Earth System
Science
Data

Open Access

Discussions

CAMELS-DK: Hydrometeorological Time Series and Landscape Attributes for 3330 Catchments in Denmark

Jun Liu[1], Julian Koch[1], Simon Stisen[1], Lars Troldborg[1], Anker Lajer Højberg[1], Hans Thodsen[2], Mark F. T. Hansen[1], Raphael J. M. Schneider[1]

[1]Department of hydrology, Geological Survey of Denmark and Greenland, Copenhagen, Denmark
[2]Department of Bioscience, Aarhus University, Silkeborg, Denmark

Correspondence to: Jun Liu (juliu@geus.dk)

Abstract

Large samples of hydrometeorological time series and catchment attributes are critical for improving the understanding of complex hydrological processes, hydrological model development and performance benchmarking. CAMELS (Catchment Attributes and Meteorological time series for Large Samples) datasets have been developed in several countries and regions around the world, providing valuable data sources and testbeds for hydrological analysis and new frontiers in data-driven hydrological modelling. Regarding the lack of samples from low-land, groundwater-dominated, small-sized catchments, we develop an extensive repository of a CAMELS-style dataset for Denmark (CAMELS-DK). This CAMELS addition is the first containing both, gauged and ungauged catchments as well as detailed groundwater information. The dataset provides dynamic and static variables for 3330 catchments from various hydrogeological datasets, meteorological observations, and a well-established national-scale hydrological model. The dataset is enhanced with streamflow observations in 304 of those catchments. The spatially dense and full spatial coverage, supplying data for 3330 catchments, instead of only gauged catchments, together with the addition of simulation data from a distributed, process-based model enhance the applicability of such CAMELS data. This is especially relevant for the development of data-driven and hybrid physical informed modelling frameworks. We also provide quantities related to human impact on the hydrological system in Denmark, such as groundwater abstraction and irrigation. The CAMELS-DK dataset is freely available at https://doi.org/10.22008/FK2/AZXSYP (Koch et al., 2024).

1 Introduction

Hydrometeorological time series and catchment attributes are crucial for understanding and modelling hydrological systems (McMillan et al., 2018; Andersson et al., 2015). Long-term records of water cycle variables, such as precipitation, evapotranspiration (ET), streamflow, and groundwater levels, give profound insights into dynamics and trends of water movement. These records form the basis of supporting water resource management and climate change adaptation. Sufficient

hydrological data are required by scientists and engineers to manage water resources effectively and to make accurate predictions of hydrological extremes (Van Loon, 2015). Catchment attributes provide information on the physical characteristics of catchments, such as topography, soil type, land use, and geology, which are important for understanding how catchments respond to meteorological events and are relevant for land-use planning and catchment management. The similarity

35 of catchment attributes is useful in comparative hydrology studies, which facilitate the transfer of knowledge from data-rich to ungauged catchments or regions where direct hydrological measurements are scarce or non-existent (Singh et al., 2014; Sawicz et al., 2011; Tegegn et al., 2022; Tshimanga et al., 2022). Bing able to predict in ungauged catchments enhances our ability to manage water resources efficiently and sustainably.

The availability of open-access environmental data is increasing, however, data sources are distributed across different

40 platforms and stored in various formats, which requires further efforts in data collection and pre-processing. Many state authorities and governmental research institutes comply with open data policies, significantly breaking down the barriers related to data sharing issues that have existed in the hydrology community for decades (Kibler et al., 2014). Numerous platforms and websites offer hydrological data from various sources, which greatly benefits large-scale studies. For example, the Global Runoff Data Base provides streamflow measurements from over 10,000 hydrological stations (GRDC;

45 http://www.bafg.de/GRDC/), the HydroSHEDS database offers a suite of hydrographic data (https://www.hydrosheds.org/), the European Centre for Medium-Range Weather Forecasts provides multiple climate reanalysis products (https://www.ecmwf.int/), and Hydroweb supplies satellite-derived water levels for global rivers and lakes (https://hydroweb.theia-land.fr/). However, large-scale studies face a key challenge: They require a vast amount of data for development, and users often spend considerable time and effort navigating through different platforms for data collection and

50 employing various programming packages for data pre-processing.

Many studies are, consequently, focused on creating large-sample hydrology (LSH) datasets providing streamflow data for a large number of catchments datasets following the Findable, Accessible, Interoperable, and Reusable (FAIR) principle (Wilkinson et al., 2016). Newman et al. (2015b) created a comprehensive hydrometeorological dataset, which includes daily forcings and hydrologic response data for 671 small- to medium-sized basins throughout the contiguous United States

55 (CONUS). This dataset was further enriched by Addor et al. (2017), who conined the term "Catchment Attributes and Meteorology for Large-sample Studies" (CAMELS) to describe their collection of catchment attributes for the same basins. Since then, CAMELS datasets have been developed in several countries, including Chile (Alvarez-Garreton et al., 2018), Brazil (Chagas et al., 2020), Great Britain (Coxon et al., 2020), Australia (Fowler et al., 2021), and Switzerland (Höge et al., 2023). Some other datasets provide hydrometeorological timeseries and catchment characteristics similar to CAMELS

60 conventions, but with diverging naming, such as LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe (LamaH-CE, Klingler et al., 2021) and Iceland (LamaH-Ice, Helgason and Nijssen, 2023), historical hydro-meteorological time series and signatures for 24 catchments in Haiti (Simbi, Bathelemy et al., 2023), a large-scale benchmark dataset for data-driven streamflow forecasting (WaterBench-Iowa, Demir et al., 2022), China Catchment Attributes and

Meteorology dataset (CCAM, Hao et al., 2021). These datasets are open access, well-formatted, and encompass a wide range of comprehensive variables related to hydrological processes (Klingler et al., 2021). They provide information for hydrological studies and water resources management (Jehn et al., 2020; Meyer Oliveira et al., 2023; Frame et al., 2021; Tang et al., 2023), but also serve as reference for the development of hydrological models, as well as the training and testing of data-driven algorithms (Nearing et al., 2024; Mai et al., 2022; Kratzert et al., 2019, 2021, 2018; Yin et al., 2022; Liu et al., 2023b). Details about recent progress and their applications have been summarized by Addor et al. (2020). Kratzert et al. (2023) introduced the Caravan platform, which consolidates the national CAMELS datasets into a singular dataset derived from global sources, which further improve the accessibility of the dataset. However, Clerc-Schwarzenbach et al. (2024) concerned the Caravan dataset, which uses ERA5-Land reanalysis data, for exhibiting an unrealistically high potential evapotranspiration and a significant discrepancy in precipitation relative to the original CAMELS datasets. Such differences in meteorological forcings influence model results. Consequently, they advocated for the augmentation of the Caravan dataset with the forcing data present in the original CAMELS dataset, underscoring the ongoing relevance of the original CAMELS development.

While the already existing CAMELS datasets cover a wide range of hydroclimatic conditions and catchments characteristics it can still be considered incomplete in terms of the catchment diversity. Especially, the lack samples from low-lying, small-size, and groundwater-dominated catchments. In the existing CAMELS datasets with a total number of 3308 catchments (CH: 331, GB:671, CL:516, BR: 897, AUS:222, US: 671), only 7.5% of the catchments (250 out of 3308) have an elevation lower than 100 m, and 9.5 % of the catchments have an area size smaller than 50 km^2. Hydrological regimes, with respect to their baseflow and peak flow differ between small and large catchments. The same holds for low lying catchments, where hydrological regimes are more affected by groundwater related processes. Hence, samples of small-sized and groundwater dominated catchments are necessary to increase the diversity of current existing CAMLES family. Additionally, information and data from well-established physical hydrological models have not yet been provided by previous LSH datasets. Simulated runoff, obtained from hydrological models, offers a valuable benchmark dataset and can potentially also be provided at ungauged basins.

It is worthy to mention that LSH provides an invaluable foundation to scientists aiming to apply data-driven machine learning (ML) techniques in hydrology, as they undoubtedly offer an ideal environment for benchmarking, training, and testing ML algorithms. CMAELS and similar LSH have already contribute to ML advancements, such as predicting streamflow (Wilbrand et al., 2023), transfer learning (Ma et al., 2021), testing advanced algorithms (Yin et al., 2022), hybrid modelling (Espinoza et al., 2023) and global-scale flood forecasting (Nearing et al., 2024). In the context of ML-based hydrological modelling, physical-informed data driven algorithms, which combine the strength of traditional physics-based approaches with ML/DL models, show often enhanced performance (Konapala et al., 2020; Liu et al., 2023a). Therefore, hydrological information from well-established physically based hydrological models (if existing), would further benefit applications of CAMELS datasets. Previous CAMELS datasets provide simulated streamflow from conceptual hydrological models, such as LamaH-CE, which is insufficient in many Danish River systems. Groundwater levels close to ground surface are found throughout the country,

reliable surface water modelling in Denmark always requires information of groundwater dynamics (Duque et al., 2023; Koch et al., 2021; Schneider et al., 2022).

