https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

CAMELS-FI: hydrometeorological time series and landscape properties for 320 catchments in Finland

Iiro Seppä[1], Carlos Gonzales Inca[1], Jari Uusikivi[2], Petteri Alho[1]

[1]Department of geography and geology, University of Turku, Turku, 20014 Finland
[2]Marine and freshwater solutions, Finnish Environment Institute, Helsinki, Finland

Correspondence to: Iiro Seppä (iielse@utu.fi)

Abstract. Comprehensive, large sample hydrological datasets, such as CAMELS (Catchment Attributes and MEteorology for Large-sample Studies), have provided the basis for advances in many aspects of hydrological research in recent years. They can be utilised for several purposes, such as training data-driven hydrological models, comparisons between regions dominated by different types of hydrological processes and testing of general validity of hydrological theories. The value of these datasets is in combining a multitude of data sources into one, easily accessible and usable, harmonised and high-quality package. We present CAMELS-FI, an extensive hydro-meteorological dataset for 320 catchments in Finland. It combines hydrological and meteorological time series with biophysical and human influence catchment attributes in a format that enables comparisons between catchments within the dataset but also between earlier CAMELS datasets. CAMELS-FI includes a diverse set of catchments with human influence varying from near natural to heavily regulated. CAMELS-FI is available at https://doi.org/10.5281/zenodo.15853357 (Seppä et al., 2025).

1 Introduction

One of the main goals of hydrology as a science is to understand and represent hydrological processes well enough to be able to simulate hydrological processes across the huge natural variations of hydrological settings, and multiple spatiotemporal scales (Gupta et al., 2014). Achieving this requires complementing detailed short-term investigations at a few locations with large-scale and long-term studies that help identifying general trends and patterns. These sorts of research approaches require large-scale runoff datasets, which have only become widely available in the past decade (e.g., do Nascimento et al., 2024; Turner et al., 2025). The general benefits of these datasets are manifold. Collection and proper quality control of large datasets with many variables is very time consuming, and using existing datasets helps reach results faster. Dedicated large sample hydrologic datasets are also typically well documented, citable, versioned and easy to bulk download. They follow the FAIR (findable, accessible, interoperable and reusable) principles (Wilkinson et al., 2016). Together these factors make it easy for researchers to use the data in a way that suits them. This has scientific benefits beyond individual researchers and studies. Since many researchers can use same datasets, commensurability and repeatability of studies is enhanced, and new models can be benchmarked against existing ones without the need to implement and rerun all the previous models.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

However, global or continental hydrologic datasets are often limited in many aspects. Firstly, global hydrology datasets are limited to global weather, soil, geology and land cover datasets, which often have relatively coarse resolution and might fail to take into account locally relevant conditions (Clerc-Schwarzenbach et al., 2024). Secondly, quality and time series length of streamflow measurements vary wildly between different countries (Hasan et al., 2024). Thirdly, the number of stream gauges (subsequently referred to as gauges) in these datasets is often limited compared to the actual number of gauges. For example, the Global Runoff Data Center (GRDC) includes only 111 gauges from Finland, while the total count is over 650 current or historical gauges (Hydrology API, 2024). Some global or continental datasets are also unable to provide the raw streamflow values due to licensing limitations and have to instead provide just hydrologic signatures (e.g., do Nascimento et al., 2024), which can lead to incorrect conclusions in some conditions (McMillan et al., 2023).

Nationwide hydrological datasets aim to be useful for large-scale hydrology while avoiding many of the pitfalls of global datasets. National agencies of many countries produce and openly publish a lot of different high-quality environmental datasets that cover their country of origin, but do not typically extend beyond national borders. Thus, nationwide data allows for high-quality, consistent data and relatively large catchment counts at the same time. One of the most widespread national scale approaches is the CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) datasets. They combine daily streamflow measurements with catchment averaged daily meteorology and more slowly changing "static" attributes, such as topography and soil types. They also provide attributes and metadata in English, which expands the number of researchers that are able to understand and use the data, compared to the underlying datasets which are quite often only accessible in the native language(s) of the particular country. Since the publication of the first CAMELS dataset for the continuous USA (Addor et al., 2017; Newman et al., 2015), CAMELS datasets have expanded into a family of relatively commensurable datasets from Australia (Fowler et al., 2021a), Brazil (Chagas et al., 2020), China (Hao et al., 2021), Chile (Alvarez-Garreton et al., 2018), Denmark (Liu et al., 2025), France (Delaigue et al., 2025), Germany (Loritz et al., 2024), Great Britain (Coxon et al., 2020a), India (Mangukiya et al., 2024), Sweden (Teutschbein, 2024) and Switzerland (Höge et al., 2023). However, each CAMELS is distinct in minor ways in the catchment selection and the variables it provides due to data availability, goals of the creators and differences in the dominant drivers of hydrological processes. For example, CAMELS-US contains three different sets of metorological forcings (Newman et al., 2015), CAMELS-FR provides multitude of morphometric attributes (Delaigue et al., 2025), and CAMELS-CH contains more information on glaciers than other CAMELS (Höge et al., 2023). To improve international comparisons, streamflow data and catchments of some of the CAMELS datasets have also been combined with global meteorology, geophysical and societal datasets to form part of a semi-global dataset called CARAVAN (Kratzert et al., 2023). However, this has resulted to a degradation of catchment attribute quality, highlighting that CAMELS datasets can be optimal for applications where both relatively large sample sizes and high quality data are required (Clerc-Schwarzenbach et al., 2024).

One of the most important outcomes of CAMELS datasets has been a breakthrough in deep learning rainfall-runoff modelling. Multiple studies have confirmed that training a long short-term memory (LSTM) network on large-sample data improves accuracy and robustness of runoff predictions compared to using data limited to one or a few catchments (Gauch et

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

al., 2021; Kratzert et al., 2018, 2019b, 2024). Crucially, LSTM networks trained on CAMELS datasets have been consistently shown to outperform calibrated conceptual and distributed, process-based models in both gauged and ungauged settings (Kraft et al., 2025; Kratzert et al., 2019a). CAMELS datasets have also proved to be potent for other hydrological research tasks. Exploring effects of changes in the environment, such as of droughts or urbanization, to runoff at large scales has been a common application (Fowler et al., 2021a; Han et al., 2022). They have also been used for broad array of other tasks, among them robust calibration and testing of traditional lumped rainfall–runoff models (Bloomfield et al., 2021; Fowler et al., 2021b), analysing relationship of dominant hydrological processes to hydrologic signatures (McMillan et al., 2022) and scrutinising popular performance metrics in hydrology (Clark et al., 2021; Klotz et al., 2024). Although CAMELS are created for large-sample studies, they are also convenient data sources for single-catchment studies, since they cover a large portion of rivers of interest in their country, which saves considerable time from gathering and preprocessing. This is attested by the number of studies that have chosen to select just one or two catchments from a CAMELS dataset (Aguilar et al., 2021; Knoben and Spieler, 2022; Robins et al., 2022).

2 Objectives

Finland has a wide range of high-quality, openly available environmental datasets and long time series of streamflow data. However, these datasets have not been synthesized earlier into one consistent catchment-oriented hydrometeorological dataset that covers most of Finland and is accessible from one location. To resolve this data gap, we present CAMELS-FI, combining daily discharge values from 320 rivers with daily meteorology and various slower changing static attributes.

Our first central objective has been consistency and high quality of data. To ensure this, the same source datasets have been used for all catchments.

Our second central objective has been to create a dataset that is comprehensive, diverse and useful to as many catchment oriented applications in Finland as possible. Thus, we decided to be more inclusive than exclusive when creating the catchment selection criteria (see section 3.1). Despite this, we had to make an exception with catchments where large portions are located outside Finland, as their inclusion would have contradicted our aim for consistency. As a result of inclusivity, CAMELS-FI includes catchments of all regulation levels and time series of varying lengths. This means that CAMELS-FI is applicable for more tasks, but often requires the end user to further create a selection that is suitable to their task.

Additionally, we have strived to ensure compatibility with previous CAMELS to the possible extent, since each CAMELS has something unique. We decided to most closely follow CAMELS-CH and CAMELS-GB for selecting most of the attributes and naming conventions, and CAMELS-SE for soil attributes and lack of hydrolithological data (Coxon et al., 2020; Höge et al., 2023; Teutschbein, 2024). The file structures, file types, file names and directory structure of the data follow the convention of CAMELS-GB in order to facilitate easier integration of CAMELS-FI with existing software that interfaces with CAMELS data.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

3 Catchments

3.1 Catchment selection

The Finnish Environment Institute (SYKE), hydropower companies and regional Centres for Economic Development, Transport and the Environment (ELY centres) maintain hundreds of hydrological gauges with long time series in Finland. The data are shared openly through SYKE's Hydrology API (Hydrologiarajapinta 2025; https://rajapinnat.ymparisto.fi/api/Hydrologiarajapinta/1.0). Clear indicators on the level of regulation are provided in the human influence attributes to aid catchment selection for specific use cases (see section 5.6).

