Grill, 2013; DTM: Lehner et al., 2008; shoreline and political boundaries: Wessel and Smith, 1996; NOAA, 2017); **(b)** observed river network (CARTHAGE: French water agencies, 2017b); **(c)** lithology (GLiM v2.0: Hartmann and Moosdorf, 2012a, b; *su*: unconsolidated sediment; *ss*: siliciclastic sedimentary rocks; *py*: pyroclastics; *sc*: carbonate sedimentary rocks; *sm*: mixed sedimentary rocks; *va*: acid volcanics; *vi*: intermediate volcanics; *vb*: basic volcanics; *pa*: acid plutonics; *pi*: intermediate plutonics; *pb*: basic plutonics; *mt*: metamorphics; *wb*: water bodies; *ig*: ice and glaciers; *nd*: non-defined); **(d)** land cover (CLC: French Ministry of the Environment, 1990); **(e)** mean annual total precipitation over the 1991–2020 period (SAFRAN: Quintana-Segui et al., 2008; Vidal et al., 2010); **(f)** mean air temperature over the 1991–2020 period (SAFRAN).

precipitation ranges from $500 \, \mathrm{mm \, yr^{-1}}$ for the driest regions (e.g., Mediterranean coasts) to more than $2000 \, \mathrm{mm \, yr^{-1}}$ in mountainous regions (see Fig. 2e and f). While precipitation usually varies in a progressive manner over much of the country, a few mountain ranges are characterized by strong differences: one can mention for example the Cévennes range in the south, where average rainfall reaches $2000 \, \mathrm{mm \, yr^{-1}}$, located not far from the Crau lowlands, where average rainfall is less than $500 \, \mathrm{mm \, yr^{-1}}$. A similar situation exists near the Rhine valley in Alsace, where average rainfall in the Massif des Vosges reaches $2000 \, \mathrm{mm \, yr^{-1}}$, while it is less than $500 \, \mathrm{mm \, yr^{-1}}$ in the Colmar plain. As far as temperature is concerned, it is probably enough to mention the mild winter temperatures on the ocean coast and the Mediterranean sea, the lowest temperatures being reached on the high mountain ranges, where a few glaciers still exist despite the global warming trend.

Metropolitan France's hydrographic network (Fig. 2b) is mainly organized around four major rivers (the Loire, Seine, Garonne, and Rhône), whose catchments cover more than 60 % of mainland France. The Rhône and the Garonne catchments are transboundary, with their sources in Switzerland and Spain, respectively. Other transboundary rivers are found in France, such as the Meuse (crossing France, Belgium, and the Netherlands), the Rhine (crossing Switzerland, Liechtenstein, Austria, Germany, France, and the Netherlands), and the Roya (crossing France and Italy), among others (e.g., the Saar, the Oise). The French rivers are characterized by different hydrological regimes, with rain-dominated catchments, both rain- and snow-dominated catchments, snow-dominated catchments, Mediterranean catchments, and groundwater-dominated catchments. The most upstream parts of a few mountainous catchments are still influenced by glaciers. In metropolitan France, glaciers are located in the Alps and in the Pyrenees. Most of them cover rather small areas, especially when related to the area of catchments. The two main glaciers are the Argentière glacier and the Mer de Glace, in the Mont Blanc massif. Vincent et al. (2019) showed that, due to climate change, these two glaciers lost 34 and 45 m of water-equivalent depth, respectively, since the beginning of 20th century, representing 25 % and 32 % of their thickness, respectively. Some other glaciers are located in the Écrins and Vanoise massifs.

Most of France's catchments are impacted by human activities through the presence of large dams (for water supply, hydropower, or flood and low-flow management), river abstraction, and groundwater pumping for agricultural, industrial, and drinking water use; therefore we aimed to exclude the most influenced catchments. The locations of the main French reservoirs are shown in Fig. 2a. Note that a large number of small artificial water bodies exist in France for various uses (recreation, irrigation, boating, fishing, etc.), but often with more limited information available.

2.3   Main data producers

The French hydrosystems have been monitored for several decades through various observational networks, maintained by different organizations. These mainly include national agencies such as Météo-France for climatic data, the French geological survey (Bureau de recherches géologiques et minières, BRGM) for geological and hydrogeological information, the National Institute of Geographic and Forest Information (Institut national de l'information géographique et forestière, IGN) for geographic and forest information, and different state services and independent producers (e.g., EDF, Électricité de France; CNR, Compagnie nationale du Rhône; universities) for hydrological data. These streamflow data are made available by the central service for hydrometeorology and inundation forecasting (Service central d'hydrométéorologie et d'appui á la prévision des inondations, SCHAPI).

3   Catchment boundaries

The first step to build the CAMELS-FR dataset was to delineate the contours of the catchment boundaries, which were then used to calculate hydroclimatic time series at this scale.

3.1   Making a flow direction grid to delineate catchment boundaries at the national scale

Hydrological analysis and modeling require climatic forcing averaged at the catchment scale to be associated with the streamflow measurements recorded at the outlet of a catchment. It was therefore necessary to identify the geographic extent of all French catchments.

To do so, we used a flow direction grid, i.e., a matrix summarizing the topological relationships for metropolitan France. This grid is derived from a topographic analysis of two digital terrain models (DTMs). For continental metropolitan France, we used the DTM from the Shuttle Radar Topography Mission (SRTM) project with a resolution of $100 \, \mathrm{m} \times 100 \, \mathrm{m}$ (Rabus et al., 2003; Farr et al., 2007), and for Corsica, we used the BD ALTI v1.0 from the IGN (2001) with a resolution of $25 \, \mathrm{m} \times 25 \, \mathrm{m}$.

Because no DTM is error-free, we used an "observed" river network to force the geometry of the theoretical river network to be closer to reality. To constrain the flow directions via the stream-burning method, we used the CARTHAGE river network (French water agencies, 2017b) with removed channels. This method consists of initially burning the vector network in the DTM by artificially lowering the altitude of the pixels belonging to the network. For more details about the method, see Bourgin et al. (2010).

https://doi.org/10.5194/essd-17-1461-2025

Earth Syst. Sci. Data, 17, 1461–1479, 2025

3.2 Geographic repositioning of hydrometric stations on the flow direction grid

When the metadata linked to the gauging stations are well filled in, the geographical coordinates (latitude, longitude, and altitude) and the area of the upstream catchment of the station are available. However, this information is not always consistent with the flow direction grid that we made. Sometimes the resolution of a DTM is not sufficient, and sometimes the gauging station's coordinates are erroneous or not sufficiently precise. Geographically repositioning the hydrometric stations on the calculated flow direction grid is therefore often necessary. Each hydrometric station must be linked with the "river" pixel of the flow direction grid. Only then can the contour and area of the catchment be delineated.

For the gauging stations that are not positioned in a "river" pixel, we searched the theoretical position of the gauging station in the 24 surrounding pixels (i.e., radius of 200 m for continental, metropolitan France) by iteratively calculating the catchment area located upstream of each surrounding pixel. The objective is then to minimize the difference between the area indicated by the data producer (if available) and the area calculated on the flow direction grid. This method allows us to automatically place a large part of the hydrometric stations on the flow direction grid.

However, errors still exist at some stations. Therefore, all outlets were checked manually using a graphical user interface (GUI) we developed for this specific goal (Génot and Delaigue, 2018). This GUI (see screenshot on Fig. 3; not made available with the dataset) provides means to analyze the geographical location of the hydrometric station by visualizing maps (e.g., maps or aerial photos) and the comparison between the observed and the theoretical river networks, but also by facilitating the comparison of catchment areas.

4 Selection of the catchment set

The second step to build the CAMELS-FR dataset was to select the catchments based on the following criteria: (i) hydrometric time series availability (Sect. 4.1), (ii) limited influence of artificial reservoirs (Sect. 4.2), (iii) consistency in catchment areas (Sect. 4.3), (iv) sufficient streamflow quality (Sect. 4.4). Starting from 4667 catchments, the successive application of these four selection criteria resulted in sub-sets of 1313, 1055, 1031, and 654 catchments, respectively.

4.1 Hydrometric time series availability

We excluded catchments with less than 30 years of complete data over the 1970–2021 period. This record length was arbitrarily chosen because it allows robust statistical analyses. A year was considered complete if less than 20 % of the data were missing.