The points presented above have motivated us to compile and introduce CAMELS-DK for Danish catchments. Denmark covers roughly 43,000 km^2, the topography is flat (highest point is 170 m above sea level), climate is temperate with precipitation evenly distributed over the year (annual average precipitation ranging between 600 mm in the east to 1000 mm in the west). The flat terrain and wet climate generate about 69,000 km of river courses and 195,000 lakes (https://eng.mst.dk/water/aquatic-environment/lakes-and-watercourses). CAMELS-DK consistent data for over 3300 catchments, which cover almost the entire land area of Denmark. The dataset has a median catchment area of 19.61 km^2 with a median elevation of 31.94 m. Around 10%, i.e. 304 of these catchments are gauged and contain observed runoff, but the entirety of catchments in this dataset contains consistent simulated hydrological data, including simulated runoff and groundwater dynamics, in addition to hydroclimatic forcing and catchments attributes. These simulations are from a spatially distributed hydrological model, the National Hydrological Model of Denmark (DK-model), which has been thoroughly validated against data from 304 streamflow stations and approximately 40,000 groundwater wells. The DK-model provides a satisfying and consistent simulation of streamflow and incorporates a sophisticated 3D representation of groundwater processes, including a detailed hydrogeological model. With the first CAMELS dataset to provide a broad range of consistent and high-quality simulated data, we intend to generate a testbed for hybrid or physics-aware ML developments to further accelerate the development of such model types.

The objective of this study is to describe the extensive repository of hydrometeorological time series and catchment attributes in Denmark. We followed the routine of previous CAMELS dataset and ensure this dataset is comparable and interoperable to the already existing CAMELS datasets. Additionally, some new features are introduced which have not been found in the previously published CAMELS datasets, such as the incorporation of simulation results (shallow/deep groundwater levels, irrigation) of a spatially distributed hydrological model and features of hydrogeological model and observed groundwater abstraction for irrigation. The dataset follows the FAIR guiding principles for scientific data management and stewardship. We believe this dataset will enrich existing CAMELS database with catchments that have different features and hydrological regimes than the already existing datasets, which respect to groundwater influence and size. The paper is organized as follows: Section 2 describes the catchment delineation. Section 3 presents the dynamic variables including climate forcings, observed streamflow, and simulations from the hydrological model. Section 4 presents the sources and features of catchment attributes of topography, soil types, land use, and geology. In Section 5, we discuss the hydrological processes in Denmark based on the provided dataset. Section 6 contains a short summary of the paper.

2 Catchment delineation

Denmark has been divided into 3351 topographical catchments, referred to as ID15v2.5 catchments (Fig. 1). The division was originally prompted by the need to know the topographic areas upstream monitoring stations, lakes, and marine waterbodies.

Thus, enabling the calculation of runoff and loads of nutrients, suspended matter and other chemical components coupled to the corresponding topographical area. For modelling purposes, the observed catchments were further subdivided and

130 supplemented by a delineation also for unmonitored catchments. A topology for the sub-catchments was established, describing the downstream relations.

The catchment delineation seeks to provide sub catchment areas at a size of about 15 km^2 (hence the name ID15) where possible. However, the criteria that all lake outlets and monitoring stations must coincide with an outlet from a ID15 sub-catchment introduces sub catchments sometimes considerably smaller than 15 km^2. Alongside the establishment of new lakes

135 or additional monitoring stations the delineation is subject to periodically updates, with additions of new or alteration of existing catchment boundaries. The most-update version ID15v2.5 includes 3351 sub-catchments with a median area of 12.98 km^2, varying from less than one square kilometer to more than 100 km^2 for the ID15 containing the largest Danish Lake. Both very large and some very small ID15s were created to accommodate lake catchments and monitoring stations. The delineation of the ID15 catchments follows topographic catchments. However, especially in flat areas and areas with modified streams and

140 canals, local knowledge of such alterations was accounted for to delineate the actual surface water catchments. Additionally, ID15 catchments also coupled with the National Hydrological Model of Denmark (details are in Section 3.3), time series of simulated streamflow are available at catchment outlets where stream are included in the national hydrological model.
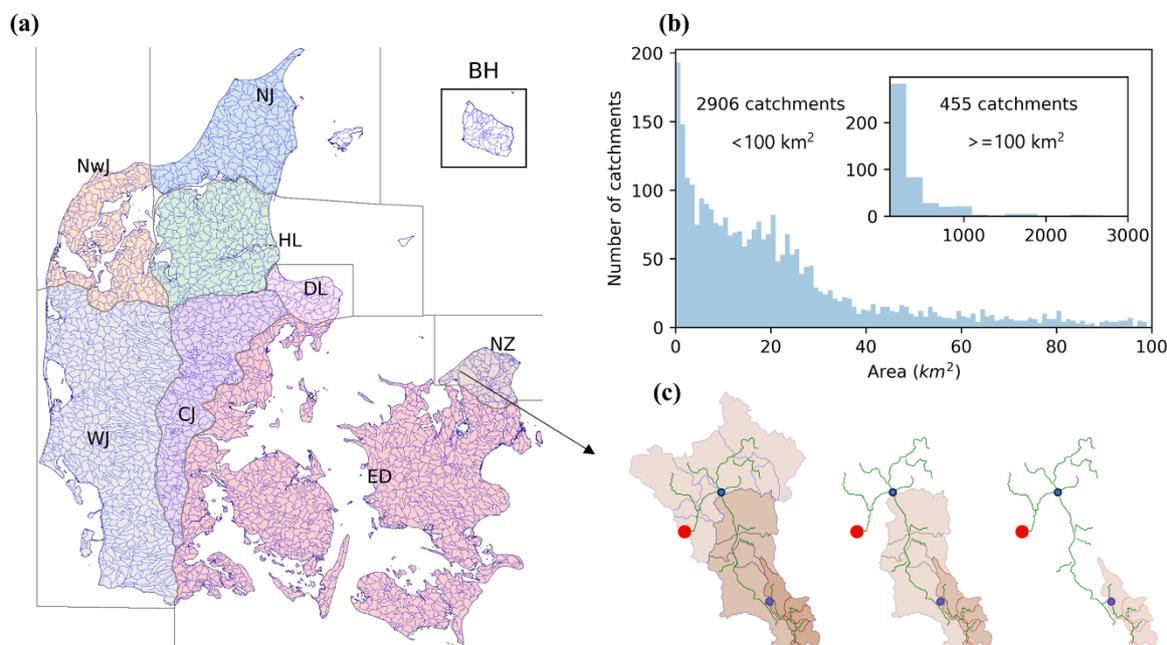
Figure 1. CAMELS-DK domain. (a) ID15 catchments and geological regions of Denmark, including northern Zealand (NZ), eastern Denmark (ED), western Jutland (WJ), central Jutland (CJ), Djursland (DL), Himmerland (HL), northwestern Jutland (NwJ), Northern Jutland (NJ), and Bornholm (BH), (b) distribution of catchment area, and (c) an example of river channels and catchment boundaries delineated by ID15 shapefile in Northern Zealand. The green lines are river channels, red dot indicates the location of a hydrological station downstream the river section, blue dots are examples basin outlets, and the corresponding catchment upstream drainage area are filled with light colour.

5

3 Dynamic variables

3.1 Climate data

The climate time series are based on national gridded products produced by the Danish Meteorological Institute (DMI). Datasets with daily timesteps going back to the year 1989 are based on observations from a network of in-situ weather stations. Daily rainfall time series, aggregated to catchment scale with an area weighted mean function, were derived from the 10 km^2 'Klimagrid Danmark' data from DMI (Scharling, 1999b). The gridded dataset is based on an inverse-distance interpolation of daily observations of precipitation at DMI's stations. The station density changes throughout the years, but generally moves between 200 and 500 stations, distributed across all of Denmark. As many hydrological applications are extremely sensitive to the exact amount of rainfall, the raw observations of precipitation were corrected for sensor/gauge under-catch. The correction was applied dynamically, accounting for rainfall intensity, wind speed and temperature (solid or liquid precipitation). CAMELS-DK provides corrected precipitation at ID15 scale based on freely available uncorrected precipitation observations, wind speed and temperature data in 10km grid format and the correction approach described by (Stisen et al., 2011). The uncorrected or DMI corrected precipitation data are not part of this dataset. Similarly to the precipitation dataset, DMI also provides gridded datasets of wind speed, air temperature and potential evapotranspiration (Scharling, 1999a). All variables are provided at daily timesteps; however, at an original resolution of 20 km^2. Potential evapotranspiration is calculated based on the Makkink formula, specifically modified for Danish conditions.

Table 1. Dynamic variables in CAMELS-DK. The time series are in daily scale, and spatially aggregated with the mean function.