The catchments included in CAMELS-FI were selected according to the following criteria:

1) The gauge must have had observations in 1995 or more recently. This is to ensure that most of the gauges have concurrent observations.
2) A Minimum of five years (1826 days) of observations required. We considered this sufficient to calculate streamflow signatures.
3) Due to aim for consistency, catchments with major areas outside of Finland were excluded due to using national datasets. However, the borders of Finland with Russia and Norway approximately (but not exactly) follow the watershed boundary in many places, leaving relatively small portions of catchments beyond the border. Therefore, we decided to include catchments with less than 5 % of the area beyond the Finnish border (n=27), according to SYKE's catchment delineation data (Valuma-aluejako, 2023). This criterion led to the deletion of some gauges at headwaters in Oulu and Kemi river catchments, while including gauges further downstream.
4) Finnish waterway systems include a considerable number of lake and river bifurcations. Some of these are natural and caused by flat topography or large and labyrinthine lake systems shaped by uneven post-glacial rebound (Kuusisto, 1984; Tikkanen, 2002). Others are manmade for the purposes of flood regulation, reservoir filling or to enable boat or log traffic (Kuusisto, 1984). We decided to exclude most gauges that are located within a bifurcated part of a watercourse, since all of the outflow of the catchment does not go past the gauge. Multiple pour points are also incompatible with the watershed delineation method used in this study (see section 3.2). However, we made exceptions in cases, where:
    a) The effect of bifurcation is minor to the total outflow measured. In these cases, the gauge was left in place, and correct watershed delineation was ensured with methods described in section 3.2 This led to some gauges immediately downstream of a bifurcation being excluded, but gauges further downstream being included. Most notable bifurcation system like this are lakes Lummene and Vesijako shared between Kokemäki and Kymi river catchments. We also decided to include the gauges that are part of the bifurcation system of Lokka and Porttipahta reservoirs, where almost all of the water from Lokka has been directed to Porttipahta and its outlet Kitinen since 1991, outside of the exceptional flood peaks in 1993, 2000 and 2010. Previously, the streamflows were more evenly distributed between the reservoirs. We included the gauge at Kammonen, Luiro (1358) at the

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

middle reach of Luiro as a "virtual gauge" 1352-1358, by removing the discharge from Lokka (gauge 1352) from the streamflow values of the following day of the gauge 1358, and removed the catchment area upstream of dam of Lokka from the catchment of the gauge 1358. The offset of one day was determined from comparing the streamflow peak timing of high discharge events at gauges 1358 and 1352 in 1993, 2000 and 2010. Virtual gauges can be identified by their id, which is in the format 0000-0000 instead of the typical (0)000 in normal gauges.

b) The bifurcating branches are in close proximity, and both had streamflow measurements. In those cases we created a "virtual gauge" by adding the values of the two gauges together and relocated the gauge just upstream or downstream of the bifurcation, depending on the measurement gauge location and watercourse geometry. Artificial structures are often involved in these situations.

In total, the selection criteria led to 320 catchments that cover 73 % of Finland's total land area (Figure 1). Gauges 1190 and 1191 were from exactly the same location, just separated temporally, and therefore we combined them together to form virtual gauge 1190-1191. Following the regional division of Korhonen and Kuusisto (2010), all hydrological regions of Finland are well represented, although the Vuoksi drainage basin in south-eastern Finland has large portions missing due to many internal bifurcations and a large part of the catchment being on the Russian side. Many of the catchments are nested, meaning that the catchment is contained within at least one larger catchment.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

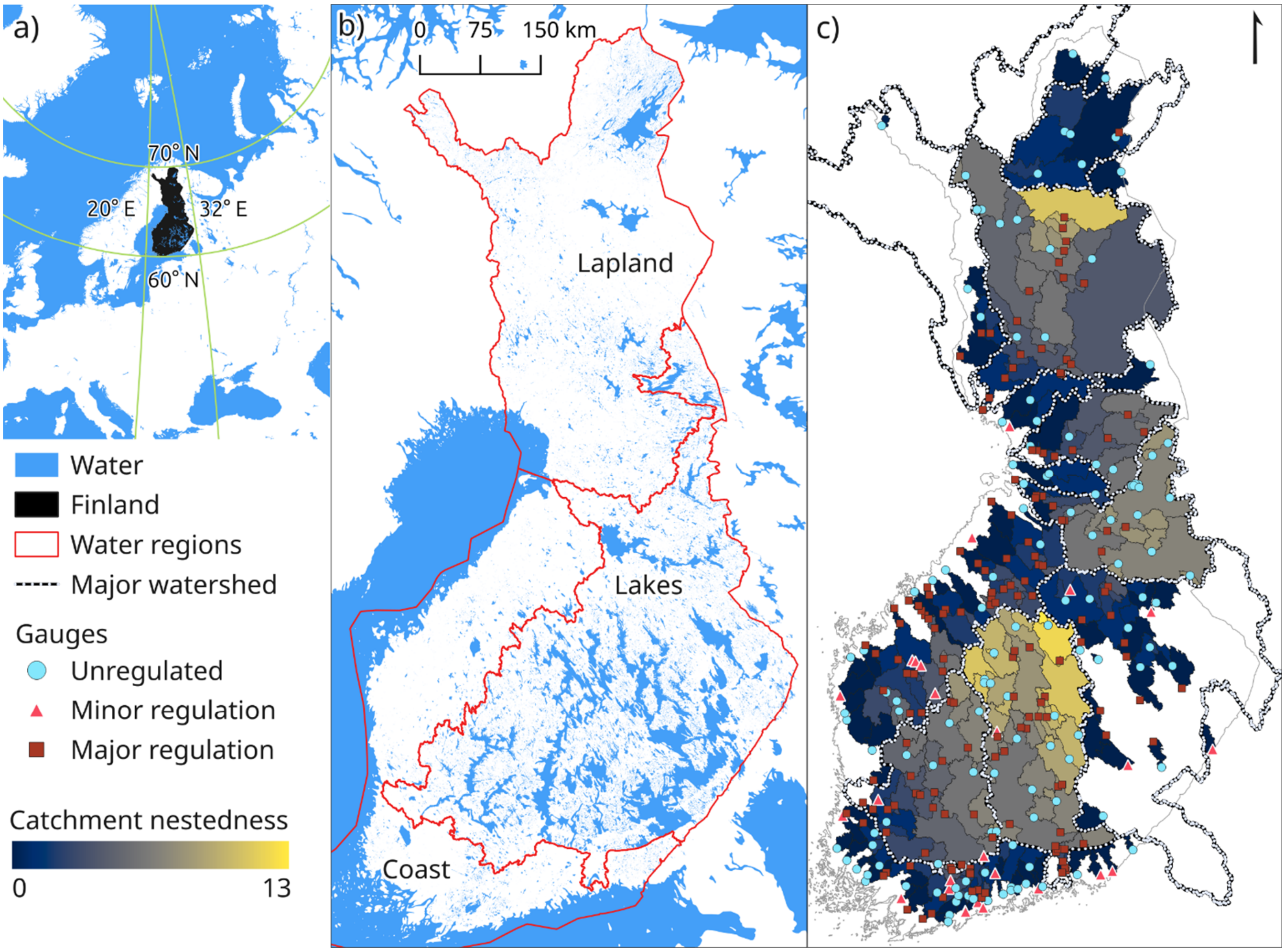
145


Figure 1: a) Finland is located in northern Europe b) The main hydrological regions of Finland according to Korhonen and Kuusisto (2010) c) CAMELS-FI includes catchments that are unregulated, minorly regulated and majorly regulated. The catchments are often heavily nested. Smaller catchments are shown on top of larger ones.

3.2 Catchment boundaries

The catchments for the selected gauges were calculated with the Python bindings of WhiteboxTools (WB tools) (2.3.6), Geopandas (1.0.1) and Rasterio (1.4.3) using the Elevation model 10 m (DEM) of the National land survey of Finland (NLS) (Bossche et al., 2024; Gillies et al., 2024; Lindsay et al., 2014; Elevation model 10 m 2019). To ensure good quality watershed delineation, SYKE's river channel and level 3 watershed division data were used as guiding features (Valuma-aluejako, 2023; Uomaverkosto, 2024). Due to the DEM only covering Finland, channel sections crossing the national border were removed, and in case of some minor bifurcations, the channel network was modified slightly. The known channels were virtually excavated one meter deeper than their surroundings, and the level 3 watershed boundaries were transformed into "walls", except where they intersect with channels. Subsequently, the DEM was then breached and depressions filled

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

with the "breach depression least cost" WB tool. Eight directional flow pointers were computed, streams were identified as places where at least 40 000 cells flow into that cell, and gauges were snapped to those streams. Some gauges had to be moved slightly so that they snapped to the correct location. After that, watersheds were calculated for each gauge. All catchment boundaries were inspected individually by visualizing them alongside other catchments and all the other datasets used in watershed delineation, followed by a visual check for inconsistencies. The inspection always included both an overview of the whole catchment and a detailed, section-by-section examination of the entire border. Where inconsistencies were identified, additional virtual walls were added and the delineation procedure was repeated until a satisfactory result was achieved. Parts of the catchments crossing international borders were added manually by selecting and merging SYKE's level 4 watershed divisions.

3.3 Gauge and catchment metadata

Diverse metadata attributes of the gauge and catchment are presented to support use. Gauge location, gauge name and basin name were downloaded from SYKE's hydrology API. Nestedness, meaning the number of larger catchments that the catchment is part of, the catchment area and the percentage of catchment outside of Finland were calculated for all catchments. The maximum nestedness is 13, and there are 77 separate basins (Figure 1c). Additionally, the hydrologic region of the watershed is also provided (Lapland, Lakes or Coast, after Korhonen and Kuusisto 2010, Figure 1b).

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

Table 1: Summary of gauge and catchment metadata in CAMELS-FI

Attribute name	Description	Unit	Data source
gauge_id	catchment identifier (corresponds to SYKE's discharge station id)	-	SYKE's hydrology API
gauge_name	gauging station name (river or lake name, often followed by more precise location)	-	
owner_id	Identifier of the maintainer of the gauge (SYKE, one of the ELY-centres or other)	-	
owner_name	Name of the maintainer of the gauge	-	
gauge_lat	gauge longitude (EPSG:4326)	°	
gauge_lon	gauge station latitude (EPSG:4326)	°	
gauge_easting	ETRS-TM35FIN coordinates (EPSG:3067), easting	m	
gauge_northing	ETRS-TM35FIN coordinates (EPSG:3067), northing	m	
basin_id	id of the drainage basin assigned by SYKE	-	
basin_name	name of the drainage basin	-	
area	area of the watershed	km^2	Derived by authors from catchments and gauge locations. Water regions based on Korhonen and Kuusisto (2010).
nestedness	how many catchments contain this catchment as subcatchment	-	
water_region_code	water region code	-	
water_region_name	water region name	-	
cross_border_perc	percentage outside of Finland.	%	
reference_gauge	is the gauge among SYKE's reference hydrometric network	yes/no	Turner et al., 2025
ice_correction	have there been manual corrections to the flow values due to ice dams	yes/no	Internal document from SYKE

4 Hydro-climatic time series

4.1 Hydrologic time series

SYKE and ELY centres maintain hundreds of gauges in Finland. Daily discharge ($m^3\ s^{-1}$) for 1961 to 2023 was downloaded from SYKE's hydrology API, and runoff ($mm\ d^{-1}$) was calculated based on catchment areas (Table 2).