4.2 Influence of artificial reservoirs

Our objective was to only select catchments displaying a level of human disturbance as low as possible. Here, we only consider influences due to artificial reservoirs. Therefore, the MADAM dataset v1.0 (Delaigue et al., 2024b) was used to estimate these influences. To build this dataset, two sources of information were used, namely the work of Payan (2007) and Payan et al. (2008) and the GEOBS datasets (OFB and partners, 2023), in order to provide locations and volumes of dams in metropolitan France (mainland and Corsica). Each time, we calculated a theoretical level of influence by estimating the equivalent water depth of the total water storage capacity of all dams in the catchment (sum of capacities divided by catchment area).

We cross-checked the two sources of information. The positions of the dams were manually checked. When the storage capacity was inconsistent among the datasets, we validated the dam volume using the website of the French Committee for Dams and Reservoirs (Comité français des barrages et réservoirs, CFBR, 2023), by collecting information from various technical documents of the operators freely available on the internet, or by directly contacting the operators.

Only the catchments with an equivalent water depth of storage capacity lower than 10 mm were selected, thus removing catchments considered highly influenced from our selection. This threshold comes from experience from past studies on this issue (e.g., Payan et al., 2008) but remains arbitrary. The actual level of influence may depend on the local context (location of the dam(s) within the catchment, management objectives of the dam(s), etc.). But we preferred to stick to a single threshold value to have a homogeneous selection criterion.

Note that other types of influence may significantly modify natural streamflow in the selected catchments, typically water withdrawals for various uses or inter-basin water transfers. However, since this information could not be accessed and processed easily, it was not considered here as selection criteria. Thus, the CAMELS-FR catchment set should not be considered to contain strictly non-influenced catchments. For this reason, we also include in the catchment attributes descriptors giving the qualitative estimation of the local and general impact of human influences as provided by the data producers.

4.3 Consistency in catchment areas

The repositioning of the catchment outlets (reported by the data producers, alongside the catchment area) on the flow direction raster was manually checked. We wished to have similar catchment areas between the automatically computed procedure using the flow direction raster and the information provided by the data producers. The catchments whose areas differ by more than 10 % were discarded from the selection. Further investigation would be needed to clarify the reasons
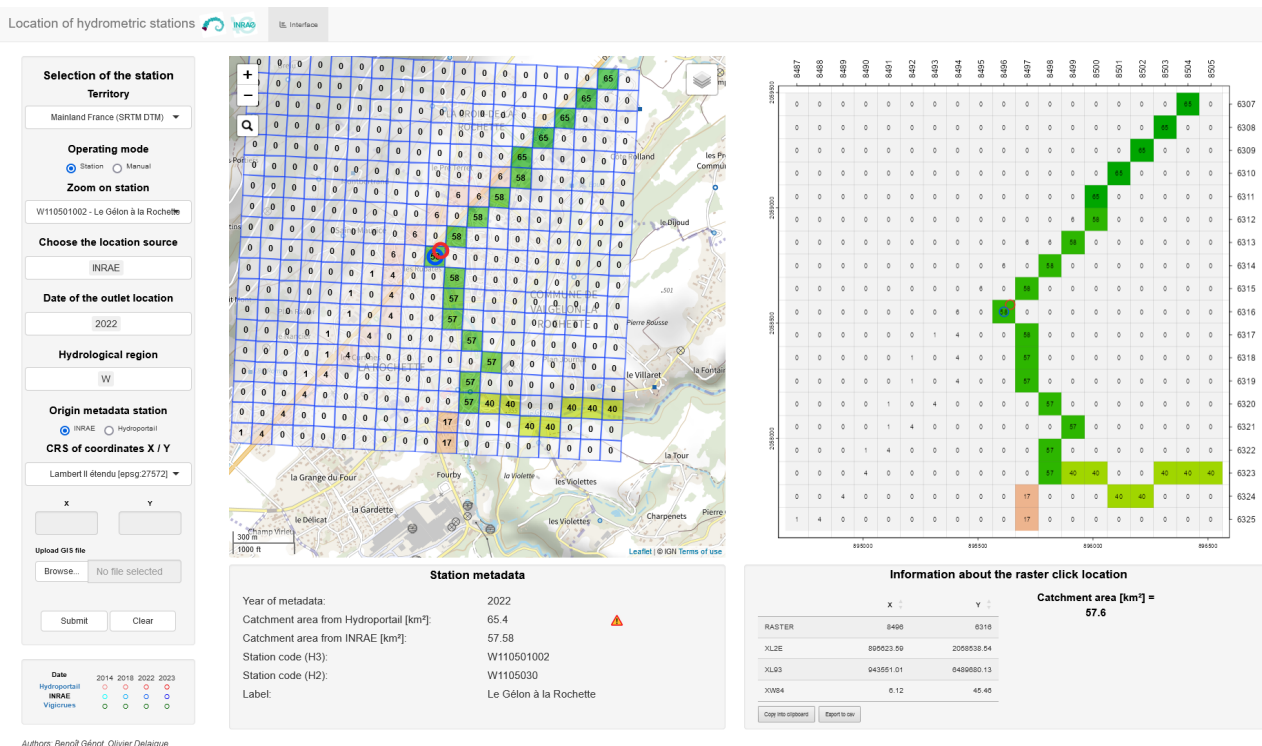
Figure 3. Screenshot of the graphical user interface helping to relocate hydrometric stations in the theoretical river network (Génot and Delaigue, 2018). The left panel is for option selection, where we can select the territory (mainland France or Corsica), a specific station by providing its code or by directly inputting the coordinates, and the metadata year. The middle panel shows the zoomed location of the station on a map layer ("Plan IGN" v2 map layer; IGN, 2020), a superposition of the raster layer with a contributing area grid to aid identification of the correct location of the hydrometric stations. The producer's gauging station is in the red circle, and our own location snapped on the theoretical river network is in the blue circle. The metadata for the searched gauging station are displayed below this map. On the right panel, the contributing area grid and the hydrometric station locations are displayed again, allowing the map to be clicked and the raster information, such as pixel location and its corresponding catchment area, to be extracted.

for these differences, which may lead to additional catchments in the CAMELS-FR dataset in the coming years. For example, there may exist two types of catchment areas (topographic and hydrogeologic) for karstic catchments in the metadata provided by the producers.

4.4 Streamflow quality inspection

Finally, a visual analysis of streamflow time series was performed in order to identify obvious errors in the streamflow series, such as flow interpolation, sudden drops, and noises not referenced as such. The time series were evaluated by four observers. If at least two observers considered a time series to be incorrect, it was removed from the dataset. If only one observer deemed a time series incorrect, it underwent group re-evaluation to get a consensus on the data quality. Catchments with such errors have been discarded from the dataset, pending data correction from the data producers. This qualitative analysis of streamflow series may lead to the addition (or removal) of catchments in the CAMELS-

FR dataset in the coming years. The difficulty of detecting non-natural records in streamflow time series has been recently illustrated and discussed by Strohmenger et al. (2023) on a large set of French catchments.