Time series class	Time series name	Description	Unit	Sources
Climate variables	precipitation	Daily accumulated precipitation	mm/d	DMI's Frie Data
	temperature	Daily mean temperature	°C	(https://www.dmi.dk/frie-
	pet	Daily potential evapotranspiration (Makkink)	mm/d	data)
Hydrological variables	Qobs	Observed streamflow	m^3/s	DK-model (https://dennationalehydrologiskemodel.dk/)
	Qsim	Simulated streamflow	m^3/s	
	DKM_wcr	average water content in root zone	-	
	DKM_dtp	phreatic depth to surface layer	m	
	DKM_eta	actual evapotranspiration	mm/d	
	DKM_ rec	Total recharge to SZ	mm/d	
	DKM_sdr	SZ drainage flow from point	m^3/s	
	DKM_sre	SZ exchange flow with river	m^3/s	
	DKM_gwh	groundwater head in major deeper aquifer layers	m	
	DKM_irr	Irrigation	m^3/s	

Groundwater abstraction	abstraction	Annual abstractions from groundwater	m³/s	Jupiter database (https://www.geus.dk/produkter-ydelser-og-faciliteter/data-og-kort/national-boringsdatabase-jupiter)

3.2 Observed streamflow

170   Streamflow is provided by Danish EPA (under the Ministry of the Environment and Food) and made publicly available by Aarhus University through the surface water database of the (https://odaforalle.au.dk/). Water levels are measured sub-daily (minutes) by hydraulic sensors at hydrological stations and aggregated to daily values. Daily streamflow is calculated by the sub daily water levels applying these to the established dynamic rating curves. The rating curves vary throughout the year due to changes in river cross sections caused by erosion and sedimentation or regulation, and by vegetation growth and -cutting.

175   Mean daily water discharges are calculated from sub-daily time series and are accessible through (https://odaforalle.au.dk/). Quality control is performed by the Environmental Protection Agency. Streamflow time series are available for over 1000 hydrological stations spread across Denmark. However, many stations are unsuited for use in hydrological modelling. Some station may have very limited time series length, other stations may have dubious discharge values. Others may show anthropogenic influence (e.g. owing to the operation of sluices or pumping stations), which cannot be represented adequately

180   in hydrological models. This meant that, based on experience with the calibration and validation of the DK-model, the observations dataset was limited to 304 stations (Stisen et al., 2020). Additionally, some of the stations (4 stations) were located in the middle of a ID15 catchment, these stations were excluded from the dataset. The spatial distribution of hydrological stations is displayed in Fig 2a. Small catchments (area $< 20 \text{ km}^2$, 37 stations) are located in the upstream of larger river systems and in coastal areas. Large catchments (area $> 200 \text{ km}^2$) are mainly located on the Jutland peninsular. The largest river of

185   Denmark, River Gudenå, in central Jutland, has a catchment area of 2602 km² at the most downstream station.

Data availability of the 304 stations involved in CAMELS-DK is shown in Fig 2. Most of the stations have observed streamflow available during the entire years during 1989 to 2005 (Fig. 2b). However, many hydrological stations . Around 200 stations have available streamflow data for 90% of the days during the period of 1990 to 2019. Most of the stations (around 280) have daily streamflow data for 90% of the period 1990 to 1999, while the data availability declined during the period 2000 to 2019.

190
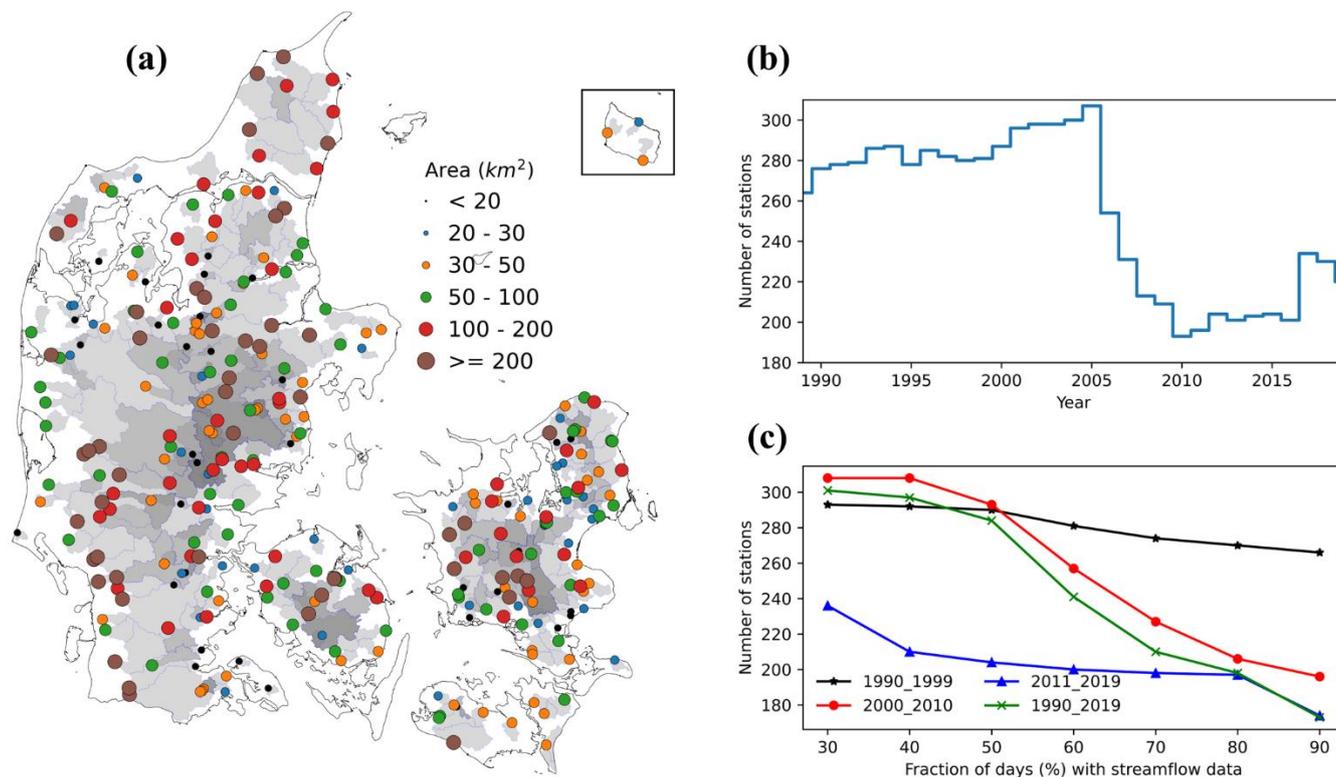
Figure 2. Streamflow data availability for all gauges contained in the CAMELS-DK dataset. (a) The distribution of hydrometrical stations and adherent catchments, (b) the number of stations have available streamflow over the entire year, and (c) the availability of observed streamflow for different period shown by the number of stations with percentage of available streamflow.

3.3 DK-model simulations

The National Hydrological Model of Denmark (hereafter referred to as DK-model) is an integrated groundwater surface water model covering most of the Danish land area of approximately 43,000 km^2. It has been continuously developed at the Geologic Survey of Denmark and Greenland (GEUS) since the late 1990's (Højberg et al., 2013; Stisen et al., 2020; Henriksen et al., 2019; Koch et al., 2021; Schneider et al., 2022). It has been used in public consultancy projects, targeting, for example, water resource, evaluation of the impact of climate change on the hydrological cycle, or nitrate retention. However, the DK-model as a whole or in parts also has been and is being used actively in various national and international research projects, further pushing its development (Martinsen et al., 2022; Mahmood et al., 2023; Soltani et al., 2021; Seidenfaden et al., 2022) .

The DK-model is setup in the MIKE SHE model code (Abbott et al., 1986; DHI, 2020), which is used to fully couple a 3D description of subsurface flow to 2D overland flow, a simple two-layer description of the unsaturated zone as well as 1D kinematic routing of streamflow. It runs as a transient model at a maximum simulation timestep of 24 hours, and, for the historic period covered by this dataset, with daily climate forcing of precipitation, potential evapotranspiration and temperature (as described in section 3.1). In its most recent version, which serves as a basis for the data presented here, it exists at horizontal

resolutions of both 100 m and 500 m (Henriksen et al., 2021),where we are using results from the 500 m version. Special attention has been paid the representation of surface and surface-near processes (shallow groundwater). The model has been

210 calibrated for the period 2000 to 2010 against 304 timeseries of daily streamflow observations (mostly identical to the ones provided in this dataset), and against groundwater head observations from roughly 40,000 intakes distributed across the entire country accounting for roughly 700,000 individual observations (some intakes provide single measurements, others timeseries of measurements). The most significant anthropogenic impacts are accounted for in the model: Groundwater extraction for domestic and industrial water use is included, using data on annual average extractions from waterworks in the national well

215 database Jupiter (https://eng.geus.dk/products-services-facilities/data-and-maps/national-well-database-jupiter). Groundwater extraction for irrigation in agriculture, which occurs predominantly in the western parts of Denmark, is accounted for based on dynamic calculations of soil water deficit in MIKE SHE (Danapour et al., 2021). Where sewage plants streamflow their outflow into streams, these are included as point sources in the stream network, based on yearly averages. As roughly half of the agricultural land in Denmark is artificially drained, saturated zone drainage is included (Møller et al., 2018). Due to the

220 coarse grid resolution and the related inability to represent the drainage infrastructure explicitly as well as the effects of microtopography, and other aspects such as urban sewer systems acting as further drainage, the DK-model includes drainage across most of Denmark. It is parameterized using drain depths and time constants distributed based on the BASEMAP land use map. Generated drain flow is routed to the river network.