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Table 2: Summary of time series variables in CAMELS-FI

Type	Attribute	Description	Unit	Data source
Hydrological time series	discharge_vol	catchment discharge, calculated from water level and channel geometry at gauge	$m^3 s^{-1}$	SYKE's Hydrology API
	discharge_spec	catchment-specific discharge converted to millimetres using catchment area	$mm d^{-1}$	
Meteorological time series	precipitation	catchment daily averaged precipitation	$mm d^{-1}$	Daily gridded climatology for Finland, FMI
	pet	If the snow depth is larger than zero: ERA5 snow evaporation. Otherwise, months 4 to 9: FMI; months 10 to 3: potential evaporation from ERA5-Land	$mm d^{-1}$	Daily gridded climatology for Finland, FMI; ERA5, ERA5-Land
	pe_era5_land	catchment daily averaged potential evaporation, calculated as pan evaporation	$mm d^{-1}$	ERA5-Land
	snow_evaporation	catchment daily averaged evaporation from snow	$mm d^{-1}$	ERA5
	pet_fmi	catchment daily averaged potential evapotranspiration, calculated for well watered 12 cm grass.	$mm d^{-1}$	Daily gridded climatology for Finland, FMI
	temperature_gmin	catchment daily averaged minimum temperature near ground	°C	
	temperature_min	catchment daily averaged minimum temperature at 2m	°C	
	temperature_mean	catchment daily averaged air temperature at 2m	°C	
	temperature_max	catchment daily averaged maximum air temperature at 2m	°C	
	radiation_global	catchment daily averaged global radiation sum	$kJ m^{-2}$	
	humidity_rel	catchment daily averaged relative humidity	%	

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Type	Attribute	Description	Unit	Data source
	snow_depth	catchment daily averaged snow depth at 6:00 UTC	cm	
	swe	catchment daily averaged snow water equivalent, ERA5-Land	mm	ERA5-Land
	swe_cci3-1	catchment daily averaged snow water equivalent, ESA Snow Climate Change Initiative, version 3.1	mm	ESA Snow Climate Change Initiative, version 3.1

The duration of the provided hydrological observation series varies from 5 to 63 years, with a median of 45 years and three-quarters of the gauges have had observations for 30 years or more (Figure 2).

The 17 catchments that are part of SYKE's hydrological reference network and the global ROBIN (Reference Observatory of Basins for International hydrological climate change detection) initiative are indicated in metadata attributes (Turner et al., 2025). Extra effort has gone into ensuring that these gauges have very high-quality observations, and they are largely unaffected by human disturbances.

Supplementary daily streamflow values are also provided for a few locations where artificial bifurcation from one catchment to another occurs. For these gauges, the only static attributes provided are the gauge metadata, and no watershed is provided. The locations are: Päijänne Water Tunnel (1092 & 1093); Paimio River to Aura River (1115); Lokka reservoir to Luiro River (1352) and Saimaa Canal to Rakkola River (3683). Overall, the effects for both source and recipient catchments are usually relatively minor, although the effect might be noticeable during low streamflow. These artificial bifurcations are separated from the rest of the data and are not counted as catchments in CAMELS-FI.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

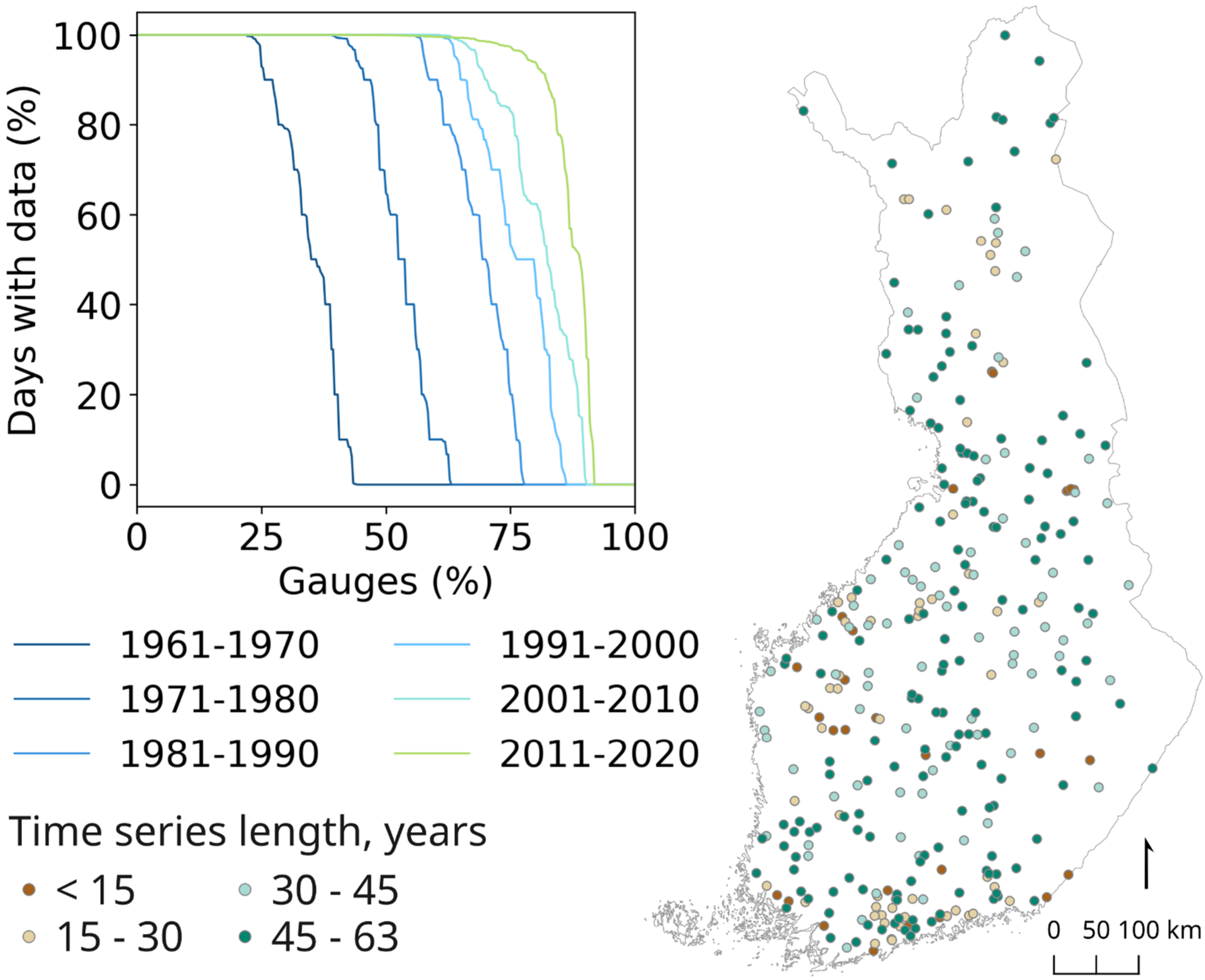


Figure 2: Time series length of hydrologic observations in CAMELS-FI. The number of gauges has increased gradually over the years, but some have also been decommissioned.

4.2 Meteorologic time series

Daily meteorological attributes from 1961 to 2023 were derived for the catchments mainly from the Finnish Meteorological Institute's (FMI) 1 km x 1 km gridded daily climatology of Finland (Aalto et al., 2016). The product has been created by applying Kriging with external predictors of lake and sea cover, elevation and relative altitude to FMI's observation network. It should be noted that minimum temperature near ground is missing from the original dataset from 16 October 2022 onwards, which is why it is also missing from CAMELS-FI. Potential evapotranspiration (PET) was only available for the months from April to September from 1981 onwards from FMI (Pirinen et al., 2022). Thus, we also include snow

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

evaporation from ERA5 and potential evaporation (PE) from ERA5-Land (Hersbach et al., 2020; Muñoz-Sabater et al., 2021). FMI and ERA5-Land use different approaches to estimate PET, as ERA5-Land simulates pan evaporation and FMI approximates well-watered grassland with Penman–Monteith equation. ERA5 would have included a PET estimate that is more similar to FMI's approach, but at the time of downloading the data, ERA5 had an issue that prevented transpiration in forested areas, leading to severe PET underestimates in almost all of Finland (https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Knownissues), last accessed 20 May 2025). In addition to providing the aforementioned data, we combined the data into one convenience PET attribute from 1981, which was also used for calculating climatic signatures (see section 5.2). This was done by using snow evaporation when the snow depth of the catchment was over zero, and filling the snow-free days with FMI PET if it was available; the remainder was completed by ERA5-Land potential evaporation. ERA5-Land PE estimates are consistently larger than the FMI PET, which causes an unrealistic increase in evaporation in October for the combined PET. However, the yearly total values are closer to the actual evaporation (Precipitation - Runoff) for the combined product instead of ERA5-Land.

FMI does not include snow water equivalent (SWE) in their gridded product, only snow depth. However, as snowmelt typically causes most major high streamflow events in Finland, we felt that including it was necessary. The SWE product in ERA5-Land has been benchmarked to perform overall better than other available products, so we chose it (Mudryk et al., 2025; Muñoz-Sabater et al., 2021). However, as there are many SWE products that have been benchmarked relatively similarly in non-mountainous regions, and there are some differences between the products, we decided to also include ESA's CCI3-1, which is available from the beginning of 1979, usually from October to May (Luojus et al., 2024). It is provided typically for every other day and has occasional gaps of two or three days, so gaps of less than three days were interpolated linearly to obtain daily values. See section 7 for a brief comparison between the different snow products.