5 Hydroclimatic time series

5.1 Climatic time series

The SAFRAN atmospheric reanalysis (Quintana-Segui et al., 2008; Vidal et al., 2010) was used as a source of daily climatic time series, aggregated at the catchment scale. SAFRAN is a mesoscale analysis system of near-surface atmospheric variables from the SAFRAN–ISBA–MODCOU hydrometeorological model chain (SIM2; Le Moigne et al., 2020). The SAFRAN data are provided annually by Météo-France at a daily time step from 1958 to the present, and there are no missing data. The reanalysis methodology is regularly updated, and data are updated retroactively (since 1958). SAFRAN uses surface observations combined with data
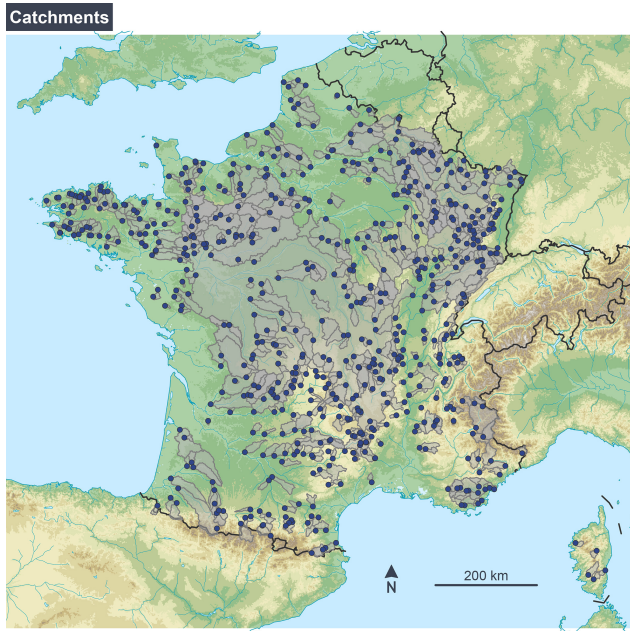
Figure 4. Location map of the 654 catchments and their outlets of the CAMELS-FR dataset v1 (river network: Lehner and Grill, 2013; DTM: Lehner et al., 2008; shoreline and political boundaries: Wessel and Smith, 1996; NOAA, 2017).

from meteorological models, in particular the ERA reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF). Climate variables are analyzed by 300 m altitudinal steps. They are then interpolated on a regular grid with a resolution of 8 km × 8 km (note that catchment rainfall is interpolated from ground measurements). The climatic data cover the whole metropolitan territory (mainland France and Corsica) as well as areas at the borders to correspond to the catchments of rivers flowing into France, with a total of 9892 pixels. An evaluation of the SAFRAN reanalysis is provided in Vidal et al. (2010).

The SAFRAN variables used in the CAMELS-FR dataset are solid precipitation, liquid precipitation, air temperatures (minimum and maximum), wind speed, specific air humidity, and atmospheric and visible radiation. In addition, these climatic data feed a soil–vegetation–atmosphere (SVAT) model called ISBA (Le Moigne et al., 2020) to produce surface variables, such as snow water equivalent and a soil wetness index. Table 1 lists these variables and if they represent daily means or daily accumulations. Note that the aggregation time window may differ between the variables considered, which may have an impact on modeling results.

Using these variables, we calculated (i) three time series of potential evaporation estimates using three commonly used formulas – Penman (Penman, 1948), Penman–Monteith (Monteith, 1965), and Oudin (Oudin et al., 2005) – and (ii) a soil moisture index as computed by the GR4J rainfall-runoff model (Perrin et al., 2003; computed with a reservoir

of 275 mm). Note that our versions of the Penman and the Penman–Monteith formulas use a snow-dependent albedo.

The catchment-scale climatic time series are obtained by spatial aggregation of the SAFRAN pixels intersecting the catchment contours. The SAFRAN variables are aggregated at the catchment scale using a weighted average, based on the proportion of pixels within the catchment boundary.

5.2 Hydrometric time series

The SCHAPI (2022) released the Hydroportail application for accessing hydrometric data at the national scale in January 2022. Hydroportail classifies the hydrometric entities into three levels: hydrometric site, hydrometric station, and hydrometric sensor. The hydrometric site refers to a portion of a watercourse on which flows are considered homogeneous. Several hydrometric (gauging) stations can be associated with a single site. A hydrometric station may include several sensors. The main benefit of dividing the hydrometric entities into these levels is to give the producers the possibility of setting a calendar (i.e., periods of validity of each station and sensor) at the site level and to unify the corresponding streamflow data to produce a consolidated time series with a better qualification procedure.

The CAMELS-FR dataset daily streamflow time series were retrieved from the Hydroportail website using the hydroportail R package (Delaigue, 2022). Streamflow data were retrieved at the station level to get all the data, even for sites where calendar information is missing or incomplete. For stations with sub-daily streamflow data, mean daily streamflows are calculated using a trapezoidal method, thus assuming a linear variation in streamflow between successive instantaneous streamflows. This was done directly by the data producer or by the Hydroportail procedure.

Streamflow data are available in the CAMELS-FR dataset in liters per second and were also converted to water depth (in millimeters per day) using the catchment area derived from the DTM (and not producer's area). As climatic data, the streamflow time series are provided at daily, monthly, and yearly time steps.

5.3 Rules for time series aggregation

The time series are provided at daily, monthly, and yearly time steps. For the calculation of monthly time series, a month was considered missing if it had more than 3 d of missing data. The time series at a yearly time step are provided using the hydrological year (considered to start on 1 October) as the basis of aggregation. Therefore the hydrological year of 1970 (first year of the dataset), which starts on 1 October 1969, is deemed incomplete because the first 3 months are missing. Moreover, these yearly data are computed based on the above-mentioned monthly times series. If 1 month is missing, the whole year is considered missing.

Earth Syst. Sci. Data, 17, 1461–1479, 2025

https://doi.org/10.5194/essd-17-1461-2025

Table 1. List of daily SIM2 variables used in the CAMELS-FR dataset and their aggregation methods at a daily time step.

SIM2 product	Variable	Unit	Aggregation methods	Aggregation time window (UTC)
SAFRAN	Solid precipitation	$mm\,d^{-1}$	Accumulation	06:00–06:00
SAFRAN	Liquid precipitation	$mm\,d^{-1}$	Accumulation	06:00–06:00
SAFRAN	Air temperature	°C	Mean	00:00–24:00
SAFRAN	Daily minimum air temperature	°C	–	00:00–24:00
SAFRAN	Daily maximum air temperature	°C	–	00:00–24:00
SAFRAN	Wind speed	$m\,s^{-1}$	Mean	00:00–24:00
SAFRAN	Specific air humidity	$g\,kg^{-1}$	Mean	00:00–24:00
SAFRAN	Atmospheric radiation	$J\,cm^{-2}\,d^{-1}$	Accumulation	00:00–24:00
SAFRAN	Visible radiation	$J\,cm^{-2}\,d^{-1}$	Accumulation	00:00–24:00
ISBA	Snow water equivalent	$mm\,d^{-1}$	Accumulation	06:00–06:00
ISBA	Soil wetness index	–	Mean	06:00–06:00

Table 2. Number of catchment attributes in each class for the CAMELS-FR dataset v1 (the "other" class is related to data access information).

Class	No.
Location and topography	91
Climatic indices	51
Hydrometry	18
Hydrological signatures	24
Hydrogeology	11
Geology	17
Soil	25
Land cover	10
Intervention degree	7
Other	1

6 Catchment attributes

The CAMELS-FR dataset v1 contains 255 attributes, organized into 10 classes (see Table 2), described in the following sub-sections. A spreadsheet listing each catchment attribute is included in the CAMELS-FR dataset.

6.1 Location

Most of the location attributes are information given by the data producer and extracted from Hydroportail. Such attributes include the station codes, names, location, type of station, etc. The data producer gives, for most catchments, an estimate of catchment area. Note that for some catchments this area is significantly different from the one derived from the DTM analysis (see Sect. 4.3). Other catchment attributes were computed for the CAMELS-FR dataset. They include station coordinates (after repositioning on the DTM-derived river network), catchment area (derived from the DTM), and information on the station nestedness (indicating whether the catchment is nested, the number of stations downstream, etc.).

6.2 Topography

Two DTMs (NASA and the National Geospatial-Intelligence Agency, NGA, DTM from the SRTM for catchments located on the continent and the BD ALTI v1.0 (IGN) for catchments in Corsica) were used to estimate various attributes describing catchment topography, drainage density, and morphometry. Among the topography attributes are the mean and the percentile distribution for elevation, slope, distance to catchment outlet, and topographic index. The methodology used to calculate the topographic index follows Ducharne (2009), which reformulates the index used in TOPMODEL (Beven and Kirby, 1979) to become a dimensionless index. The percentage of catchment slope classes (flat, gentle, moderate, strong, steep, and very steep) and orientation and the classical drainage density, which is measured as the ratio of total length of stream channels to catchment area (Horton, 1932), are also provided. Among the morphometry attributes, several catchment shape indicators are provided. These include three variants of basin form factor (Horton, 1932; Zăvoianu, 1978, cited by Zăvoianu, 1985, p. 109; Subramanya, 2013, p. 172), the compactness coefficient (Fitzpatrick, 2017), the circularity ratio proposed by Miller (1953) (cited by Zăvoianu, 1985, p. 104), the catchment relief ratio (Fryirs and Brierley, 2013, p. 35), and two variations of elongation ratio: (i) using a circle as a reference (Schumm, 1956) and (ii) using the catchment area as a reference (Fryirs and Brierley, 2013, p. 34).