Due to its nature, the DK-model simulates the hydrological cycle in an integrated manner, allowing to retrieve information

225 about various compartments from its results. With hydrological processes in Denmark in general, and streamflow in particular, being strongly groundwater-dependent (Duque et al., 2023), we provide, amongst others, groundwater related variables from the DK-model simulations. We could previously show that data-driven methods for streamflow estimation benefit from information contained in simulated groundwater levels across Denmark (Liu et al., 2023a). Hence, for the CAMELS-DK dataset, we provide a set of simulation results from the DK-model at daily timesteps (Table 1). Those include: (1) The simulated

230 streamflow at the outlet of each ID15 catchment. (2) The water content in the root zone of the two-layer unsaturated zone module of MIKE SHE. Due to the simplified description of the unsaturated zone, this accounts for the average water content across the entire root zone, where the root depth in the DK-model is variable in space and time dependent on soil type, vegetation type and season. (3) The same unsaturated zone module also provides actual evapotranspiration. (4) Recharge from the unsaturated to the saturated zone accounts for exchange between unsaturated and saturated zone, including bypass flow,

235 and exchange between the saturate zone and overland flow or evaporation (when the saturated zone is at the surface). Here, positive values indicate flow downwards to the saturated zone, whereas negative values indicate upwards flow from the saturated zone, occurring for example in areas of upwelling groundwater. (5) The simulated exchange between streamflow in rivers and the saturated zone gives an indication of the groundwater baseflow contribution to streamflow. Most Danish rivers can be classified as predominantly gaining streams. (6) Another important part of the hydrological cycle in Denmark is artificial

240 drainage, which we therefore included in the present dataset. Direct groundwater contribution together with flow from saturated zone drainage account for most of the streamflow generated in the DK-model. Finally, the state of the groundwater is included

9

by two variables: The depth to phreatic, which represents the depth below the surface to (7) the uppermost simulated groundwater table (which can be a perched groundwater table), as well as (8) the simulated groundwater head in the main deeper aquifer (typically in a few tens of metres depth). The former gives an indication of surface-near interactions between

245 groundwater and surface water, reacting more quickly to precipitation events, whereas the latter typically shows a more delayed response to climate.

3.4 Groundwater abstraction

Domestic and industrial water supply in Denmark is almost exclusively covered by groundwater. Data on groundwater abstraction from water works are reported to the national well database Jupiter (Hansen and Pjetursson, 2011), as aggregated

250 yearly values per well or well field. Those data are used as input to the DK-model and are also provide as part of the CAMELS-DK dataset. Due to the close coupling between surface and subsurface processes in Denmark, groundwater extraction locally impacts streamflow, as for example acknowledged in national water resources assessments (Henriksen et al., 2023, 2008).

4 Catchment attributes

4.1 Location and topography

255 The dataset is organized based on the ID15 catchment. All the catchments are identified with an eight-digital name according to the shapefile. The locations of catchment outlets are provided as longitude and latitude with the projection of WGS 84 / UTM zone 32N (EPSG:32632). Flow direction indicates the downstream catchments, the column was filled with –9999 for most downstream catchments. The number of upstream catchments and the list of upstream catchment IDs are also provided. The area of the individual ID15 catchments and their entire upstream accumulated area originate from the provided shapefile.

260 Gauge types indicate if the catchment is gauged or ungauged, when 'True' indicates that the catchment is a gauged catchment, observed and simulated streamflow are available in the dataset and 'False' indicates only simulated streamflow is available. Details about the observed data, such as the length and monitoring data of the data, are provided. Catchment topography described using various statistics derived from elevation and slope are also provided in the attributes table, see Table 2. Please be aware that catchment zonal statistics are based on the entire upstream area.

265

Table 2. Catchment location and topography

Attribute name	Description	type/Unit	Data source
catch_id	Catchment identifier (eight-digit ID15 catchment name)		
catch_outlet_lon	Catchment outlet longitude		
catch_outlet_lat	Catchment outlet latitude		
catch_flow_dir	Catchment flow direction		
catch_accum_number	The total number of upstream accumulated ID15 catchment units		

Earth System
Open Access
Science
Data
Discussions

catch_area	Catchment accumulated area	m2	
gauged_type	Boolean values to indicate the catchment is gauged (True) or ungauged (False)		
gauge_record_pct	If gauged catchment, the time percentage with available observed streamflow	%	
ele_min	Minimum elevation of the catchment	m	
ele_median	Median elevation of the catchment	m	
ele_mean	Mean elevation of the catchment	m	
ele_max	Maximum elevation of the catchment	m	
slope_min	Minimum slope of the catchment	m/km	10m DEM
slope_median	Median slope of the catchment	m/km	
slope_mean	Mean slope of the catchment	m/km	
slope_max	Maximum slope of the catchment	m/km	
pct_flat_area	Percentage of catchment area with slope smaller than 3m/km	%	

4.2 Climate indices

Climate indices are based on the aforementioned grided datasets of precipitation, temperature, and potential evapotranspiration.
270 To provide consistency with previous CAMELS datasets, we compute the same climatic indices for all catchments in CAMELS-DK as Addor et al. (2017), such as the mean value of daily precipitation, seasonality, and frequency. Details of these climate indices are listed in Table 3. It is important to note that in CAMELS-DK climatic indices are calculated for water years from 1989 to 2019, which is consistency to the availability of observed streamflow.

275 **Table 3. Climate indices**

Attribute	Description	Unit
p_mean	Mean of daily precipitation	mm/d
t_mean	Mean of daily temperature	$\circ$C
pet_mean	Mean of daily potential evapotranspiration	mm/d
aridity	aridity (PET /P, ratio of mean PET to mean 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); values close to 0 indicate uniform precipitation throughout the year)	-
frac_snow	fraction of precipitation falling as snow (i.e., on days colder than 0 $\circ$C)	%
high_prec_freq	frequency of high precipitation days ($\geq$5 times mean daily precipitation)	d/yr
high_prec_dur	average duration of high precipitation events (number of consecutive days $\geq$5 times mean daily precipitation)	d
high_prec_timing	season during which most high precipitation days ($\geq$5 times mean daily precipitation) occur	season

Earth System
Science
Data

low_prec_freq	frequency of dry days (< 1 mm/d)	d/yr
low_prec_dur	average duration of dry periods (number of consecutive days < 1 mm/d)	d
low_prec_timing	season during which most dry days (< 1 mm/day) occur	season

4.3 Streamflow signatures

Hydrologic signatures are based on the observed daily streamflow at 304 hydrological stations and DK-model simulated streamflow at the outlet of 2942 catchments (388 catchments have no surface river flows). We calculated 13 hydrological signatures presented by Addor et al. (2017), which is consistent with the other CAMELS datasets (Alvarez-Garreton et al., 2018; Chagas et al., 2020; Höge et al., 2023). These signatures are listed in Table 4. Additionally, nine signatures developed by McMillan (2020) that relate to groundwater storage, groundwater dynamics and baseflow, nine signatures showing infiltration excess and saturation excess of overland flow and, 15 signatures promoted by Gnann et al. (2021) are included as hydrological signatures (Table 4). Hydrologic signatures are calculated for the flow time series available during water years from 1989 to 2019 using MATLAB code provided by Gnann et al, (2021).

Table 4. Hydrographical signatures based on observed and simulated streamflow.

Attribute	Description	Unit
q_mean	mean daily streamflow	mm/timestep
TotalRR	total runoff ratio (runoff divided by the precipitation)	
QP_elasticity	streamflow precipitation elasticity (sensitivity of streamflow to changes in precipitation at the annual timescale, using the mean daily streamflow as reference)	
FDC_slope	slope of the flow duration curve (between the log- transformed 33rd and 66th streamflow percentiles)	
BFI	baseflow index (ratio of mean daily baseflow to mean daily streamflow, hydrograph separation performed using the Ladson et al., 2013 digital filter	
HFD_mean	mean half-flow date (date on which the cumulative streamflow since 1 October reaches half of the annual streamflow)	day of year
Q5	5% flow quantile (low flow	mm/timestep
Q95	95% flow quantile (high flow	mm/timestep
high_Q_frequency	frequency of high-flow days (> 9 times the median daily flow	
high_Q_duration	average duration of high-flow events (number of con- secutive days > 9 times the median daily flow	timestep
low_Q_frequency	frequency of low-flow days (< 0.2 times the mean daily flow	
low_Q_duration	average duration of low-flow events (number of consec- utive days < 0.2 times the mean daily flow	timestep
zero_Q_frequency	frequency of days with Q=0	

4.4 Land use and land cover

We provide both national and continental data sources for land use and land cover (LULC). The LULC attributes are derived from two datasets: the CORINE Land Cover (CLC) from the European Copernicus program (Büttner et al., 2004) and Basemap04 developed by Aarhus University (Levin, 2022). CAMELS-DK calculates the catchment-scale percentage of key

LULC classes. The aggregated areal percentages of five land cover types are provided for all catchments, including the percentages of forest, crops, water, urban areas, and wetlands (Table 5).