The catchments' daily values were extracted by calculating the areal average of the grid for each catchment, considering all pixels that touch the catchment to avoid problems with the smallest catchments for coarser resolution products. ERA5 and ERA5-Land offer hourly data, which were aggregated to daily values. The portions of catchments beyond Finnish borders were not included in the catchments used for calculations, as is the case for all static catchment attributes as well.

5 Catchment attributes

5.1 Hydrologic signatures

Hydrologic signatures were calculated from the daily streamflow and precipitation data described in Section 4. Definition of water year for Finland was obtained from Irannezhad et al. (2024), meaning that the water year starts on 1 September and ends 31 August. Hydrologic signatures were calculated with available observations for the water years of the climatological standard period 1991–2020, except for a few gauges with less than five years of observations during the period, for which all

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

available data were used. All those gauges had started measurements after 2015, excluding gauge 1353, which had a gap in observations between 1989 and 2020. Hydrologic signatures were calculated using code from https://github.com/naddor/camels (last accessed 20 March 2025).

Table 3: Static attributes of the catchments in CAMELS-FI

Attribute class	Attribute name	Description	Unit	Data source
Topography	slope	catchment mean slope	%	Elevation model 10 m, NLS.
	elev_gauge	gauge elevation. Not always equal to minimum due to mines.	m a.s.l.	
	elev_min	minimum elevation	m a.s.l.	
	elev_10	10th elevation percentile	m a.s.l.	
	elev_50	median elevation	m a.s.l.	
	elev_90	90th elevation percentile	m a.s.l.	
	elev_max	maximum elevation	m a.s.l.	
	elev_range	gauge elevation subtracted from maximum	m	
Climate signatures	p_mean	Long-term mean daily precipitation	mm d^{-1}	Daily gridded climatology for Finland, FMI
	temperature_mean	Mean annual temperature	°C	
	pet_mean	long-term mean daily PET	mm d^{-1}	Daily gridded climatology for Finland, FMI; ERA5 & ERA5-Land
	aridity	aridity, calculated as the ratio of mean daily potential evapotranspiration to mean daily precipitation	-	
	p_seasonality	seasonality and timing of precipitation (estimated using sine curves to represent the annual temperature and precipitation cycles, positive (negative) values indicate that precipitation peaks in summer (winter), and values close to zero indicate uniform precipitation throughout the year). See Eq. (14) in	-	Daily gridded climatology for Finland, FMI

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Attribute class	Attribute name	Description	Unit	Data source
		Woods (2009)		
	frac_snow	fraction of precipitation falling as snow (T< 0 °C)	-	
	high_prec_freq	frequency of high-precipitation days (≥ 5 times mean daily precipitation)	d y^{-1}	
	high_prec_dur	average duration of high-precipitation events (number of consecutive days ≥ 5 times mean daily precipitation)	d	
	high_prec_timing	Season during which most high precipitation days occur, if two seasons register the same number of events, a value of NaN is given	season	
	low_prec_freq	frequency of dry days (< 1 mm d−1)	d y^{-1}	
	low_prec_dur	Average duration of dry periods (number of consecutive days < 1 mm d−1 mean daily precipitation)	d	
	low_prec_timing	season during which most dry days (< 1 mm d−1) occur. If two seasons register the same number of events, a value of NaN is given	season	
Hydrologic signatures	timeseries_number_of_years	total duration of time series in years, without gaps	y	Hydrology API, SYKE
	sign_start_date	Start date for signature evaluation	date	
	sign_end_date	End date for signature evaluation	date	
	sign_number_of _years	Number of years for signature evaluation, without gaps	y	
	sign_number_of _obs	Total number of observations	-	
	q_mean	Mean daily specific discharge	mm d^{-1}	
	runoff_ratio	Runoff ratio (ratio of mean daily discharge to mean daily precipitation)	-	

Attribute class	Attribute name	Description	Unit	Data source
	stream_elas	Stream flow precipitation elasticity (sensitivity of stream flow to changes during a year)	-	
	slope_fdc	Slope of the flow duration curve (between the log-transformed 33rd and 66th stream flow percentiles) (Yadav et al., 2007). Value is NaN if over a third of the observations are zero.	-	
	baseflow_index_landson	base flow index (ratio of mean daily base flow to daily discharge, using the Ladson et al. (2013) digital filter	-	
	baseflow_index_lfstat	base flow index (ratio of mean daily-base flow to daily discharge, using the lfstat implementation of Tallaksen and Van Lanen (2004)		
	hfd_mean	Mean half-flow date (number of days since 1 September at which the cumulative discharge reaches half of the annual discahrge)	d	
	Q5	5 % flow quantile (low flow)	mm d^{-1}	
	Q95	95 % flow quantile (high flow)	mm d^{-1}	
	high_q_freq	frequency of high-flow days (> 9 times the median daily flow)	d y^{-1}	
	high_q_dur	average duration of high-flow events (number of consecutive days > 9 times the median daily flow)	d	
	low_q_freq	frequency of low-flow days (< 0.2 times the mean daily flow)	d y^{-1}	
	low_q_dur	average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	d	

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

Attribute class	Attribute name	Description	Unit	Data source
	zero_q_freq	fraction of days with Q = 0	-	
Soil	bedrock_perc	Percentage of rocky outcrops of total area	%	Superficial deposits of Finland 1:200 000 and Superficial deposits thickness 1:1000 000, GTK
	coarse_perc	Percentage of coarse grained (glaciofluvial) deposits of total area	%	
	silt_perc	Percentage of silt dominated soils of total area	%	
	till_perc	Percentage of till dominated soil of total area	%	
	clay_perc	Percentage of clay dominated soil of total area	%	
	peat_perc	Percentage of peat dominated soil of total area	%	
	soil_depth	Mean soil depth to bedrock	m	
Land cover	crop_perc	Percentage of agricultural land of total area	%	CORINE Land Cover 2000, 2006, 2012, 2018, SYKE
	grass_perc	Percentage of grassland of total area	%	
	shrub_perc	Percentage of shrubland of total area	%	
	dwood_perc	Percentage of deciduous forests of total area	%	
	ewood_perc	Percentage of evergreen forests of total area	%	
	urban_perc	Percentage of impermeable or poorly permeable human made surfaces and built areas of total area	%	
	inwater_perc	Percentage of water areas of total area	%	
	bares_perc	Percentage of bare land of total area	%	
	wetland_perc	Percentage of wetlands of total area	%	
Human influence	num_inhabitants	population count		Population grid data 2023, Statistics Finland
	dens_inhabitants	population density	km^{-2}	
	num_dam	dam count, includes active and	-	Topographic database,

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Attribute class	Attribute name	Description	Unit	Data source
		passive dams		NLS
	num_reservoir	reservoir count	-	lake API, SYKE
	reservoir_cap	reservoir capacity	1000 m^3	
	num_regulation_other	Count of other water regulation permits than reservoirs	-	
	regulation_level	how strongly the catchment is regulated	-	

5.2 Climatic signatures

Climatic signatures of the catchments were calculated for water years 1991–2020 from daily precipitation, PET, and mean temperature, using code from (https://github.com/naddor/camels, last accessed 20 March 2025). Additionally, mean temperature was calculated for all the catchments for the same period. We chose to use the combined PET that was described in section 4.2 since it gave more realistic yearly mean potential evaporation values than the potential evaporation estimate of ERA5-Land

5.3 Elevation

Topography descriptors were derived from the NLS's DEM, without hydrological preprocessing (Elevation model 10 m, 2019). Please note that gauge elevation may differ from the lowest elevation of the catchment due to mines, which is why both are provided.

5.4 Land cover

Land cover can change significantly on a decadal scale. Thus, instead of providing one value for each land cover class, we follow the example of CAMELS-CH (Höge et al., 2023), and provide values for all four CORINE Land Cover (CLC) products made for Finland, from the years 2000, 2006, 2012 and 2018 (Corine maanpeite 2000, 2002; Corine maanpeite 2006, 2008; Corine maanpeite 2012, 2014; Corine maanpeite 2018, 2020). Linearly interpolated yearly values for 2000–2018 are also provided. Unfortunately, we were unable to find earlier harmonized land cover time series datasets, so our land cover time series covers only the most recent decades. CLCs are produced nationally with harmonized criteria set by the European Environment Agency. We used the Finnish versions of CLCs, which have higher resolution and a smaller minimum feature area than the Europe-wide products. They have been created using existing datasets at SYKE combined with automated satellite image interpretation. We created a similar classification to CAMELS-CH and CAMELS-SE, which also used CLC as their land cover data. Each CLC edition has seen some small changes to the classes present, but in our

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

opinion these are largely inconsequential for hydrological tasks. Exact CLC classes used for each year are presented in Appendix A1.

5.5 Soil

Soil class areal portions were calculated for each class based on the Superficial deposits of Finland 1:200 000 (2018) soil classification map by the Geological Survey of Finland (GTK). It should be noted that the areal percentages of soil classes do not total 100, because water and built areas were excluded. This data is different from the soil information provided by most other CAMELS, since GTK classifies soils based on the dominant grain size, instead of providing weight portions of different grain sizes. However, Sweden has a similar soil classification system, and we decided to follow the soil classes present in CAMELS-SE (Teutschbein, 2024). Descriptions of the exact soil classes used are presented in Appendix A2. Global or Europe-wide soil datasets were considered unsuitable for Finland due to their coarse resolution. In terms of groundwater, the most important landforms in Finland are eskers and terminal moraines (Fin. s*alpausselkä*), which form the most important aquifers in Finland due to their high sand and gravel content (Katko et al., 2006). These landforms have a thin and long shape, meaning that they are poorly represented by coarse resolution datasets. Soil depth was averaged from the Superficial deposit thickness 1:1 000 000 product from GTK (Superficial deposit thickness 1:1 000 000, 2018).