6.3 Climatic indices

Numerous climatic attributes were estimated: annual mean of different variables (air temperature, potential evaporation, precipitation) over the entire studied period, aridity and seasonality indices, measure of the asynchronicity between the precipitation and potential evaporation, etc. The attributes that use potential evaporation are computed three times using the Penman, Penman–Monteith and Oudin formulas. The attributes of heavy-precipitation-day frequency

(frequency of high-precipitation days, i.e., larger than 5 times the mean daily precipitation) and intensity (strongest rainfall on record) and dry-day frequency (average duration of dry periods, i.e., number of consecutive days with precipitation lower than $1 \, \mathrm{mm \, d^{-1}}$) were also computed. Time series of statistics at an annual time step using hydrological year are also provided, including annual daily maximum precipitation and annual seasonality index. As mentioned in Sect. 5.3, the first year (1970) is always missing. Furthermore, maps with a resolution of $1 \, \mathrm{km} \times 1 \, \mathrm{km}$ of hourly and daily rainfall intensity, developed by Arnaud et al. (2008) from the SHYREG-Pluie database (Base de données SHYREG-Pluie ©, INRAE, 2016, all rights reserved), were used to calculate a normalized indicator of rainfall intensity at the catchment scale. This indicator is defined as the ratio of the hourly to daily rainfall intensity with 10-year return period (Poncelet, 2016).

6.4 Hydrometry

Several CAMELS-FR dataset attributes are related to streamflow data quality estimated for each gauging station. Some of these attributes are provided by the data producers. For example, the overall level of uncertainty is qualified by the producer for three flow ranges (high, mean, and low flow). Moreover, several quality codes are assigned to each streamflow value, depending on the way the value was estimated (classical flow estimation using a water level measurement and a rating curve or a value reconstituted a posteriori) (see Sect. 5.2).

Note that this information lacks consistency throughout the CAMELS-FR catchment set, since this qualification is done by regional services, which have different qualification methods according to the regional hydroclimatic context.

Additional attributes were estimated for each station to describe the number of gaps over the studied period (e.g., percentage of missing data in the total period, 1 January 1970 to 31 December 2021) and the overall percentage of streamflow data flagged as questionable or as unqualified by the data producer. A simple algorithm was also applied to identify potential errors due to linear interpolation between two values.

Finally, we attributed to each station an overall estimation of low-flow quality based on the visual analysis of temporal streamflow series: stations with low flows showing suspicious data or behavior have been identified; 52 % of catchments in the CAMELS-FR dataset v1 exhibit such issues.

6.5 Hydrological signatures

As for catchment climate, various hydrological attributes were estimated to describe the main catchment hydrological features. These include catchment aridity (i.e., the ratio of mean daily potential evaporation to mean daily precipitation), catchment yield, streamflow elasticity to precipitation, and catchment seasonality (e.g., month with the mini-

mum mean monthly streamflow). The baseflow index (BFI) was estimated following three different approaches: (i) the method proposed by Ladson et al. (2013) that uses a digital filter, (ii) the method proposed by Gustard and Tallaksen (2008) through linear interpolation of a 5 d non-overlapping streamflow minima (computed with the lfstat R package; Laaha and Koffler, 2022), and (iii) the method proposed by Pelletier and Andréassian (2020) that uses a conceptual quadratic reservoir model (computed with the baseflow R package; Pelletier et al., 2021). For the latter, before calculating the BFI, we filled the gaps in time series with GR6J-CemaNeige (Pushpalatha et al., 2011; Valéry et al., 2014a, b) using the Penman (Penman, 1948), Penman–Monteith (Monteith, 1965), and Oudin (Oudin et al., 2005) formulas. However, we provide the BFI using only the Oudin potential evaporation formula because there were no significant BFI differences when using other formulas. The other attributes that use potential evaporation are computed three times using the three selected formulas. Time series of statistics at an annual time step are also provided for annual maximum daily streamflow and the mean monthly annual minimum streamflow (QMNA) using hydrological and calendar year, respectively.

6.6 Hydrogeology

The average catchment permeability and porosity were estimated thanks to the GLobal HYdrogeology MaPS database (GLHYMPS v2.0) (Huscroft et al., 2018). The national hydrogeological reference map (BDLISA v3) (Brugeron et al., 2018; BRGM, 2022) was used to calculate the karstic portion of the catchments and the percentages of each catchment covered by different hydrogeological formations (e.g., bedrock, sedimentary, alluvial zones).

6.7 Geology

The Global Lithological Map database (GLiM v1.0) (Hartmann and Moosdorf, 2012a, b) was used to characterize the catchment geology. This database, available at a 0.5° spatial resolution, provided estimates of (i) the dominant geological class of each catchment and (ii) the percentages of different lithologies (e.g., metamorphics, carbonate sedimentary rocks; see Fig. 2c).

6.8 Soil

Soil characteristics were described using four datasets. First, the Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers v1 produced by Pelletier et al. (2016) was cropped in order to estimate, for each catchment, (i) the mean soil depth and (ii) the distribution (percentiles) of the soil depths over the catchment. The European Soil Database Derived data (ESDD) (JRC et al., 2013a, b) provide soil attributes with a resolution of $1 \, \mathrm{km} \times 1 \, \mathrm{km}$ for a

topsoil layer and a subsoil layer with the boundary at 30 cm soil depth for clay, sand, silt, and organic carbon content; bulk density; coarse fragments; and total available water content (TAWC). We therefore, aggregated these two layers from ESDD using the weighted arithmetic mean in cells based on their soil depth for each pixel. We consider the attribute "depth available to roots" to be the correspondent soil depth for each pixel. For the aggregation of TAWC, instead of calculating the weighted arithmetic mean, we calculated the sum of the two layers. Then we cropped the raster for each catchment to extract the percentiles, mean, and skewness of each attribute for the topsoil and the aggregated layer. The EU-SoilHydroGrids v1.0 (Tóth et al., 2017) with a resolution of $250\,\text{m} \times 250\,\text{m}$ was used to extract the saturated hydraulic conductivity. This dataset has seven soil layers at 0, 5, 15, 30, 60, 100, and 200 cm depth. The first four layers, which are less than or equal to 30 cm depth, are considered to be topsoil. Therefore, we aggregated these layers to provide information on topsoil (using the first four layers) and the total seven layers as well using depth-weighted harmonic mean. The hydraulic saturated conductivity was divided by 24 in order to provide the results in $\text{cm}\,\text{h}^{-1}$ and then was cropped for each catchment to extract the percentiles, mean, and skewness. Finally, the TAWC was also estimated by IN-RAE over France from the Réservoir utile des sols de la France métropolitaine v1.2 database (Roman Dobarco et al., 2021; Le Bas, 2021) with a resolution of $90\,\text{m} \times 90\,\text{m}$, providing another distribution of TAWC at the catchment scale.

6.9 Land cover

The CORINE Land Cover dataset was used to estimate, for each catchment, (i) the dominant land cover class and the (ii) percentage of each class on two dates (French Ministry of the Environment, 1990, 2018). This analysis has been performed for all aggregation levels (1, 2, and 3) of the CORINE Land Cover classification.

6.10 Intervention degree

Influence level attributes gather (i) information given by the data producer on the estimated anthropogenic impacts on the observed streamflow time series (four attributes) and (ii) metrics we calculated to estimate the influence of dams present within each catchment (three attributes). Thus, we listed, for each catchment, the number of dams and the total volume of artificial storage.

7 Dataset description and add-on products

7.1 Illustration of several attributes of the CAMELS-FR dataset

Figure 5b shows the distribution of four climatic characteristics (mean annual precipitation, mean annual potential evap-

oration, seasonality index, and maximum annual precipitation) of the CAMELS-FR dataset catchment set, grouped by hydrological region (see Fig. 5a). The northwestern catchments have lower annual precipitation than those in the south, the latter being characterized by high variability in annual precipitation and also in mean annual potential evaporation. Southern catchments show lower seasonality index values, due to Mediterranean climate characterized by a winter maximum for precipitation and significant summer drought (de Lavenne and Andréassian, 2018). Figure 5c shows each CAMELS-FR dataset catchment in a Budyko-type nondimensional space (Andréassian and Perrin, 2012) and for each hydrological region, showing the diversity of hydroclimatic context. Several catchments are located over the water limit ($Q = P$) or under the energy limit ($Q = P - PET$) due to potential misestimation of precipitation, potential evaporation, or streamflow; errors in catchment area estimation; or non-conservative catchment behavior.