295 The CORINE Land Cover (CLC) product offers a pan-European land cover and land use inventory with 44 thematic classes, ranging from broad forested areas to individual vineyards (Büttner et al., 2004). The product is updated every six years with new status and change layers for 1990, 2000, 2006, 2012, and 2018. The dataset is based on multiple satellite data sources, such as Landsat, SPOT-4/5, RapidEye, and Sentinel-2. The geometric accuracy of the CLC data is 100 meters for the first dataset released in 1990, with accuracy improving to better than 100 meters in subsequent years. Basemap04 provides four

300 maps of LULC representing the years 2011, 2016, 2018, and 2021, with a spatial resolution of 10 meters. The dataset combines existing thematic geographic information, such as census mapping of state forests, maps of protected habitats, agricultural field parcel maps, and cadastre maps (Levin, 2022).

Table 5. Land use attributes.

Attribute	Description	Yeas	Unit	Data source
pct_ forest_corine_yyyy	Percentage of agriculture		%	
pct_ crop_corine_yyyy	Percentage of grass and herb vegetation		%	
pct_ water_corine_yyyy	Percentage of medium-scale vegetation	1990, 2000, 2006, 2012, 2018	%	Copernicus CORINE land cover (CLC, 2024)
pct_ urban_corine_yyyy	Percentage of deciduous forest in		%	
pct_ wetlands_corine_yyyy	Percentage of mixed forest in		%	
pct_forest_levin_yyyy	Percentage of forest		%	
pct_ crop_levin_yyyy	Percentage of agriculture		%	
pct_ water_levin_yyyy	Percentage of lakes and streams		%	Levin, G. 2022. Aarhus
pct_ urban_levin_yyyy	Percentage of urban	2011, 2016, 2018, 2021	%	University, https://dce2.au.dk/pub/TR252.pdf
pct_ naturewet_levin_yyyy	Percentage of nature, (wet, agriculture, extensive)		%	
pct_ naturedry_levin_yyyy	Percentage of nature, (dry, agriculture, extensive)		%	

305

4.5 Soil types

Physical soil properties were derived from national and continental data sources. For the former, catchment average soil texture was derived for four soil layers, i.e., 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm at ~30 m spatial resolution. The Danish dataset by Adhikari et al. (2013) is based on a regression modelling using field data for 1958 soil profiles and 17

310 environmental covariates. In addition, catchment average soil physical properties were derived from the European Soil Data Centre (ESDAC) and the dataset of 3D Soil Hydraulic Database of Europe at 1 km and 250 m resolution. A broader set of variables

was processed, including rooting depth, saturated hydraulic conductivity, among others. Soil maps of ESDAC are at 1000 m spatial resolution and are based on data from the European Soil Database in combination with data from the Harmonized World Soil Database (HWSD) and Soil-Terrain Database (SOTER). Several hydraulic parameters from the dataset of 3D Soil

315   Hydraulic Database of Europe at 1 km and 250 m resolution, such as the water content at field capacity and saturated hydraulic conductivity (Table 6).

Table 6. Catchment attributes of soil types.

Attribute	Description	Unit	Data source and References
pct_sand	Sand content	%	
pct_silt	Silt content	%	
pct_clay	Clay content	%	
pct_organic	Organic carbon content	%	
bulk_density	Bulk density	g/cm^3	European Soil Database Derived data (Hiederer, 2013a, b)
tawc	Total available water content (from PTR)	mm	
pct_gravel	Coarse Fragments	%	
root_depth	depth available for roots	m	
FC	water content at field capacity	cm^3/cm^3	
HCC	hydraulic conductivity curve	log$_{10}$ [cm/d]	3D Soil Hydraulic Database of Europe at 1 km and 250 m resolution (Tóth et al., 2017)
KS	saturated hydraulic conductivity	log$_{10}$[cm/d]	
MRC	moisture retention curve	cm^3/cm^3	
THS	Saturated water content	cm^3/cm^3	
WP	water content at wilting point	cm^3/cm^3	
pct_claynor_30	Clay percentage in layer depth of 0-30cm	%	
pct_claynor_60	Clay percentage in layer depth of 30-60cm	%	
pct_claynor_100	Clay percentage in layer depth of 60-100cm	%	
pct_claynor_200	Clay percentage in layer depth of 100-200cm	%	Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., ... & Greve, M. H. (2013). High-resolution 3-D mapping of soil texture in Denmark. Soil Science Society of America Journal, 77(3), 860-876.
pct_fsandno_30	Fine sand percentage in layer depth of 0-30cm	%	
pct_fsandno_60	Fine sand percentage in layer depth of 30-60cm	%	
pct_fsandno_100	Fine sand percentage in layer depth of 60-100cm	%	
pct_fsandno_200	Fine sand percentage in layer depth of 100-200cm	%	
pct_gsandno_30	Coarse sand percentage in layer depth of 0-30cm	%	
pct_gsandno_60	Coarse sand percentage in layer depth of 30-60cm	%	
pct_gsandno_100	Coarse sand percentage in layer depth of 60-100cm	%	
pct_gsandno_200	Coarse sand percentage in layer depth of 100-200cm	%	

Earth System
Science
Data

4.6 Hydrogeology

Hydrogeological features were obtained from the 3D hydrogeological model implemented in the DK-Model (Stisen 2019).
The model is based on a layer structure, with in total of up to 28 hydrostratigraphical layers that have been modelled based on
borehole data and subsurface geophysics. The layers delineate the occurrence of, i.e., quaternary, pre-quaternary sand and clay
as well as limestone. The bottom and top of the hydrostratigraphical layers are mapped on a 100 m grid and five key variables
were compiled for the CAMELS-DK dataset: (1) Depth to chalk specifies the vertical distance from terrain to chalk. (2)
Thickness of the uppermost aquifer accumulates quaternary sand layers that are not intermittent by clay layers thicker than 1
m. (3) Depth to uppermost aquifer specifies the vertical distance from terrain to the top of the uppermost quaternary sand layer
thicker than 1 m. 4) Thickness of uppermost clay specifies the accumulated quaternary clay layers from terrain until a sand
layer thicker than 1 m is present. 5) Thickness of uppermost sand specifies the accumulated quaternary sand layers from
terrain until a clay layer thicker than 1 m is present. All variables are stated in meters and the catchment average has been
calculated for the CAMELS-DK dataset (Table 7). The hydrostratigraphical model for the island Bornholm is developed
separately from the rest of Denmark. Bornholm's geology is significantly different with Pre-Cambrian granite-gneisses and
Cambrian sandstones. Available data are interpreted to build a 3D voxel model for the island. Due to these differences, the
five hydrogeological features are not available for Bornholm. Moreover, the 14 classes of the surface geology map of Denmark
(Pedersen et al., 2011) were processed as catchment percentages.

Table 7. Catchment attributes of geologic features.

Attribute	Description	Unit	Data source
chalk_d	Depth to chalk	m	https://doi.org/10.2200
uaquifer_t	Thickness of the uppermost aquifer	m	8/FK2/UP1PBJ/E3WY
uaquifer_d	Depth to uppermost aquifer	m	BX &
uclay_t	Thickness of uppermost clay	m	https://doi.org/10.2200
usand_t	Thickness of uppermost sand	m	8/gpub/32631
pct_aeolain_sand	Percentage of aeolian sand, including dunes and cover sands	%	
pct_water_deposit	Percentage of freshwater deposits	%	
pct_marsh	Percentage of marsh, alternating tin	%	
pct_marine_sand	Percentage of Marine sand and clay	%	Surface Geology Map
pct_beach	Percentage of Beach ridges	%	of Denmark 1:200000
pct_sandy_till	Percentage of Sandy and gravelly till	%	
pct_till	Percentage of till, clay and fine-sandy	%	
pct_glaf_sand	Percentage of glaciofluvial sand and gravel	%	

pct_glal_clay	Percentage of glaciolacustrine laminated clay, silt and fine sand	%
pct_down_sand	Percentage of downwash sandy deposits	%
pct_glam_clay	Percentage of glaciomarine clay, silt and sand	%

5 Data Discussion

5.1 Regional variability in catchment attributes

340

Denmark has a gentle topography with many regions at a mean elevation lower than 50 m and low slopes (Fig. 3a, b). Across central Jutland, a ridge is located with elevations reaching up to about 170 m. This ridge, together with the prevailing westerly winds, is responsible for the regional precipitation patterns with higher precipitation in western Jutland, and lower precipitation in the more eastern parts of the country (Fig. 3c). Another important factor controlling hydrology across Denmark is the soil