In Finland, quaternary sediment deposits form only a thin or, in some locations, a non-existent layer above the crystalline bedrock, which conducts water very poorly outside of fractures (Karro and Lahermo, 1999). Thus, the geology beneath the soil has only a little influence on the streamflow on the surface. Due to this, no national hydrolithological data exists for Finland, and hydrolithological attributes present in many previous CAMELS cannot be provided. Sweden is similar, and CAMELS-SE does not include any hydrolithological information (Teutschbein, 2024).

5.6 Human influence

CAMELS-FI contains catchments with vastly different levels of human influence, and multiple different attributes are provided to help dataset users decide which catchments suit their needs. Catchments are divided into three regulation classes based on regulation information from SYKE's Lake API (Järvirajapinta, 2025; https://rajapinnat.ymparisto.fi/api/jarvirajapinta/1.0) and manual assessment of streamflow plots. The classes are: No active regulation (n=131), Minor active regulation (n=26) and Major active regulation (n=163). Gauge 3536 changed from majorly to minorly regulated in 2013, and was classified as minorly regulated. Not actively regulated catchments may still have dams, but the dams only dampen natural variations slightly. They also include some catchments with a regulation permit that is either not used or its effects are nearly inconsequential for the discharge at the gauge. Minor active regulation includes catchments which have regulation that influences the discharge, but the end result is still relatively close to natural. Major active regulation includes the rest of the catchments where the discharge is strongly controlled. Reservoir capacity (difference between maximum and minimum) and count, as well as the count of other regulation types, are provided. We

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

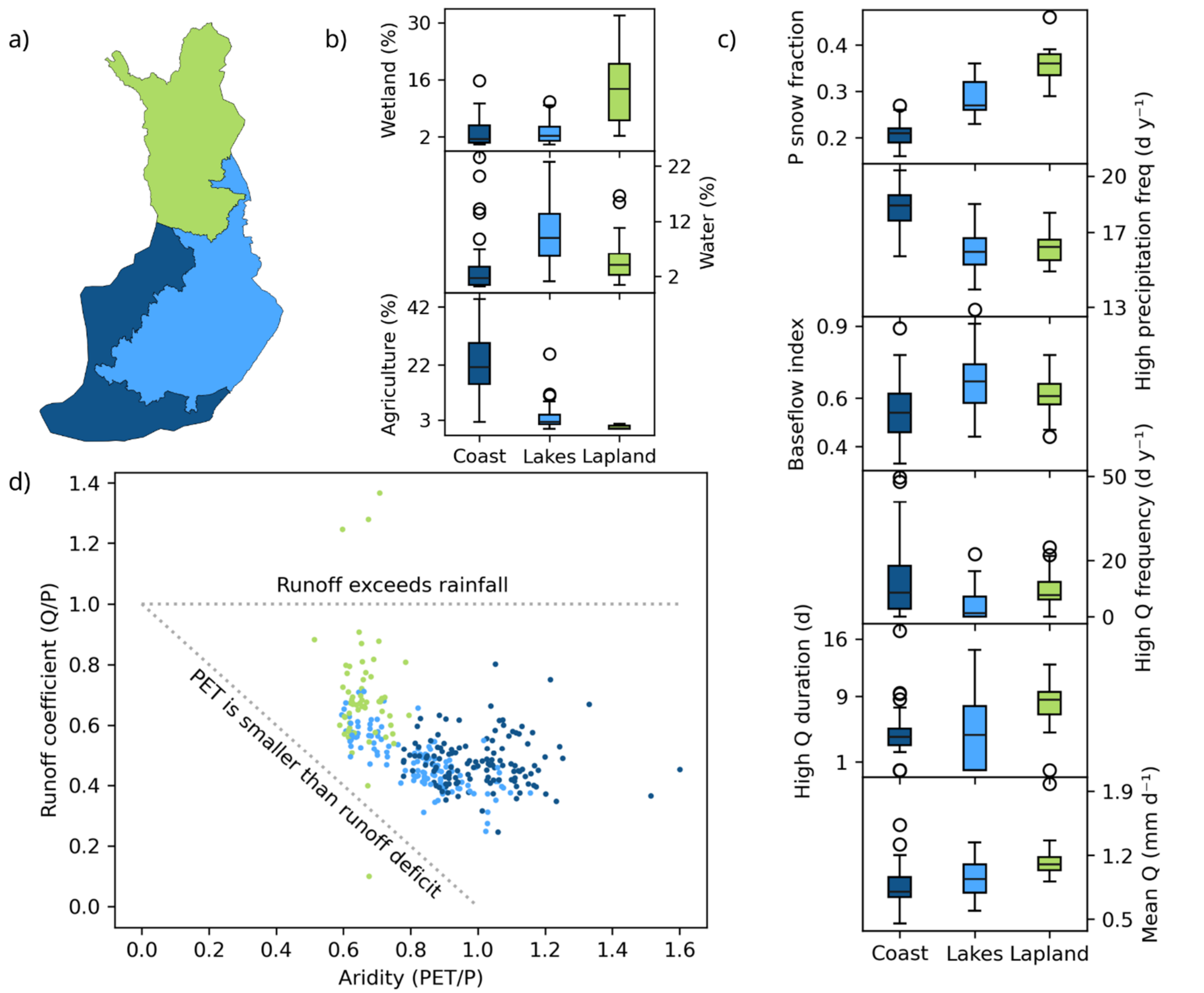
define a lake as a reservoir if the regulation permit has a known capacity; otherwise, it is deemed to be other regulation. Dam count is also provided based on NLS's Topographic Database (National Land Survey of Finland, 2024). It should be noted that the authors noticed some dams missing from the topographic database, so the given counts must be considered as approximate. The population count of each catchment, based on Statistics Finland's 1 km x 1 km grid for 2023, is also included (Statistics Finland, 2024).

6 Regional variability and catchment properties

Hydrologic regions of Finland have differences that cause significant disparities in hydrological behaviour. Rainfall is greatest in the southern coast and high hills of the northern lake region, up to 750 mm per year. The lowest rainfall occurs in northern Lapland, where rainfall can be as low as 470 mm y^{-1}. However, mean temperature and mean PET are the highest in the southwest, and decrease towards the northeast, which results in Lapland and the northern lake region having the highest runoff (Figures 3 & 5). The proportion of snow from rainfall follows a similar pattern.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions



305 **Figure 3: a) Map of the colours used to mark hydrological regions for the rest of the figure b) Variation in land cover between regions c) Variation in climate and flow regime between regions. Baseflow calculated with the methodology of Ladson et al. (2013) d) Gauges plotted by runoff coefficient and aridity. Inspiration for the figure from figure 2 of Coxon et. al. (2020)**

Flood patterns also differ between the regions. Lapland and the northern lakes typically have only one single flood event in the spring caused by snowmelt, while this is true for the coasts only in the coldest winters (Figure 4). During milder winters, winter floods caused by (partial) snowmelt are common on the coasts, while northern parts may receive increased or normal snowfall due to warmer air being able to retain more water. Catchments with a high lake portion have dampened and elongated flood peaks and an overall higher baseflow index. Catchments on the coast are typically smaller or have fewer lakes than the other two regions, causing catchments there to react more rapidly to rainfall or snowmelt events. The coasts also receive, on average, more high precipitation events, which can cause local flooding, especially in cities.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