Figure 6 presents maps and distributions of three climatic attributes (mean annual precipitation, mean annual potential evaporation, and seasonality index). One can clearly identify the regions with higher-than-average mean precipitation (Fig. 6a): Brittany in the west and all the mountain ranges (i.e., Vosges, Jura, Alps, Massif central, and the Pyrenees). Note that the spatial distribution of extreme precipitation shows a clear structure along a northwest–southeast axis (Fig. 6b), while potential evaporation is structured along a north–south axis, with some variations induced by the mountain ranges (Fig. 6c). Streamflow-related indices (Fig. 6d–f) do not show clear large-scale patterns because they reflect the complex interplay between the climatic factors and the pedo-lithologic factors (see Fig. 2b and c).

7.2 CAMELS-FR add-on products

Two side products were developed around the CAMELS-FR dataset, to provide tools to visualize catchment-scale time series and synthetic summaries to give the user a good overview of the data.

- CAMELS-FR graphical fact sheets (Delaigue et al., 2024a): static plots summarizing hydroclimatic, topographical, hydrogeological, and land cover data (available in English and French) (see example Fig. 7);

- CAMELS-FR time series dynamic graphs (Delaigue et al., 2024d): HTML pages with dynamic plots of hydroclimatic time series (click-and-drag zooming and "onmouseover" display of legends and values; available in English and French) (see example Fig. 8).
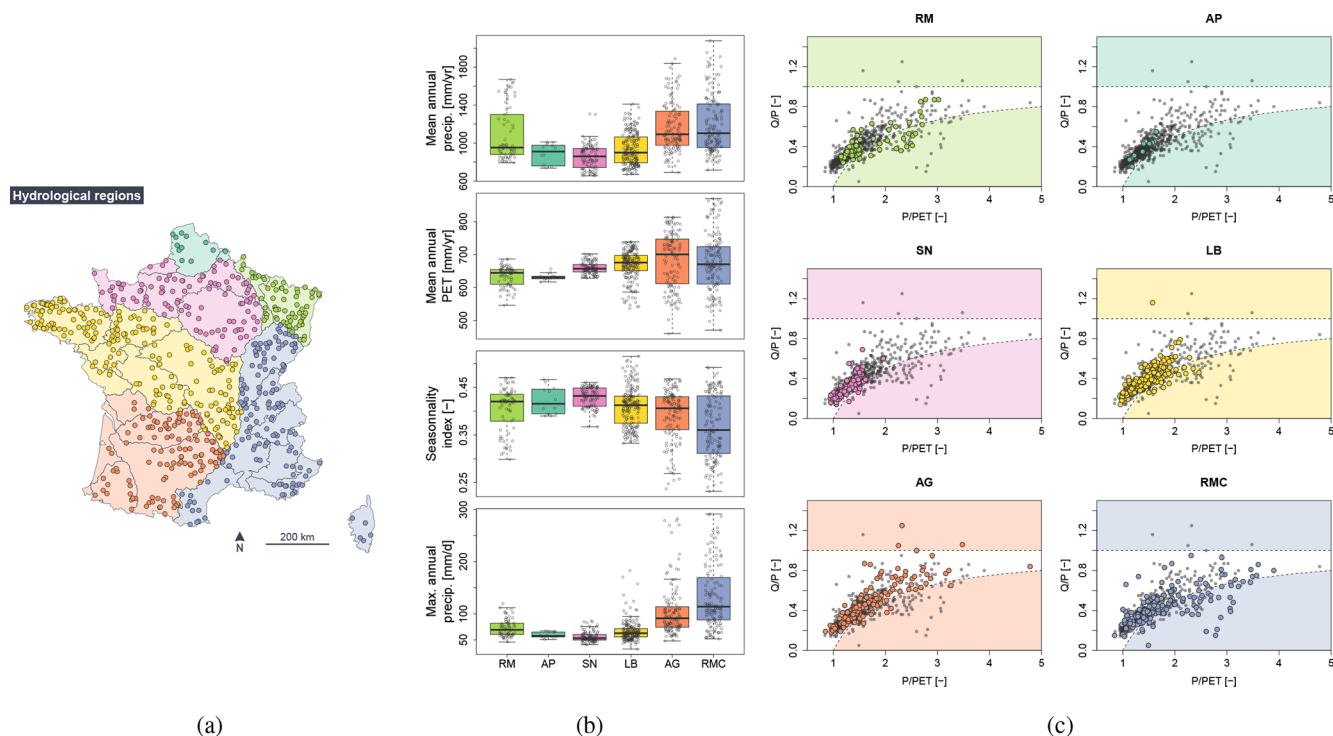
Figure 5. Distribution of CAMELS-FR catchment set over hydrological regions (AG: Adour–Garonne; AP: Artois–Picardie; LB: Loire–Bretagne; RM: Rhin–Meuse; RMC: Rhône–Méditerrannée–Corse; SN: Seine–Normandie; hydrological region boundaries: French water agencies, 2017a): **(a)** geographical locations, **(b)** boxplots of climatic indices, **(c)** projection onto a Budyko-like space (large colored dots: catchments inside the given hydrological region; small gray dots: all the other catchments).

8   Data availability

The CAMELS-FR dataset (Delaigue et al., 2024c) can be freely downloaded from the French governmental research data warehouse (https://entrepot.recherche.data.gouv.fr/, last access: 4 April 2025) using the following digital object identifier: https://doi.org/10.57745/WH7FJR. The CAMELS-FR time series dynamic graphs (Delaigue et al., 2024d) and the CAMELS-FR graphical fact sheets (Delaigue et al., 2024a) also have their own digital object identifier, https://doi.org/10.57745/KK2SVJ and https://doi.org/10.57745/HBQWP5, respectively. Most of the variables included in the dataset are based on the French State Open License version 2.0 (https://www.etalab.gouv.fr/licence-ouverte-open-licence, last access: 23 October 2024), which is compatible with the CC BY license. A few are based either on CC BY licenses version 3.0 or 4.0 or specific licenses (provided with the dataset). The CAMELS-FR dataset is distributed under the CC BY license version 4.0 (https://creativecommons.org/licenses/by/4.0, last access: 23 October 2024). Each modification to the dataset can be tracked, as the version number will be automatically updated. Older releases will remain available, even if the data have been updated. A "NEWS" file will be updated in order to track changes between versions.
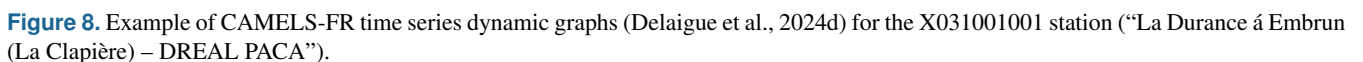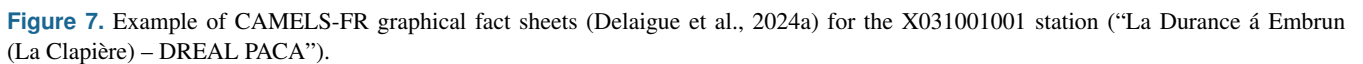
9   Conclusions and perspectives

The CAMELS-FR dataset v1 gathers data over 654 catchments located in France, including daily hydroclimatic time series over the 1970–2021 period and 255 attributes split into 10 classes (see Table 2). These catchments represent a significant diversity of hydroclimatic contexts (e.g., snow- or groundwater-dominated catchments, Mediterranean catchments). The catchment selection was based on four criteria: (i) streamflow data availability over the 1970–2021 period, (ii) limited artificial influences (quantified based on upstream dam storage only), (iii) consistency between catchment area provided by the data producer and estimated by the DTM analysis, and (iv) visual analysis of the streamflow time series (at a daily time step). We did not consider any selection criteria based on hydrological model efficiency.