345

type: soils are most clayey in eastern Denmark, central and northwestern Jutland (Fig. 3d), and least clayey (mostly sandy) in western Jutland. Higher percentage of clay soil generally relates to lower hydraulic conductivities in the topsoil, leading to higher groundwater levels and more artificial drainage. It also leads to larger amount of water available for evapotranspiration and lower runoff. In contrast, the thickness of uppermost sand is high in West Jutland, north Jutland, and Himmerland (Fig. 3e). The depth to chalk aquifers is high (> 250 m) in northwestern Jutland, central and western Jutland, but lower in North

350

Zealand and Djursland with typical depths of around 60 m (Fig. 3f). The temperate climate with precipitation clearly exceeding evapotranspiration, low-lying topography, and geology dominated by clay till or sandy meltwater deposits result in high groundwater levels and abundant groundwater resources in Denmark.
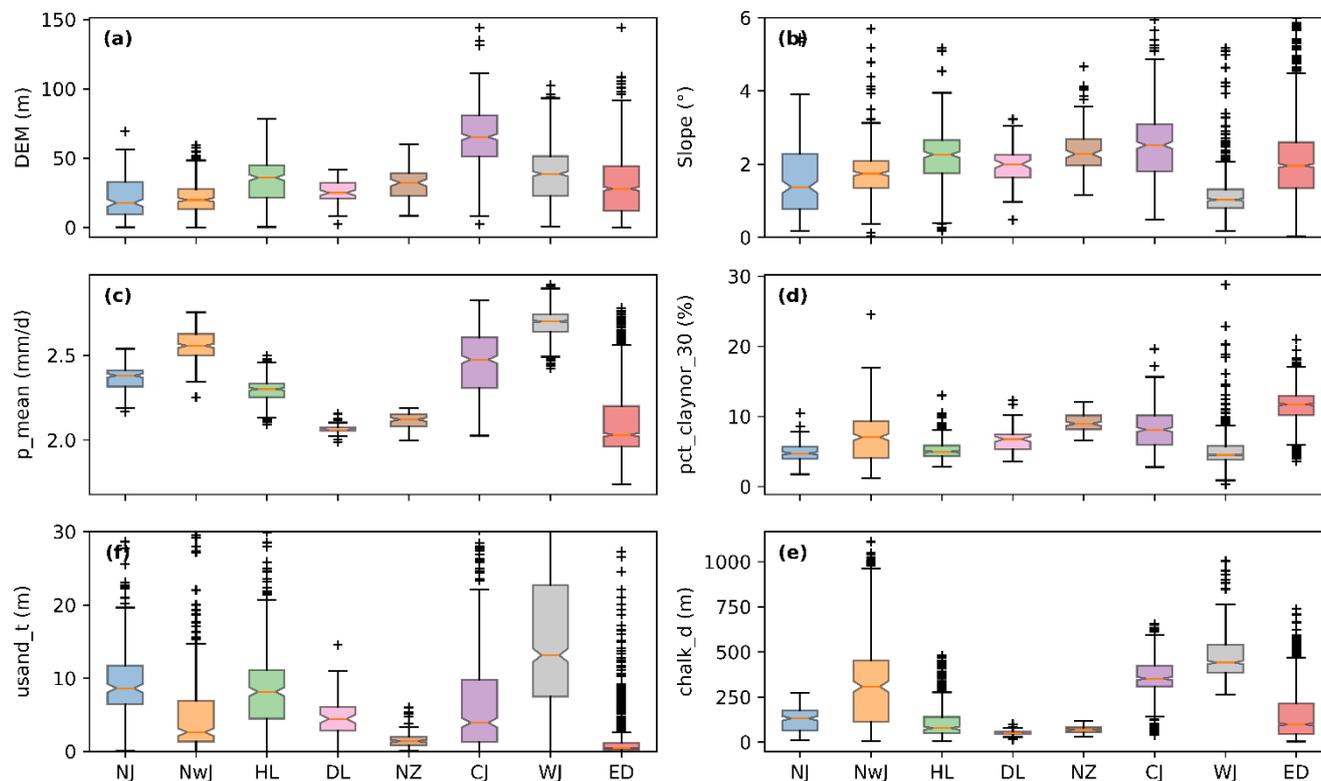
Figure 3. Regional distributions of (a) mean elevation, (b) slope, (c) mean daily precipitation, (d) Clay percentage in layer depth of
355 **0-30cm, (e) depth of chalk, and (f) Thickness of uppermost sand in geo-regions (NJ: North Jutland, NwJ: northwestern Jutland, HL: Himmerland, DL: Djursland, NZ: North Zealand, CJ: Central Jutland, WJ: Western Jutland, and ED: East Denmark).**

5.2 Surface water dynamics

CAMELS-DK provides observed streamflow data from 304 hydrological stations and simulated streamflow data for 2,942
360 catchments. However, 388 catchments lack observed or simulated streamflow data because they are located upstream or in coastal areas without distinct land surface river channels. The simulated streamflow is based on the DK-model, which was calibrated using observed streamflow data from the period 2000-2010, see section 3.3. Fig. 4 shows the Sutcliffe and Kling-Gupta efficiency (KGE) and Nash–Sutcliffe model efficiency coefficient (NSE) for the DK-model's simulated hydrography, referenced against observations from 1990-1999. The mean NSE is 0.46, and the median value is 0.64. The mean KGE is 0.64,
365 and the median value is 0.69, indicating that the simulated streamflow is satisfactory at many stations, with the simulations and observations showing temporal coherence and consistency. However, the simulations demonstrate lower performance for high-flow and low-flow conditions, as indicated by lower NSE values in central and northern Jutland (Fig.4). Consequently, deep learning methods and hybrid schemes have been investigated to improve streamflow forecasting (Liu et al., 2023a). Additionally, the DK-model's performance in ungauged catchments has not been fully validated due to the computational
370 intensity of model calibration.
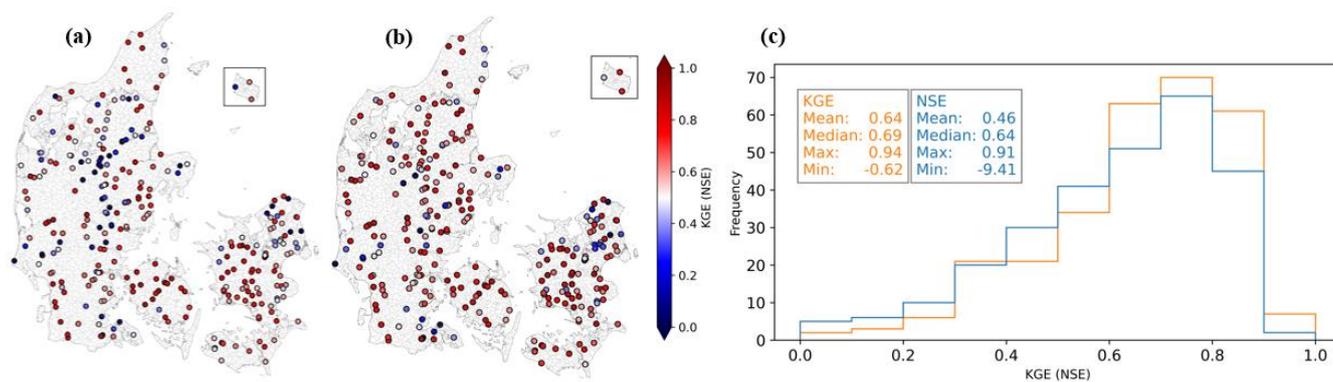
Figure 4. The performance of DK-model simulated streamflow at 304 hydrological stations. (a) spatial distribution of Nash–Sutcliffe model efficiency coefficient, (b) spatial distribution of the Kling–Gupta efficiency (KGE), and (c) histogram of NSE and KGE.