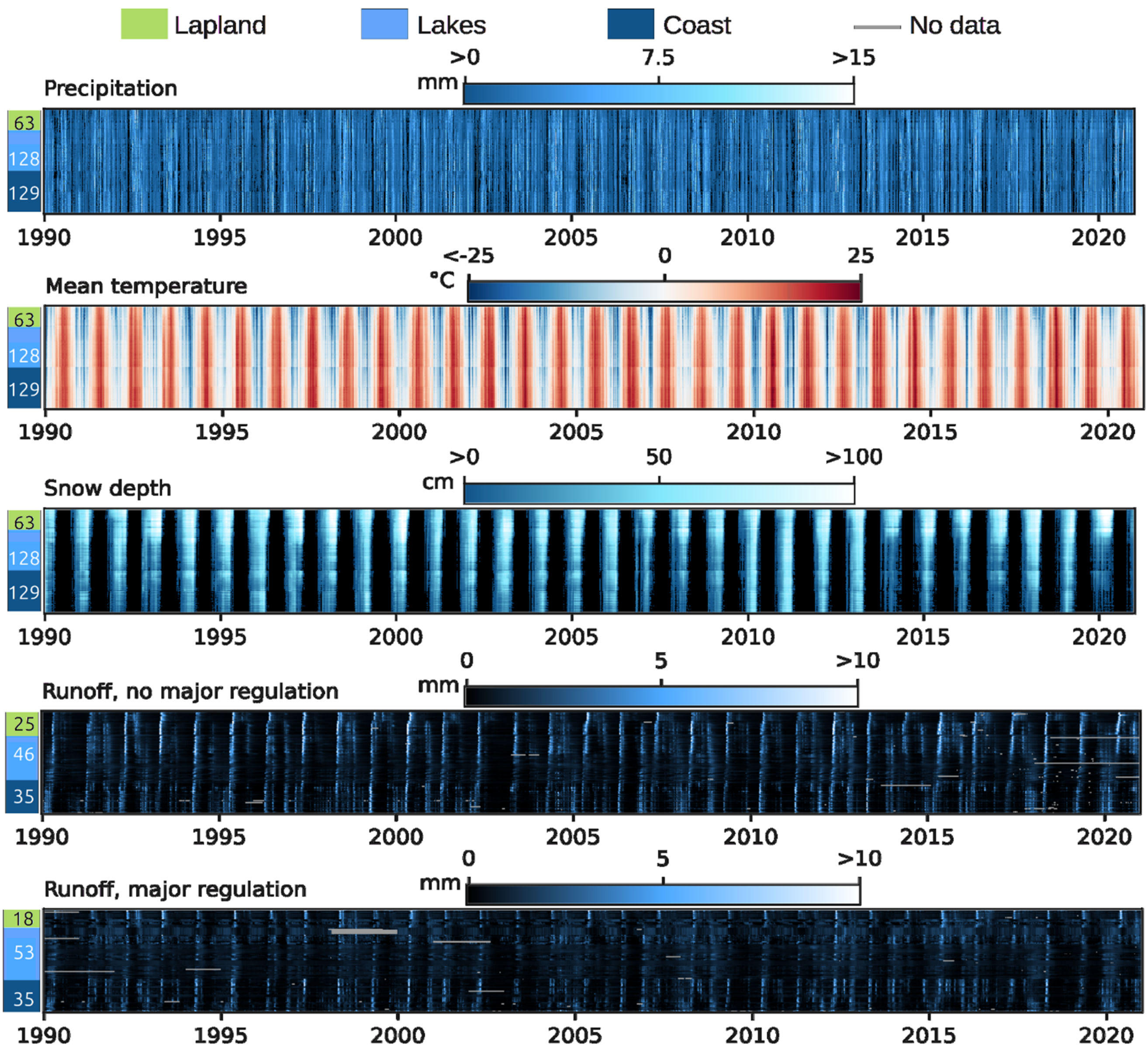


Figure 4: Key hydrometeorological time series for the years 1990–2020. One row of pixels corresponds to the time series of one catchment, and one column of pixels to state of all catchments at the same four day period (resampled from one day values due to resolution limitations of the figure). Catchments have been vertically sorted first into hydrological regions (Lapland, Lakes or Coast) and then by mean temperature, placing colder catchments at the top of each region. The hydrological region bar on the left shows the region of the catchments on its right, as well as number of catchments per region. For visual clarity, only gauges with fewer than 800 missing days during the period were chosen for plotting runoff and catchments with major regulation were plotted separately. The value ranges do not show the full ranges of the variables to place emphasis on typical behavior of catchments instead of extremes. Full ranges are described in Appendix B. Precipitation and snow depth are shown as black when they have been zero to distinguish them from days with values slightly above zero

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Open Access Earth System Science Data Discussions

Evergreen forests are the dominant land cover for most of Finland, except for the northernmost area in Lapland. The largest differences between the main hydrological regions are that the Coast encompasses most of the agricultural areas, Lapland has the largest portion of wetlands and the Lake area contains the highest percentage of lakes, although especially Lapland and some areas of the Coast also have significant lake areas (Figures 3 & 5). Overall, changes to the land cover of Finland since the start of the millennium have been minor. The most notable short-term transformations comprise the areally fragmented cycle of forest logging and subsequent regeneration.

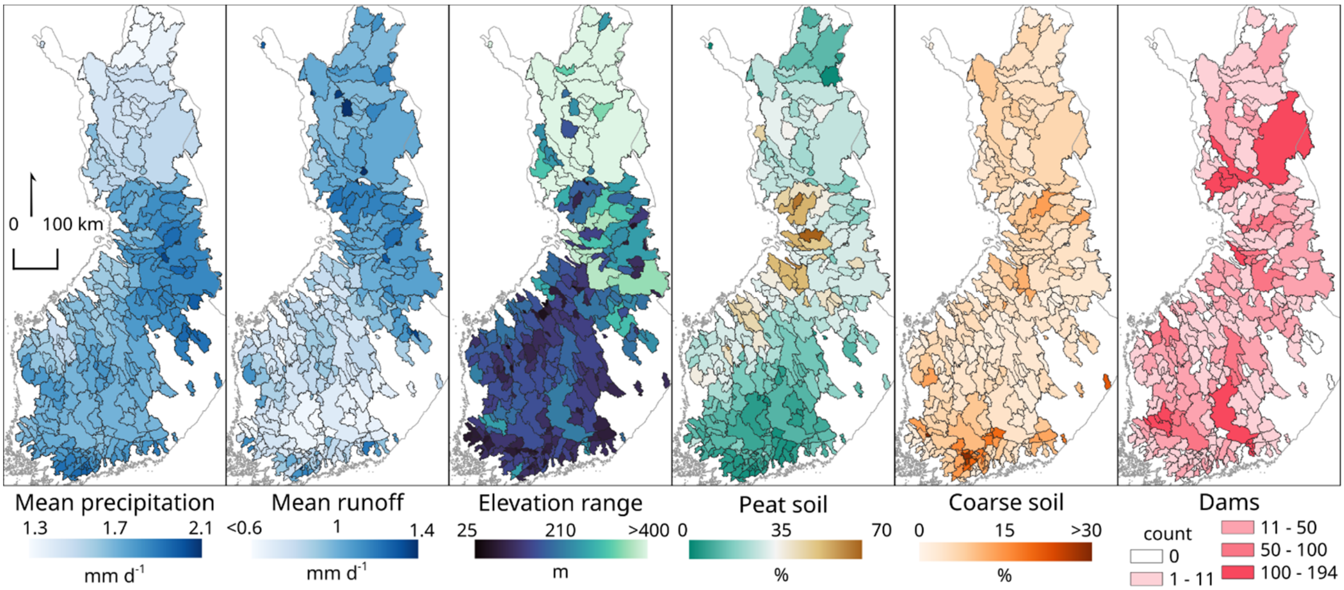


Figure 5: Multiple attributes exhibit great spatial variation over Finland. It should be noted that color scale for daily mean runoff has been limited to 0.6 mm on the low end due to anomalously low average runoff of gauge 3727 (0.19 mm d^{-1}). Similarly, shown elevation range was capped at 400 m (max. 804 m) to keep variation outside of Lapland visible and coarse soil percentage was capped at 30 % (max. 48 %) due to a few small outlier catchments covered by major glaciofluvial deposits.

Eastern Finland and Lapland have considerably higher hills than the rest of Finland. Ostrobothnia has quite flat topography, which is accompanied by a large portion of the top layer of soil being peat, even in areas that support forests. The rest of Finland has varied topography with smaller, but relatively steep hills. Sandy and gravelly soils that are associated with terminal moraines and eskers are dispersed throughout the country but are especially concentrated on the southern border of the Coast and Lakes regions, which is also the reason for the general location of watersheds there. Catchments with the least human influence are limited in size and mostly located near basin watersheds.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

7 Uncertainty and consistency of hydrometeorological observations

The quality of streamflow observations varies both in time and between the gauges. Streamflow measurements from (hydroelectric) dams have generally had high quality from the 1960s. For other locations, the streamflow values are based on rating curves, which started to be updated regularly only in the 1990s for gauges maintained by SYKE. This may cause inaccuracies in the earlier decades due to changes in the morphology of the river channel. Discharge measurements that form the basis for the rating curve also improved substantially around the same time, especially in large rivers due to progress in discharge measurement technology. Measurements were automated from 1984 to the early 2000s. Before the 1980s, the value of the day was read manually at 08:00 in the morning. Automation enabled a shift to using daily mean values instead. There is significant variation between different ELY centres in quality. Some are very close to SYKE, others do not update their rating curves as regularly.

The wintertime raw streamflow values are influenced by ice cover almost every winter for all of the gauges that are not part of a dam. The effects of ice are especially significant during ice jams. The wintertime streamflow values are corrected afterwards by an expert at the SYKE for 61 catchments in CAMELS-FI. Some of the gauges only receive occasional corrections for ice jam occurrences. Only one person has been appointed to handle the task at any time. When the person changes, the new one is trained for the task, but the task is so manual that each person leaves their mark on the corrections. A computer-aided system has been in use since 1997, and the current system since 2008; previously the correction was done using a semi-logarithmic graph. Gauge metadata includes information of whether ice corrections are made for the gauge.

A simple comparison on the agreement of Precipitation, PET and streamflow as well as different snow products is presented here to probe the general agreement, or lack thereof, of different data sources. PE or PET is on average larger in the boreal zone than actual evapotranspiration (AET) (Peltoniemi et al., 2015). This prevents comparisons that would directly calculate water balance using different subsets of precipitation, streamflow and AET, since CAMELS-FI does not include AET estimates. However, some general agreement indicators can still be used. For example, over long timescales precipitation should always exceed runoff and difference of them should not exceed PET, assuming no water flows through catchment borders and water storages remain unchanged. These “bounded water balance” indicators were calculated for all catchments in CAMELS-FI using p_mean, pet_mean and q_mean meteorologic and hydrologic indices (table 3). The meteorological and streamflow data are within expected bounds for most of the catchments in CAMELS-FI (Figure 4d). However, runoff exceeds precipitation for gauges 1369, 3335 and 3728, and runoff is anomalously low for 3727. The gauge 3335 is probably explained by the mountainous terrain of the catchment, since it is a known issue that rain gauges underestimate precipitation during snowfall in windy conditions (Saltikoff et al., 2015). The other three catchments are heavily regulated and have relatively short observation periods within the reference climatological period 1991–2020, and thus changes to water storages are likely causes. The reservoir lake of 3727 is also directly upstream of the reservoir of 3728, so they are coupled. Following our goal of inclusivity, we decided to not remove them from the data, as they can still be useful for many tasks.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

The FMI's snow depth and two SWE products, ERA5-Land and CCI3-1, were compared by counting the average number of snow days during water years 1991–2020. A day was considered a snow day if snow depth exceeded 1 cm, or SWE was over 1 mm. ERA5-land estimated on average 189.4 snow days, over three weeks more than CCI3-1 (161.9 d y^{-1}) or FMI's snow depth (163.6 d y^{-1}). This pattern holds generally in all parts of Finland. The similar averages of FMI's snow depth and CCI3-1 are probably explained by the fact that CCI3-1 combines ground station snow depth observations with microwave satellite observations (Takala et al., 2011). It should be noted that on the catchment level there is substantial variation between all products, which is probably due in part to the different grid sizes of the products and in part by methodological differences.