The CAMELS-FR dataset was designed to be a "living" dataset. Several changes and updates are planned in subsequent versions: time series lengthening, streamflow value correction by the data producers, addition of "new" catchments (e.g., from French overseas territories). Versions with data at finer temporal and spatial resolutions may also be developed. A web app devoted to the selection of catchment subsets using hydroclimatic, topographical, hydrogeological, or land cover criteria is currently under development.
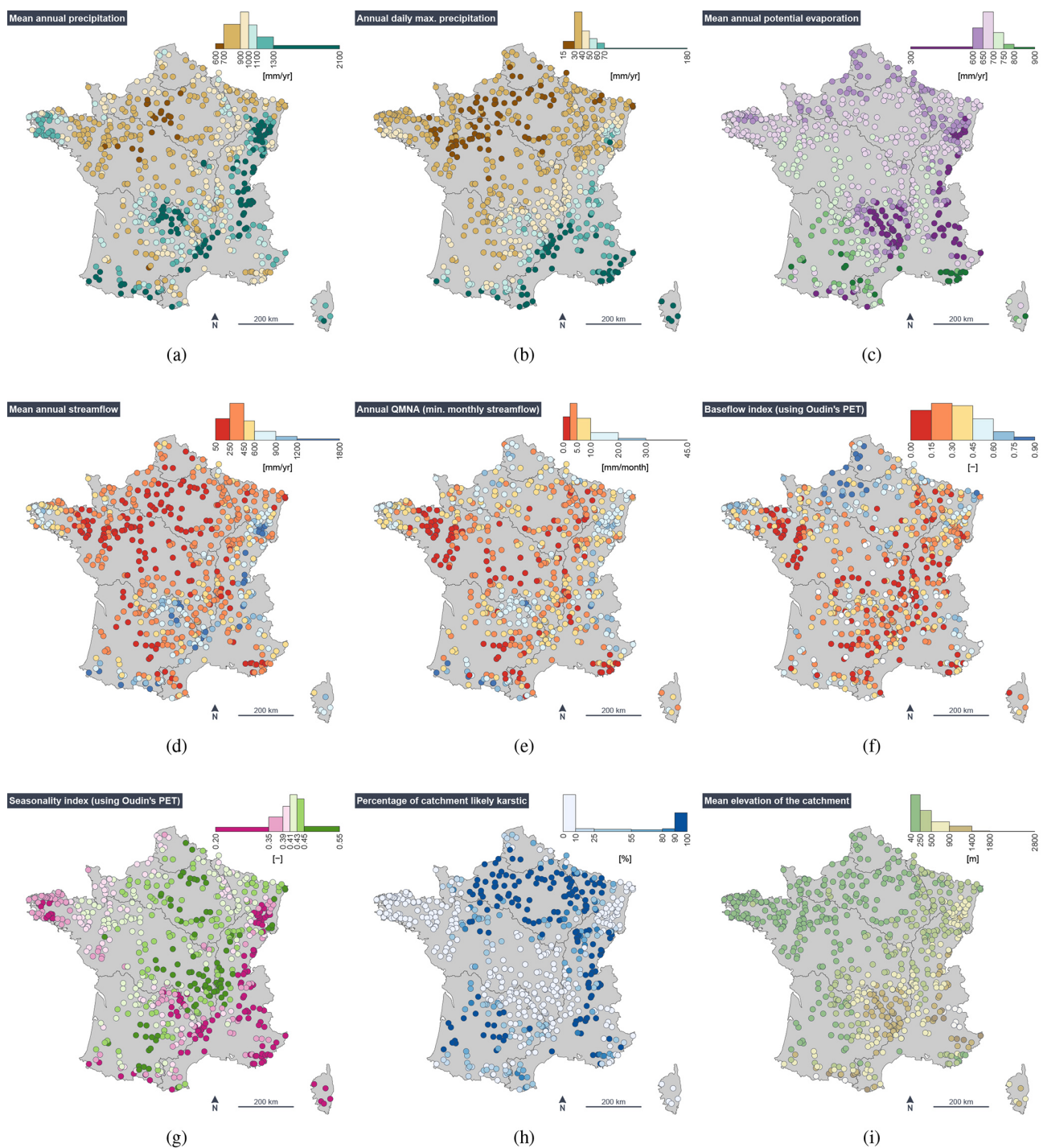
Figure 6. Maps and distributions of a few catchment-scale attributes (hydrological region boundaries: French water agencies, 2017a). **(a)** Mean annual precipitation, **(b)** mean annual daily maximum precipitation, **(c)** mean annual long-term potential evaporation (Oudin method), **(d)** mean annual streamflow, **(e)** mean annual minimum monthly streamflow (QMNA), **(f)** baseflow index using Pelletier and Andréassian (2020), **(g)** seasonality index (de Lavenne and Andréassian, 2018), **(h)** percentage of catchment that is likely karstic, **(i)** mean elevation of the catchment. Note that for those indices requiring a potential evaporation estimate, we consistently used the Oudin method.

Figure 7. Example of CAMELS-FR graphical fact sheets (Delaigue et al., 2024a) for the X031001001 station ("La Durance á Embrun (La Clapière) – DREAL PACA").

Figure 8. Example of CAMELS-FR time series dynamic graphs (Delaigue et al., 2024d) for the X031001001 station ("La Durance á Embrun (La Clapière) – DREAL PACA").

A CAMELS-FR dataset extension to the world-wide Caravan initiative (Kratzert et al., 2023) is also foreseen.

References

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.

Andréassian, V. and Perrin, C.: On the ambiguous interpretation of the Turc-Budyko nondimensional graph, Water Resour. Res., 48, W10601, https://doi.org/10.1029/2012WR012532, 2012.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrol. Earth Syst. Sci., 13, 1757–1764, https://doi.org/10.5194/hess-13-1757-2009, 2009.

Andréassian, V., Perrin, C., Parent, E., and Bárdossy, A.: The Court of Miracles of Hydrology: can failure stories contribute to hydrological science?, Hydrolog. Sci. J., 55, 849–856, https://doi.org/10.1080/02626667.2010.506050, 2010.

Arnaud, P., Lavabre, J., Sol, B., and Descouches, C.: Regionalization of an hourly rainfall generating model over metropolitan France for flood hazard estimation, Hydrolog. Sci. J., 53, 34–47, https://doi.org/10.1623/hysj.53.1.34, 2008.

Arsenault, R., Bazile, R., Ouellet Dallaire, C., and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrol. Process., 30, 2734–2736, https://doi.org/10.1002/hyp.10880, 2016.

Beven, K. J. and Kirby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrol. Sci. B., 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Bourgin, P.-Y., Lobligeois, F., Peschard, J., Andréassian, V., Le Moine, N., Coron, L., Perrin, C., Ramos, M.-H., and Khalifa, A.: Description des caractéristiques morphologiques, climatiques et hydrologiques de 4436 bassins versants français. Guide d'utilisation de la base de données hydro-climatique, Technical report, IRSTEA, https://hal.science/hal-02596718 (last access: 12 September 2024), 2010.

BRGM: BDLISA. Base de donnée des limites des systèmes aquifères, version 3.0, Bureau de recherches géologiques et minières (BRGM), Eaufrance [data set], https://bdlisa.eaufrance.fr/ (last access: 12 November 2023), 2022.

Brugeron, A., Paroissien, J., and Tillier, L.: Référentiel hydrogéologique BDLISA version 2 : Principes de construction et évolutions. Rapport final, Tech. Rep. BRGM/RP-67489-FR, Bureau de recherches géologiques et minières (BRGM), 2018.

CFBR: CFBR website, Comité français des barrages et réservoirs (CFBR), https://www.barrages-cfbr.eu/ (last access: 15 January 2023), 2023.

https://doi.org/10.5194/essd-17-1461-2025

Earth Syst. Sci. Data, 17, 1461–1479, 2025

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleis-chmann, 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.

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, 2020.

de Lavenne, A. and Andréassian, V.: Impact of climate seasonality on catchment yield: A parameterization for commonly-used water balance formulas, J. Hydrol., 558, 266–274, https://doi.org/10.1016/j.jhydrol.2018.01.009, 2018.

de Marsily, G., Combes, P., and Goblet, P.: Comment on 'Ground-water models cannot be validated', by L. F. Konikow & J. D. Bredehoeft, Adv. Water Resour., 15, 367–369, https://doi.org/10.1016/0309-1708(92)90003-K, 1992.