375
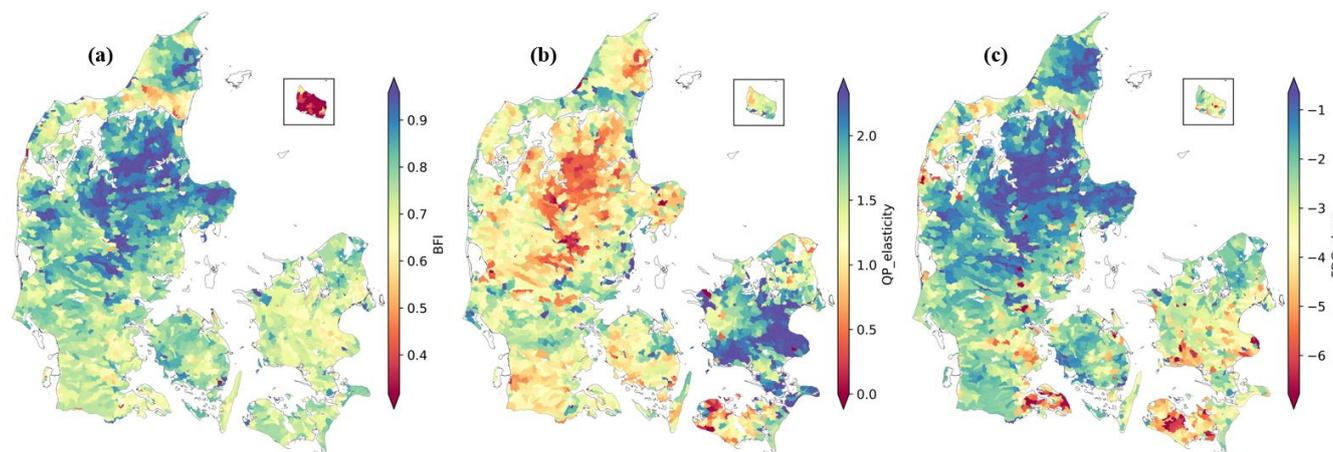
Figure 5. Spatial distribution of (a) Baseflow index, (b) Streamflow-precipitation elasticity, and (c) Slope of the flow duration curve.

380   Groundwater and surface water are closely linked in Denmark (Duque et al., 2023). Fig. 5 presents three hydrographic signatures derived from DK-model simulations: Baseflow Index (BFI), Streamflow-Precipitation Elasticity (QP-elasticity), and the Slope of the Flow Duration Curve (FDC_slope). BFI represents the proportion of streamflow that occurs as baseflow. The value is higher in Himmerland, Djursland, and central Jutland, indicating that a significant portion of the streamflow comes from groundwater seepage (Fig. 5a). The groundwater levels are lower in these areas, resulting in the slow and steady

385   release of groundwater into the streams, thereby exhibiting more stable flow conditions throughout the year. Groundwater buffers against short-term variations in precipitation in these areas. QP-elasticity indicates the sensitivity of streamflow to changes in precipitation, which is low in central Jutland and high in eastern Denmark. This suggests significant regulation of

streamflow by groundwater. The FDC_slope also shows higher values in eastern Denmark, indicating a highly variable stream

flow that is largely due to the quick runoff of rainfall into the stream (Fig. 5c).
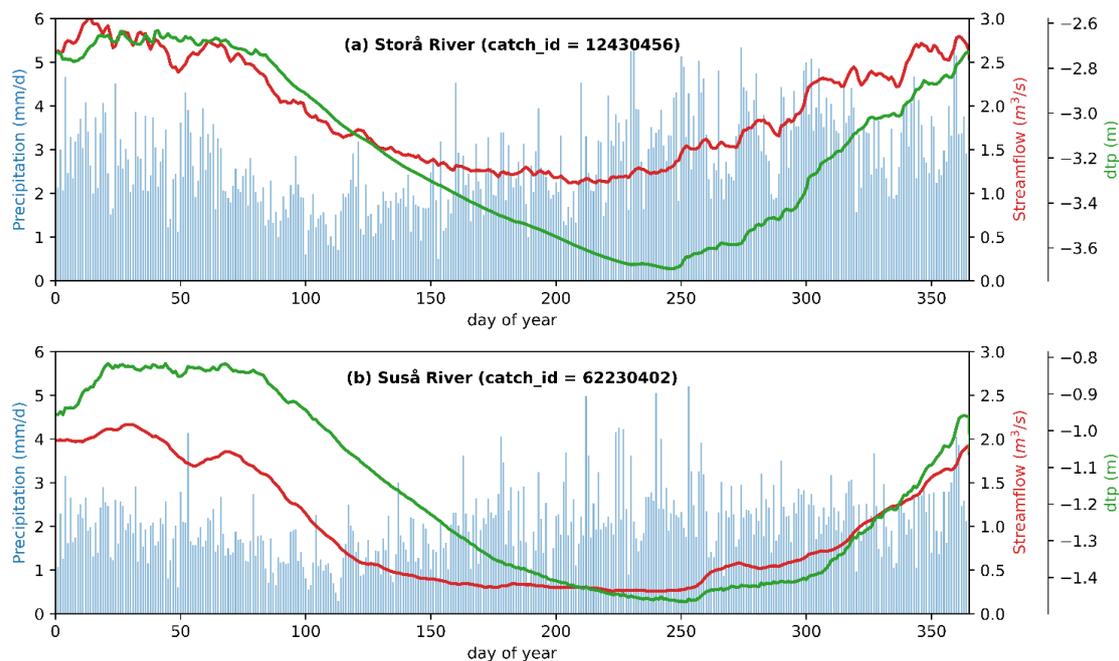
Figure 6. Climatology of precipitation observed streamflow, and DK-model simulated shallow groundwater levels (dtp) in (a) Storå River (the largest river in West Jutland), and (b) Suså River (the largest in East Denmark).

Groundwater – surface water interaction is significant at temporal scales as well. Fig. 6 shows the climatology (1990 - 2019) of precipitation, observed streamflow and depth to phreatic layers in Storå River (largest river in Western Jutland) and Suså River (largest river in East Denmark), close to each of their outlets. The spatial variability of precipitation is relatively small, and the seasonality is similar across Denmark. The two examples shown in Fig.6 have a comparable amplitude of precipitation during the year with highest precipitation occurring during the fall and winter months, and lowest precipitation during spring. Streamflow, though, has a more pronounced seasonality, due to higher evapotranspiration during summer, leading to lower runoff coefficients. Even in the climatology, the peaky streamflow response of the sandier Storå catchment becomes apparent, compared to the more clayey Suså catchment. Moreover, summer baseflow in the Storå catchment is higher than in the Suså catchment (see also BFI in Fig. 5a, and clay percentage in Fig. 6d), which again is due to the better connectivity between (shallow and deep) groundwater and rivers due to the more sandy/higher conductivity subsurface conditions in large parts of Western Jutland. Linked to similar processes, differences in groundwater dynamics can be seen, with the Storå catchment exhibiting lower groundwater tables, but higher seasonal amplitudes of the two example catchments. The steady decline of groundwater levels during the summer months is due to elevated levels of evapotranspiration, limiting groundwater recharge during that time. Groundwater recharge typically starts again around the beginning of September, and streamflow levels follow closely and begin to rise from summer baseflow conditions around the same time.

19

5.3 Groundwater dynamics

410 Groundwater supplies nearly 100 percent of Danish domestic and industrial water use. Groundwater monitoring and modelling are thus important for sustainable water management in Denmark. Fig.7 shows catchment-aggregated variables from DK-model simulations, such as depth to phreatic layers (Fig. 7a), groundwater extraction for irrigation (Fig. 7b), and reported groundwater abstractions (Fig. 7c) from waterworks. The spatial pattern of phreatic depth shows the groundwater level is highest in eastern and southern Denmark, and lowest in northern Denmark. In some catchments located in Himmerland and

415 central Jutland, the average phreatic depth is around 15 meters blow land surface (Fig. 7a). An example of the time series of phreatic depth in a catchment in Himmerland shows an annual amplitude of approximately 5 meters. The groundwater table is higher in winter and lower in summer, which aligns with the seasonal variations in precipitation and evapotranspiration. Precipitation deficits can reduce the amplitude of phreatic depth, especially during winter months such as in 1996. Catchments with more shallow groundwater depths show typically smaller yearly amplitudes. Groundwater extraction for domestic and