The values of the two SWE products were also compared by subtracting ERA5-Land from CCI3-1 (Figure 6). ERA5-Land provides SWE values that are on average 11.3 mm higher than CCI3-1. However, the difference has exceeded 100 mm every year in the snowier parts of Finland. This is most likely caused by CCI3-1's known issue with severely underestimating SWE values higher than 150 mm due to limitations of microwave satellite measurements (Takala et al., 2011).

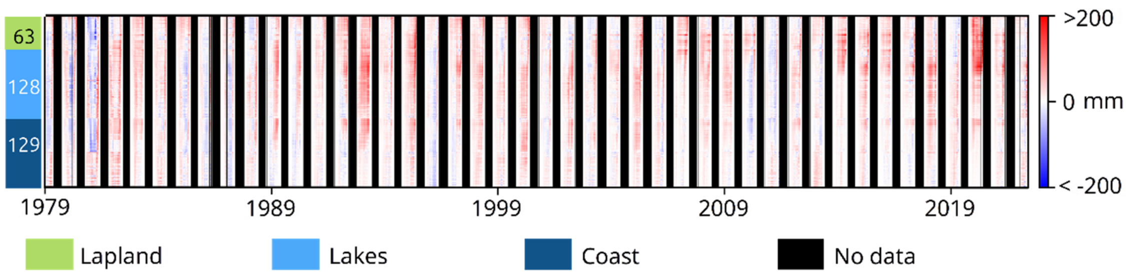


Figure 6: Difference between SWE estimates of ERA5-Land and CCI3-1. Positive (negative) values indicate that ERA5-Land is larger (smaller) than CCI3-1. Each row of pixels corresponds to one catchment and one column of pixels to state of all catchments at the same four day period (resampled from one day values due to resolution limitations of the figure). Gauges have been sorted vertically within regions by mean temperature.

8 Data availability

CAMELS-FI can be accessed at Zenodo under https://doi.org/10.5281/zenodo.15853357 (Seppä et al., 2025). All the code produced for creating CAMELS-FI is available at the dataset's code repository at https://github.com/iiroseppa/CAMELS-FI. Please note that the process involved some manual work, which means that the data are not fully reproducible by code alone.

9 Conclusions

This study introduces CAMELS-FI, a large-sample, open-source catchment dataset for Finland, encompassing 320 catchments. CAMELS-FI provides a consistent, high quality, comprehensive and easily usable dataset of daily hydro-meteorological time series and attributes on topography, land cover, anthropogenic influence, climate, hydrology and soil for a diverse set of catchments. The dataset follows the format of other CAMELS datasets, especially those from Great Britain, Switzerland and Sweden, as closely as possible, facilitating cross-dataset comparisons and software integration.

The dataset's extensive spatiotemporal coverage and detailed attributes advance cold region hydrological science by:

I) Facilitating robust large-scale studies on the impacts of snow, ice and lakes on water cycles, and the hydrological responses to climate change in northern latitudes.

II) Enabling the training of state-of-the-art rainfall-runoff deep learning models, which could lead to e.g. improved flood mitigation, enhanced water resource management and more efficient hydropower production.

III) Conserving time of researchers from complex and computationally expensive preprocessing of data.

IV) Offering a common dataset for benchmarking performance of different models in snowy and lake-dominated catchments.

While CAMELS-FI offers a wealth of data, there are opportunities for future expansion and enhancement. These include adding new types of information, for instance morphometrics or water quality attributes, such as phosphorus, nitrogen or carbon loads. We believe the simple structure of the dataset makes adding new attributes straightforward, and we also encourage possible expansions to be shared openly. Another possible development opportunity would be the creation of "CAMELS-Nordic", since Sweden, Norway and Finland share many cross-border datasets. This would also allow including cross-catchment borders that were excluded from CAMELS-FI.

10 Author contributions

IS conceived CAMELS-FI, gathered and processed data, and was responsible for most of the writing and all the figures. IS, CGI and PA designed the article. JU provided expertise on hydrological observations, provided SYKE's internal metadata about gauges, and helped especially with writing sections 4.1 and 7. CGI, JU and PA suggested improvements to the article. CGI and PA supervised the work.

11 Competing interests

The authors declare that they have no conflict of interest.

12 Acknowledgements

This research was a part of the Ministry of Education and Culture's Doctoral Education Pilot under Decision No. VN/3137/2024-OKM-6 (Digital Waters (DIWA) Doctoral Education Pilot related to the DIWA Flagship (decision no.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

359247) funded by the Research Council of Finland's Flagship Programme) and with Flagship Programme funding granted by the Research Council of Finland for Digital Waters Flagship (decision no. 359247).

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

Appendix A: Class generalizations

Table A1: CLC lvl three classes used for each land cover class in CAMELS-FI. Please note that some of the classes are not present in all editions.

	Crops	Grass	Shrub	Dwood	Ewood	Urban	Inwater	Bares	Wetland
CLC codes	211	141	324	311	312	111	511	131	411
	222	142		311	312	112	512	132	412
	243	231			312	121		331	
		244			313	123		332	
		321			313	124			
		322			313	133			
						142			

Table A2: GTK's soil classes that were combined for CAMELS-FI

	Bedrock	Coarse	Silt	Till	Clay	Peat
Original class	bedrock outcrop	coarse grained	fine grained	till	clay	peat < 0.3 m
	bedrock with less than 1 m of till on top				gyttja	peat > 0.3 m, < 0.6 m
	fragmented rock				gyttja and fine particles	peat > 0.6 m

Appendix B: Time series value ranges

Table B1: Value ranges of time series variables

Attribute	Min	Max	Unit
discharge_vol	0	4824	$m^3 s^{-1}$
discharge_spec	0	62.18	$mm d^{-1}$

precipitation	0	97.8	mm d^{-1}
pet	-1.23	45.06	mm d^{-1}
pe_era5_land	-1.16	63.29	mm d^{-1}
snow_evaporation	-1.23	2.19	mm d^{-1}
pet_fmi	0.1	7.3	mm d^{-1}
temperature_gmin	-53	27.2	°C
temperature_min	-50.1	22.3	°C
temperature_mean	-48	28.5	°C
temperature_max	-45.3	35.9	°C
radiation_global	0	35155.6	kJ m^{-2}
humidity_rel	24.7	100	%
snow_depth	0	184.4	cm
swe	0	469,38	mm
swe_cci3-1	0	457,36	mm

References

Aalto, J., Pirinen, P., and Jylhä, K.: New gridded daily climatology of Finland: Permutation-based uncertainty estimates and temporal trends in climate, Journal of Geophysical Research: Atmospheres, 121, 3807–3823, https://doi.org/10.1002/2015JD024651, 2016.

Addor, N.: camels [code] https://github.com/naddor/camels, 2020, last access, 18 September 2025.

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

Aguilar, G., Valdés, A., Cabré, A., and Galdames, F.: Flash floods controlling Cu, Pb, As and Hg variations in fluvial sediments of a river impacted by metal mining in the Atacama Desert, Journal of South American Earth Sciences, 109, 103290, https://doi.org/10.1016/j.jsames.2021.103290, 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, Hydrology and Earth System Sciences, 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Bloomfield, J. P., Gong, M., Marchant, B. P., Coxon, G., and Addor, N.: How is Baseflow Index (BFI) impacted by water resource management practices?, Hydrology and Earth System Sciences, 25, 5355–5379, https://doi.org/10.5194/hess-25-5355-2021, 2021.

Bossche, J. V. den, Jordahl, K., Fleischmann, M., Richards, M., McBride, J., Wasserman, J., Badaracco, A. G., Snow, A. D., Ward, B., Tratner, J., Gerard, J., Perry, M., cjqf, Hjelle, G. A., Taves, M., ter Hoeven, E., Cochran, M., Bell, R., rraymondgh, Bartos, M., Roggemans, P., Culbertson, L., Caria, G., Tan, N. Y., Eubank, N., sangarshanan, Flavin, J., Rey, S., and Gardiner, S.: geopandas: v1.01 [code], https://github.com/geopandas/geopandas 2024.

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 System Science Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, 2020.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The Abuse of Popular Performance Metrics in Hydrologic Modeling, Water Resources Research, 57, e2020WR029001, https://doi.org/10.1029/2020WR029001, 2021.

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?, Hydrology and Earth System Sciences, 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, 2024.

Corine maanpeite 2000: [dataset] https://ckan.ymparisto.fi/dataset/corine-maanpeite-2000, 2002.

Corine maanpeite 2006: [dataset] https://ckan.ymparisto.fi/dataset/corine-maanpeite-2006, 2008.

Corine maanpeite 2012: [dataset] https://ckan.ymparisto.fi/en/dataset/corine-maanpeite-2012, 2014.

Corine maanpeite 2018: [dataset] https://ckan.ymparisto.fi/en/dataset/corine-maanpeite-2018, 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 System Science Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

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 System Science Data, 17, 1461–1479, https://doi.org/10.5194/essd-17-1461-2025, 2025.

Elevation model 10 m, [dataset], https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/elevation-model-10-m, 2019.

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 System Science Data, 13, 3847–3867, https://doi.org/10.5194/essd-13-3847-2021, 2021a.

Fowler, K. J. A., Coxon, G., Freer, J. E., Knoben, W. J. M., Peel, M. C., Wagener, T., Western, A. W., Woods, R. A., and Zhang, L.: Towards more realistic runoff projections by removing limits on simulated soil moisture deficit, Journal of Hydrology, 600, 126505, https://doi.org/10.1016/j.jhydrol.2021.126505, 2021b.

Gauch, M., Mai, J., and Lin, J.: The proper care and feeding of CAMELS: How limited training data affects streamflow prediction, Environmental Modelling & Software, 135, 104926, https://doi.org/10.1016/j.envsoft.2020.104926, 2021.

Gillies, S. and others: Rasterio: geospatial raster I/O for Python programmers, v 1.4.3. [code], https://github.com/rasterio/rasterio, 2024.