Delaigue, O.: hydroportail: Retrieve French Hydrological Data from Hydroportail, R package version 0.1.0.9006, INRAE [code], https://gitlab.irstea.fr/HYCAR-Hydro/hydroportail (last access: 12 November 2023), 2022.

Delaigue, O., Brigode, P., Lobligeois, F., Bourgin, P.-Y., and Guimarães, G. M.: CAMELS-FR graphical fact sheets, V1, Recherche Data Gouv [data set], https://doi.org/10.57745/KK2SVJ, 2024a.

Delaigue, O., Guimarães, G. M., Brigode, P., Andréassian, V., Payan, J.-L., Steinbach, P., and Kreutzenberger, K.: MADAM: Metropolitan Area Dams, V1, Recherche Data Gouv [data set], https://doi.org/10.57745/N98NEN, 2024b.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., and Andréassian, V.: CAMELS-FR dataset, V1, Recherche Data Gouv [data set], https://doi.org/10.57745/WH7FJR, 2024c.

Delaigue, O., Génot, B., and Guimarães, G. M.: CAMELS-FR time series dynamic graphs, V1, Recherche Data Gouv [data set], https://doi.org/10.57745/HBQWP5, 2024d.

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, Zenodo [data set], https://doi.org/10.5281/zenodo.13961394, 2024a.

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, Scientific Data, 11, 879, https://doi.org/10.1038/s41597-024-03706-1, 2024b.

Ducharne, A.: Reducing scale dependence in TOPMODEL using a dimensionless topographic index, Hydrol. Earth Syst. Sci., 13, 2399–2412, https://doi.org/10.5194/hess-13-2399-2009, 2009.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, https://doi.org/10.1029/2005RG000183, 2007.

Fitzpatrick, F. A.: Watershed Geomorphological Characteristics, in: Handbook of applied hydrology, 2nd Edn., edited by: Singh, V. P., McGraw-Hill Education, New York, 44-1–44-12, ISBN 9780071835091, 2017.

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.

French Ministry of the Environment: CORINE Land Cover. Occupation des sols en France, version 1990, Data Gouv [data set], https://www.data.gouv.fr/fr/datasets/corine-land-cover-occupation-des-sols-en-france/ (last access: 12 November 2023), 1990.

French Ministry of the Environment: CORINE Land Cover. Occupation des sols en France, version 2018, Data Gouv [data set], https://www.data.gouv.fr/fr/datasets/corine-land-cover-occupation-des-sols-en-france/ (last access: 12 November 2023), 2018.

French water agencies: BD Carthage: Régions hydrographiques de Métropole 2017, version 2021-06-09, Eaufrance [data set], https://www.sandre.eaufrance.fr/atlas/srv/fre/catalog.search#/metadata/485cfb8d (last access: 12 November 2023), 2017a.

French water agencies: BD Carthage: Cours d'eau de Métropole 2017, version 2019-10-24, Eaufrance [data set], https://www.sandre.eaufrance.fr/atlas/srv/fre/catalog.search#/metadata/7381de46-42f7-42df-9abe-0ecd4b946034 (last access: 12 November 2023), 2017b.

Fryirs, K. A. and Brierley, G. J.: Geomorphic analysis of river systems: an approach to reading the landscape, 1 edn., Wiley-Blackwell, Chichester, ISBN 9781405192750, 2013.

Génot, B. and Delaigue, O.: Graphical user interface to help snapping hydrometric stations on the theoretical river network, INRAE, 2018.

Giuntoli, I., Renard, B., Vidal, J.-P., and Bard, A.: Low flows in France and their relationship to large-scale climate indices, J. Hydrol., 482, 105–118, https://doi.org/10.1016/j.jhydrol.2012.12.038, 2013.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18, 463–477, https://doi.org/10.5194/hess-18-463-2014, 2014.

Gustard, A. and Tallaksen, L. M.: Low-Flow indices, in: Manual on Low-flow Estimation and Prediction, vol. 50 of Operational hydrology report, edited by: Gustard, A. and Demuth, S., WMO, Geneva, p. 138, https://library.wmo.int/records/item/32176-manual-on-low-flow-estimation-and-prediction (last access: 4 April 2025), 2008.

Hartmann, J. and Moosdorf, N.: Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution), PANGAEA [data set], https://doi.org/10.1594/PANGAEA.788537, 2012a.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochem. Geophy. Geosy., 13, Q12004, https://doi.org/10.1029/2012GC004370, 2012b.

Hauffe, C., Brandes, C., Lei, K., Pahner, S., Körner, P., Kronenberg, R., and Schuetze, N.: CAMELS-SAX: A meteorological

and hydrological dataset for spatially distributed modeling of catchments in Saxony, https://doi.org/10.5194/egusphere-egu23-14357, conference Name: EGU23, 2023.

Hodgkins, G. A., Renard, B., Whitfield, P. H., Laaha, G., Stahl, K., Hannaford, J., Burn, D. H., Westra, S., Fleig, A. K., Araújo Lopes, W. T., Murphy, C., Mediero, L., and Hanel, M.: Climate Driven Trends in Historical Extreme Low Streamflows on Four Continents, Water Resour. Res., 60, e2022WR034326, https://doi.org/10.1029/2022WR034326, 2024.

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.: 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.

Horton, R. E.: Drainage-basin characteristics, Eos T. Am. Geophys. Un., 13, 350–361, https://doi.org/10.1029/TR013i001p00350, 1932.

Huscroft, J., Gleeson, T., Hartmann, J., and Börker, J.: Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS (GLHYMPS), version 2.0, Borealis [data set], https://doi.org/10.5683/SP2/TTJNIU, 2018.

IGN: BD ALTI, version 1.0., Institut national de l'information géographique et forestière (IGN) [data set], https://geoservices.ign.fr/bdalti/ (last access: 12 November 2023), 2001.

IGN: Plan IGN, version 2.0, Institut national de l'information géographique et forestière (IGN) [data set], https://geoservices.ign.fr/planign (last access: 12 November 2023), 2020.

INRAE: Base de données SHYREG-pluie, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) [data set], https://shyreg.pluie.recover.inrae.fr/ (last access: 8 November 2016), 2016.

JRC, IES, and Hiederer, R.: Mapping soil properties for Europe – Spatial representation of soil database attributes, Tech. rep., Joint Research Centre and Institute for Environment and Sustainability, https://doi.org/10.2788/94128, 2013a.

JRC, IES, and Hiederer, R.: Mapping soil typologies – Spatial decision support applied to the European Soil Database, Tech. rep., Joint Research Centre and Institute for Environment and Sustainability, https://doi.org/10.2788/87286, 2013b.

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.

Koch, J.: Catchment Dataset Denmark, version 1.0, GEUS Dataverse [data set], https://doi.org/10.22008/FK2/YCQXTR, 2021.

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, Scientific Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Laaha, G. and Koffler, D.: lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data, R package version 0.9.12, CRAN [code], https://doi.org/10.32614/CRAN.package.lfstat, 2022.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A Standard Approach to Baseflow Separation Using The Lyne and Hollick Filter, Australasian Journal of Water Resources, 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

Le Bas, C.: Carte de la Réserve Utile en eau issue de la Base de Données Géographique des Sols de France, version 2.3, Recherche Data Gouv [data set], https://doi.org/10.15454/JPB9RB, 2021.

Le Moigne, P., Besson, F., Martin, E., Boé, J., Boone, A., Decharme, B., Etchevers, P., Faroux, S., Habets, F., Lafaysse, M., Leroux, D., and Rousset-Regimbeau, F.: The latest improvements with SURFEX v8.0 of the Safran–Isba–Modcou hydrometeorological model for France, Geosci. Model Dev., 13, 3925–3946, https://doi.org/10.5194/gmd-13-3925-2020, 2020.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process., 27, 2171–2186, https://doi.org/10.1002/hyp.9740, 2013.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos T. Am. Geophys. Un., 89, 93–94, https://doi.org/10.1029/2008EO100001, 2008.

Linsley, R.: Rainfall-runoff models: an overview, in: Rainfall-runoff relationship, edited by: Singh, V. P., Water Resources Publications, Chelsea, Michigan, US, bookcrafters inc. edn., meeting Name: International Symposium on Rainfall Runoff Modeling, 3–22, ISBN 9780918334459, 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 Catchments in Denmark, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2024-292, in review, 2024.