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, Hydrology and Earth System Sciences, 18, 463–477, https://doi.org/10.5194/hess-18-463-2014, 2014.

Han, S., Slater, L., Wilby, R. L., and Faulkner, D.: Contribution of urbanisation to non-stationary river flow in the UK, Journal of Hydrology, 613, 128417, https://doi.org/10.1016/j.jhydrol.2022.128417, 2022.

Hao, Z., Jin, J., Xia, R., Tian, S., Yang, W., Liu, Q., Zhu, M., Ma, T., Jing, C., and Zhang, Y.: CCAM: China Catchment Attributes and Meteorology dataset, Earth System Science Data, 13, 5591–5616, https://doi.org/10.5194/essd-13-5591-2021, 2021.

Hasan, F., Medley, P., Drake, J., and Chen, G.: Advancing Hydrology through Machine Learning: Insights, Challenges, and Future Directions Using the CAMELS, Caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE Datasets, Water, 16, 1904, https://doi.org/10.3390/w16131904, 2024.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hydrologiarajapinta, https://rajapinnat.ymparisto.fi/api/Hydrologiarajapinta/1.0, last access 15 July 2025.

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 System Science Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Irannezhad, M., Abdulghafour, Z., and Sadeqi, A.: Climate Teleconnections Influencing Historical Variations, Trends, and Shifts in Snow Cover Days in Finland, Earth Syst Environ, 8, 1601–1613, https://doi.org/10.1007/s41748-024-00466-1, 2024.

Järvirajapinta, https://rajapinnat.ymparisto.fi/api/jarvirajapinta/1.0, last access 23 June 2025.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Karro, E. and Lahermo, P.: Occurence and chemical characteristics of groundwater, Geological Survey of Finland, Special Paper, 27, 85–96, 1999.

Katko, T. S., Lipponen, M. A., and Rönkä, E. K. T.: Groundwater use and policy in community water supply in Finland, Hydrogeol J, 14, 69–78, https://doi.org/10.1007/s10040-004-0351-3, 2006.

Klotz, D., Gauch, M., Kratzert, F., Nearing, G., and Zscheischler, J.: Technical Note: The divide and measure nonconformity – how metrics can mislead when we evaluate on different data partitions, Hydrology and Earth System Sciences, 28, 3665–3673, https://doi.org/10.5194/hess-28-3665-2024, 2024.

Knoben, W. J. M. and Spieler, D.: Teaching hydrological modelling: illustrating model structure uncertainty with a ready-to-use computational exercise, Hydrology and Earth System Sciences, 26, 3299–3314, https://doi.org/10.5194/hess-26-3299-2022, 2022.

Korhonen, J. and Kuusisto, E.: Long-term changes in the discharge regime in Finland, Hydrology Research, 41, 253–268, https://doi.org/10.2166/nh.2010.112, 2010.

Kraft, B., Schirmer, M., Aeberhard, W. H., Zappa, M., Seneviratne, S. I., and Gudmundsson, L.: CH-RUN: a deep-learning-based spatially contiguous runoff reconstruction for Switzerland, Hydrology and Earth System Sciences, 29, 1061–1082, https://doi.org/10.5194/hess-29-1061-2025, 2025.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resources Research, 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019a.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019b.

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.

Kratzert, F., Gauch, M., Klotz, D., and Nearing, G.: HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin, Hydrology and Earth System Sciences, 28, 4187–4201, https://doi.org/10.5194/hess-28-4187-2024, 2024.

Kuusisto, E.: Suomen vesistöjen bifurkaatiot. [The bifurcations of Finnish watercourses]., Terra, 96, 253–261, 1984.

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.

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Lindsay, J. B.: The Whitebox Geospatial Analysis Tools project and open-access GIS, in: Proceedings of the GIS Research UK, 22nd Annual Conference, The University of Glasgow, https://www.gla.ac.uk/media/Media_401757_smxx.pdf, 2014

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

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

Luojus, K., Venäläinen, P., Moisander, M., Pulliainen, J., Takala, M., Lemmetyinen, J., Derksen, C., Mortimer, C., Mudryk, L., Scwaizer, G., and Nagler, T.: ESA Snow Climate Change Initiative (Snow_cci): Snow Water Equivalent (SWE) level 3C daily global climate research data package (CRDP) (1979 - 2022), version 3.1. [dataset], https://dx.doi.org/10.5285/9d9bfc488ec54b1297eca2c9662f9c81, 2024.

Mangukiya, N. K., Kumar, K. B., Dey, P., Sharma, S., Bejagam, V., Mujumdar, P. P., and Sharma, A.: CAMELS-INDIA: hydrometeorological time series and catchment attributes for 472 catchments in Peninsular India, Earth System Science Data Discussions, 1–43, https://doi.org/10.5194/essd-2024-379, 2024.

McMillan, H., Coxon, G., Araki, R., Salwey, S., Kelleher, C., Zheng, Y., Knoben, W., Gnann, S., Seibert, J., and Bolotin, L.: When good signatures go bad: Applying hydrologic signatures in large sample studies, Hydrological Processes, 37, e14987, https://doi.org/10.1002/hyp.14987, 2023.

McMillan, H. K., Gnann, S. J., and Araki, R.: Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes, Water Resources Research, 58, e2021WR031751, https://doi.org/10.1029/2021WR031751, 2022.

Mudryk, L., Mortimer, C., Derksen, C., Elias Chereque, A., and Kushner, P.: Benchmarking of snow water equivalent (SWE) products based on outcomes of the SnowPEx+ Intercomparison Project, The Cryosphere, 19, 201–218, https://doi.org/10.5194/tc-19-201-2025, 2025.

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 System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

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.

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

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Peltoniemi, M., Pulkkinen, M., Aurela, M., Pumpanen, J., Kolari, P., and Mäkelä, A.: A semi-empirical model of boreal-forest gross primary production, evapotranspiration, and soil water — calibration and sensitivity analysis, Boreal Env. Res., 20, 151–171, 2015.

Pirinen, P., Lehtonen, I., Heikkinen, R. K., Aapala, K., and Aalto, J.: Daily gridded evapotranspiration data for Finland for 1981–2020, FMI's Clim. Bull. Res. Lett., 4, 35–37, https://doi.org/10.35614/ISSN-2341-6408-IK-2022-11-RL, 2022.

Robins, P. E., Dickson, N., Kevill, J. L., Malham, S. K., Singer, A. C., Quilliam, R. S., and Jones, D. L.: Predicting the dispersal of SARS-CoV-2 RNA from the wastewater treatment plant to the coast, Heliyon, 8, https://doi.org/10.1016/j.heliyon.2022.e10547, 2022.

Saltikoff, E., Lopez, P., Taskinen, A., and Pulkkinen, S.: Comparison of quantitative snowfall estimates from weather radar, rain gauges and a numerical weather prediction model, Boreal Environment Research, 20, 667–678, 2015.

Seppä, I., Gonzales Inca, C. A., Uusikivi, J., and Alho, P.: CAMELS-FI (1.0.0), https://doi.org/10.5281/zenodo.15853357, 2025.

Population grid data 1km x 1km 2023, 2024. https://stat.fi/tup/avoin-data/paikkatietoaineistot/vaestoruutuaineisto_1km_en.html

Superficial deposit thickness 1:1 000 000 [dataset], https://tupa.gtk.fi/paikkatieto/meta/maapeitepaksuus_1000k.html, 2018

Superficial deposits of Finland 1:200 000: [dataset] https://tupa.gtk.fi/paikkatieto/meta/maapera_200k.html, 2018.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, https://doi.org/10.1016/j.rse.2011.08.014, 2011.

Tallaksen, L. M. and Van Lanen, H. A. J.: Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater, Developments in Water Science, 48, 2004.

Teutschbein, C.: CAMELS-SE: Long-term hydroclimatic observations (1961–2020) across 50 catchments in Sweden as a resource for modelling, education, and collaboration, Geoscience Data Journal, 11, 655–668, https://doi.org/10.1002/gdj3.239, 2024.

Tikkanen, M.: Long-term changes in lake and river systems in Finland, Fennia, 180, 31–42, 2002.

Topographic database, [dataset], https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/topographic-database, 2024.

Turner, S., Hannaford, J., Barker, L. J., Suman, G., Killeen, A., Armitage, R., Chan, W., Davies, H., Griffin, A., Kumar, A., Dixon, H., Albuquerque, M. T. D., Almeida Ribeiro, N., Alvarez-Garreton, C., Amoussou, E., Arheimer, B., Asano, Y., Berezowski, T., Bodian, A., Boutaghane, H., Capell, R., Dakhaoui, H., Daňhelka, J., Do, H. X., Ekkawatpanit, C., El Khalki, E. M., Fleig, A. K., Fonseca, R., Giraldo-Osorio, J. D., Goula, A. B. T., Hanel, M., Horton, S., Kan, C., Kingston, D. G.,

https://doi.org/10.5194/essd-2025-578
Preprint. Discussion started: 16 December 2025

Laaha, G., Laugesen, R., Lopes, W., Mager, S., Rachdane, M., Markonis, Y., Medeiro, L., Midgley, G., Murphy, C., O'Connor, P., Pedersen, A. I., Pham, H. T., Piniewski, M., Renard, B., Saidi, M. E., Schmocker-Fackel, P., Stahl, K., Thyer, M., Toucher, M., Tramblay, Y., Uusikivi, J., Venegas-Cordero, N., Visessri, S., Watson, A., Westra, S., and Whitfield, P. H.: ROBIN: Reference observatory of basins for international hydrological climate change detection, Sci Data, 12, 654, https://doi.org/10.1038/s41597-025-04907-y, 2025.

Uomaverkosto: [dataset] https://ckan.ymparisto.fi/dataset/uomaverkosto, 2024.

Valuma-aluejako: [dataset] https://ckan.ymparisto.fi/dataset/valuma-aluejako, 2023.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756–1774, https://doi.org/10.1016/j.advwatres.2007.01.005, 2007.