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.: 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.

Miller, V. C. A.: Quantitative Geomorphic Study of Drainage Basin Characteristics in the Clinch Mountain Area, Virginia and Tennessee, Tech. Rep. 3, Columbia University, Department of Geology, New York, 1953.

Monteith, J.: Evaporation and environment, Symposia of the Society for Experimental Biology, Copenhagen, 1965, 19, 205–234, http://europepmc.org/abstract/MED/5321565 (last access: 4 April 2025), 1965.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

NOAA: Global Self-consistent, Hierarchical, High-resolution Geography Database (GSHHG), version 2.3.7, National Oceanic and Atmospheric Administration (NOAA) [data set], https://www.ngdc.noaa.gov/mgg/shorelines/ (last access: 12 November 2023), 2017.

OFB and partners: Géoréférenceur des obstacles [Application GEOBS, en ligne], Référentiel des obstacles à l'écoulement (Module 1, ROE) et Base de données complémentaire sur les obstacles à l'écoulement (Module 3, BDOe), version 2023-02-21, Office français de la biodiversité (OFB), Eaufrance [data set], https://geobs.eaufrance.fr/, (last access: 12 November 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, 00236, 2005.

Payan, J.-L.: Prise en compte de barrages-réservoirs dans un modèle global pluie-débit, PhD thesis, ENGREF (AgroParisTech), http://theses.hal.science/pastel-00003555 (last access: 4 April 2025), 2007.

Payan, J.-L., Perrin, C., Andréassian, V., and Michel, C.: How can man-made water reservoirs be accounted for in a lumped rainfall-runoff model?, Water Resour. Res., 44, W03420, https://doi.org/10.1029/2007WR005971, 2008.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, https://doi.org/10.5194/hess-11-1633-2007, 2007.

Pelletier, A.: Nappes et rivières: la piézométrie peut-elle améliorer la prévision des étiages des cours d'eau?, PhD thesis, Sorbonne Université, Paris, France, https://theses.hal.science/tel-03783485 (last access: 4 April 2025), 2021.

Pelletier, A. and Andréassian, V.: Hydrograph separation: an impartial parametrisation for an imperfect method, Hydrol. Earth Syst. Sci., 24, 1171–1187, https://doi.org/10.5194/hess-24-1171-2020, 2020.

Pelletier, A., Andréassian, V., and Delaigue, O.: baseflow: Computes Hydrograph Separation, R package version 0.13.2, Recherche Data Gouv [code], https://doi.org/10.15454/Z9IK5N, 2021.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z. C., Brunke, M. A., and Gochis, D.: Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers, ORNL DAAC, https://doi.org/10.3334/ORNLDAAC/1304, 2016.

Penman, H. L.: Natural evaporation from open water, bare soil and grass, P. Roy. Soc. Lond. A Mat., 193, 120–145, https://doi.org/10.1098/rspa.1948.0037, 1948.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Poncelet, C.: Du bassin au paramètre: jusqu'où peut-on régionaliser un modèle hydrologique conceptuel?, PhD thesis, Université Pierre et Marie Curie – Paris VI, Antony, France, https://theses.hal.science/tel-01529196 (last access: 4 April 2025), 2016.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, J. Hydrol., 411, 66–76, https://doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Quintana-Segui, P., Le Moigne, P., 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. Clim., 47, 92–107, https://doi.org/10.1175/2007JAMC1636.1, 2008.

Rabus, B., Eineder, M., Roth, A., and Bamler, R.: The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar, ISPRS J. Photogramm., 57, 241–262, https://doi.org/10.1016/S0924-2716(02)00124-7, 2003.

Refsgaard, J. C. and Hansen, J. R.: A good-looking catchment can turn into a modeller's nightmare, Hydrolog. Sci. J., 55, 899–912, https://doi.org/10.1080/02626667.2010.505571, 2010.

Roman Dobarco, M., Bourennane, H., Arrouays, D., Saby, N., Cousin, I., and Martin, M. P.: Réservoir utile des sols de la France métropolitaine, version 1.2, Recherche Data Gouv [data set], https://doi.org/10.15454/9IRARJ, 2021.

Schaake, J., Cong, S., and Duan, Q.: The US mopex data set, IAHS Publication Series, 307, 9–28, Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States), https://www.osti.gov/biblio/899413 (last access: 4 April 2025), 2006.

SCHAPI: Hydroportail. Site de référence d'accès aux données hydrométriques et hydrologiques en France, Service central d'hydrométéorologie et d'appui à la prévision des inondations (SCHAPI), Eaufrance [data set], https://www.hydro.eaufrance.fr/ (last access: 20 January 2025), 2022.

Schumm, S. A.: Evolution of Drainage Systems & Slopes in Badlands at Perth, New Jersey, Geol. Soc. Am. Bull., 67, 597, https://doi.org/10.1130/0016-7606(1956)67[597:EODSAS]2.0.CO;2, 1956.

Strohmenger, L., Sauquet, E., Bernard, C., Bonneau, J., Branger, F., Bresson, A., Brigode, P., Buzier, R., Delaigue, O., Devers, A., Evin, G., Fournier, M., Hsu, S.-C., Lanini, S., de Lavenne, A., Lemaitre-Basset, T., Magand, C., Mendoza Guimarães, G., Mentha, M., Munier, S., Perrin, C., Podechard, T., Rouchy, L., Sadki, M., Soutif-Bellenger, M., Tilmant, F., Tramblay, Y., Véron, A.-L., Vidal, J.-P., and Thirel, G.: On the visual detection of non-natural records in streamflow time series: challenges and impacts, Hydrol. Earth Syst. Sci., 27, 3375–3391, https://doi.org/10.5194/hess-27-3375-2023, 2023.

Strohmenger, L., Collet, L., Andréassian, V., Corre, L., Rousset, F., and Thirel, G.: Köppen–Geiger climate classification across France based on an ensemble of high-resolution climate projections, CR Geosci., 356, 67–82, https://doi.org/10.5802/crgeos.263, 2024.

Subramanya, K.: Engineering Hydrology, 4th edn., McGraw Hill Education, New Delhi, ISBN 9781259029974, 2013.

Tóth, B., Weynants, M., Pásztor, L., and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrol. Process., 31, 2662–2666, https://doi.org/10.1002/hyp.11203, 2017.

Tramblay, Y., Rouché, N., Paturel, J.-E., Mahé, G., Boyer, J.-F., Amoussou, E., Bodian, A., Dacosta, H., Dakhlaoui, H., Dezetter, A., Hughes, D., Hanich, L., Peugeot, C., Tshimanga, R., and Lachassagne, P.: ADHI: the African Database of Hydrometric Indices (1950–2018), Earth Syst. Sci. Data, 13, 1547–1560, https://doi.org/10.5194/essd-13-1547-2021, 2021.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-

accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments, J. Hydrol., 517, 1166–1175, https://doi.org/10.1016/j.jhydrol.2014.04.059, 2014a.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, J. Hydrol., 517, 1176–1187, https://doi.org/10.1016/j.jhydrol.2014.04.058, 2014b.

Vidal, J., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.: 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.

Vincent, C., Peyaud, V., Laarman, O., Six, D., Gilbert, A., Gillet-Chaulet, F., Berthier, E., Morin, S., Verfaillie, D., Rabatel, A., Jourdain, B., and Bolibar, J.: Déclin des deux plus grands glaciers des Alpes françaises au cours du XXIe siècle : Argentière et Mer de Glace, La Météorologie, 106, 49–58, https://doi.org/10.4267/2042/70369, 2019.

Wessel, P. and Smith, W. H. F.: A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res.-Sol. Ea., 101, 8741–8743, https://doi.org/10.1029/96JB00104, 1996.

Zăvoianu, I.: Morfometria bazinelor hidrografice, Editura Academiei Republicii Socialiste România, Bucharest, ISBN 9780080870113, 1978.

Zăvoianu, I.: Morphometry of drainage basins, 2nd edn., no. 20 in Developments in water science, Elsevier, Amsterdam, Oxford, New York, Tokyo, ISBN 9780080870113, 1985.

https://doi.org/10.5194/essd-17-1461-2025

Earth Syst. Sci. Data, 17, 1461–1479, 2025