420 industrial water supply is based on yearly data from the national well database, Jupiter (https://www.geus.dk/produkter-ydelser-og-faciliteter/data-og-kort/national-boringsdatabase-jupiter). Groundwater abstraction for irrigation is only available in Jutland since there is no significant irrigation activity in the other parts of Denmark. Irrigation is provided as determined dynamically, demand-based by the DK-model.
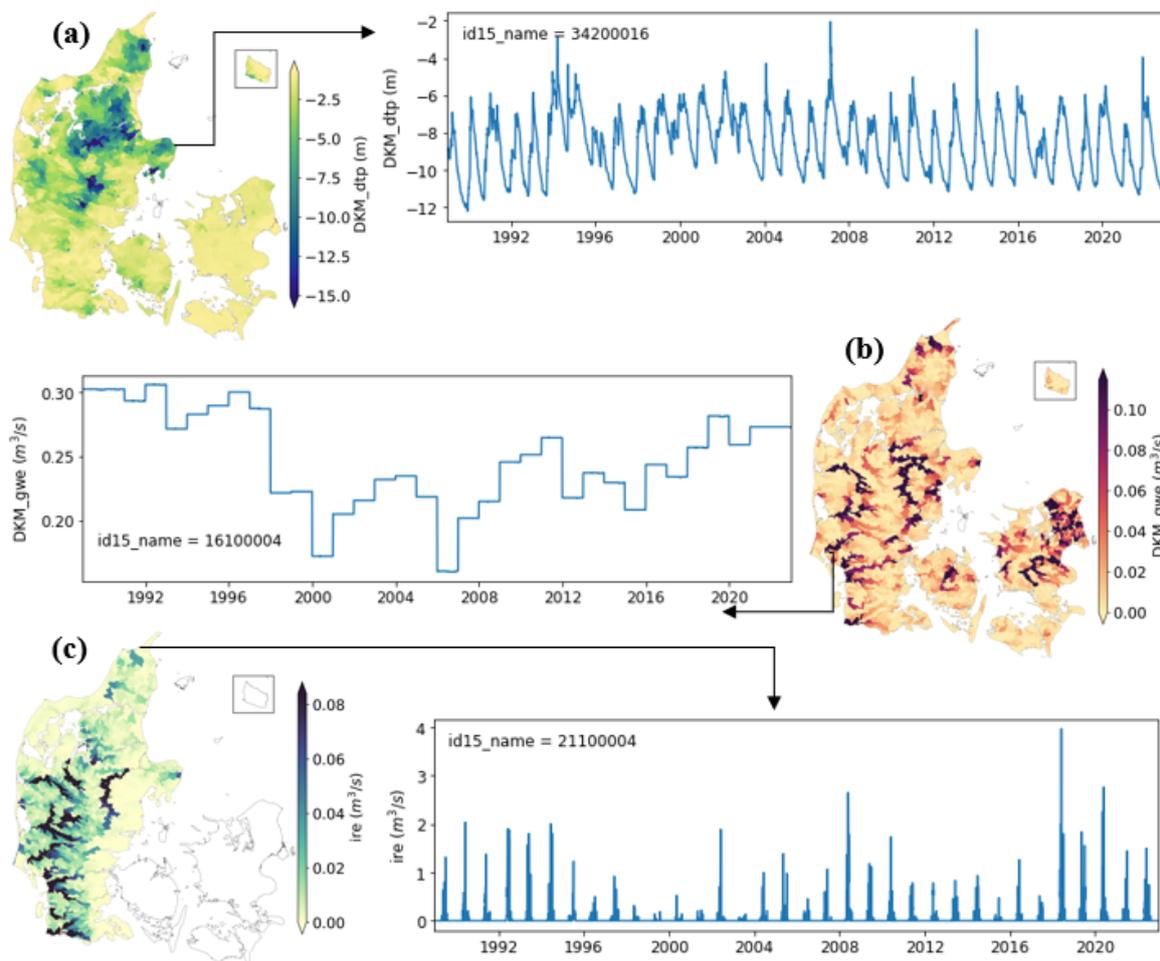
Figure 7. Groundwater dynamics, abstraction, and irrigation abstraction. (a) catchment average depth to phreatic layers with the time series of a catchment located in Himmerland, (b) catchment summary groundwater abstraction with the yearly abstraction amount time series of a catchment in western Jutland, and (c) catchment summary groundwater abstraction for irrigation, with time series of a catchment in northern Jutland.

6 Dataset structure

430   CAMELS-DK is available at https://doi.org/10.22008/FK2/AZXSYP (Koch et al., 2024). There are four folders after download and unzipping the dataset. Landscape attributes including climate indices, topography, hydrological signatures, land use, soil characteristics, and hydrogeological features are saved in a folder named 'Attributes'. These static attributes are saved in CSV files, where the index is the 8-digital identifier, and the columns are the names of variables (see Table 2-5). Timeseries of climate data, observed streamflow, DK-model simulated variables, such as depth to phreatic layers, average soil water content,

435   and actual evapotranspiration, and observed yearly groundwater abstractions are available in the folder 'Dynamics'. The timeseries data is saved separately for gauged and ungauged catchments. The shapefile of catchments boundaries, location and

the geo-regions of Denmark are provided in a folder named 'Shapefile'. The Python script of processing the time series and landscape attributes based on original datasets are provided in a folder named 'Python'.

7 Conclusion

440   In this study, we provided an updated version of CAMELS-DK, which involves various hydrometeorological variables and landscape features for 3330 catchments that covers the entirety of Denmark. Time series of hydrometeorological variables are at daily scale and span over the period from 1989 to 2023. Land cover/use are from two datasets and are available for a few years (1990, 2000, 2016, 2011, 2012, 2016, 2018, 2021), yearly groundwater irrigation abstraction data are available in Jutland. Shallow and deep groundwater levels are derived from the National Hydrologic Model of Denmark, which plays an important

445   role in hydrological systems for low-laying catchments.

The dataset is developed to assistant machine learning studies which require large samples for training, testing and validation. The dataset includes observed streamflow for 304 catchments. Other observed variables include information on, e.g, precipitation ang groundwater abstractions. The simulated variables, such as streamflow, groundwater levels, and irrigation have high accuracy in many regions of Denmark, which will benefit the development of physical-informed machine learning

450   algorithms. We provide most of the original raster data and Python scripts for data processing (level-1), based on which spatially aggregated time series and attributes were calculated for 3330 catchments (level-2). The dataset follows the FAIR rule and is convenient to use with other CAMELS-like datasets.

Appendix

455   Appendix A: Additional hydrological signatures

Attribute	Description	Unit	Reference
SnowDayRatio	Snow day ratio	–	Euser et al., (2013)
RLD	Rising limb density	1/day	
AC1	Lag-1 autocorrelation	–	McMillan, (2020) for groundwater
RR_Seasonality	Runoff ratio seasonality	–	
EventRR	Event runoff ratio	–	
StorageFraction	Ratio between active and total storage	–	
Recession_a_Seasonality	Seasonal variations in recession parameters	–	
AverageStorage	Average storage from average baseflow and storage-discharge relationship	–	
Spearmans_rho	Non-uniqueness in the storage-discharge relationship	–	
EventRR_TotalRR_ratio	Ratio between event and total runoff ratio	–	
VariabilityIndex	Variability index of flow	–	
IE_effect	Infiltration excess importance	–	McMillan, (2020) for surface water
IE_thresh	Infiltration excess threshold location (in a plot of quickflow volume vs. maximum intensity)	mm/day	
IE_thresh_signif	Infiltration excess threshold significance (in a plot of quickflow volume vs. maximum intensity)	–	
SE_effect	Saturation excess importance	–	
SE_thresh_signif	Saturation excess threshold significance (in a plot of quickflow volume vs. total precipitation)	–	
SE_thresh	Saturation excess threshold location (in a plot of quickflow volume vs. total precipitation)	mm	

SE_slope	Saturation excess threshold above-threshold slope (in a plot of quickflow volume vs. total precipitation)	–	
Storage_thresh	Storage/saturation excess threshold location (in a plot of quickflow volume vs. antecedent precipitation index + total precipitation)	mm	
Storage_thresh_signif	Storage/saturation excess threshold significance (in a plot of quickflow volume vs. antecedent precipitation index + total precipitation)	–	
BaseflowMagnitude	Difference between maximum and minimum of annual baseflow regime	mm	Gnann et al., (2021)
ResponseTime	Catchment response time	day	
FlashinessIndex	Richards-Baker flashiness idex	–	
PQ_Curve	Slopes and breakpoints in cumulative P-Q regime curve	–	
Q_n_day_max	n-day maximum streamflow	mm/day	
Q_skew	Skewness of streamflow	mm 3/ day 3	
Q_var	Variance of streamflow	mm 2/ day 2	
RecessionK_part	Recession constant of early/late (exponential) recessions	1/day	
SeasonalTranslation	Amplitude ratio and phase shift between seasonal forcing and flow cycles	–	
SnowStorage	Snow storage derived from cumulative P-Q regime curve	mm	
zero_Q_duration	Zero flow duration	day	
Q_7_day_min	7-day minimum streamflow	mm/day	
CoV	Coefficient of variation	–	
HFI_mean	Half flow interval	days	
BaseflowRecessionK	Exponential recession constant	1/d	

Appendix B. Comparison of simulated and observed signatures.

The DK-model has been calibrated jointly against observed streamflow at around 300 hydrological stations; the performance of DK-model simulated streamflow is satisfactory at many stations (Stisen et al., 2020). Here, we displayed some comparison
460 results of signatures derived from DK-model simulated streamflow and the observations. The magnitude of streamflow, such as the mean streamflow, between simulated streamflow is aligned with the observations (Fig. B1a). The mean streamflow is lower than 5 m^3/s at many stations (96%) and 18 stations are in large rivers with a daily average streamflow higher than 5 m^3/s. DK-model tends to overestimate the BFI (Fig.B1b), indicating that a slightly smoother streamflow hydrograph simulated by the DK-model. The precipitation elasticity of streamflow, which indicates the sensitivity of streamflow to precipitation, is still
465 challenged to capture accurately by the model (Fig. B1c). The slope of flow duration curves (FDC_slope) quantifies the variability of hydrographs. A steep slope indicates a highly variable stream, where flow is primarily driven by the quick runoff of rainfall to the stream. The simulated hydrography tends to overestimate the FDC_slope for highly variable streams but underestimate it for less variable streams (Fig. B1d).
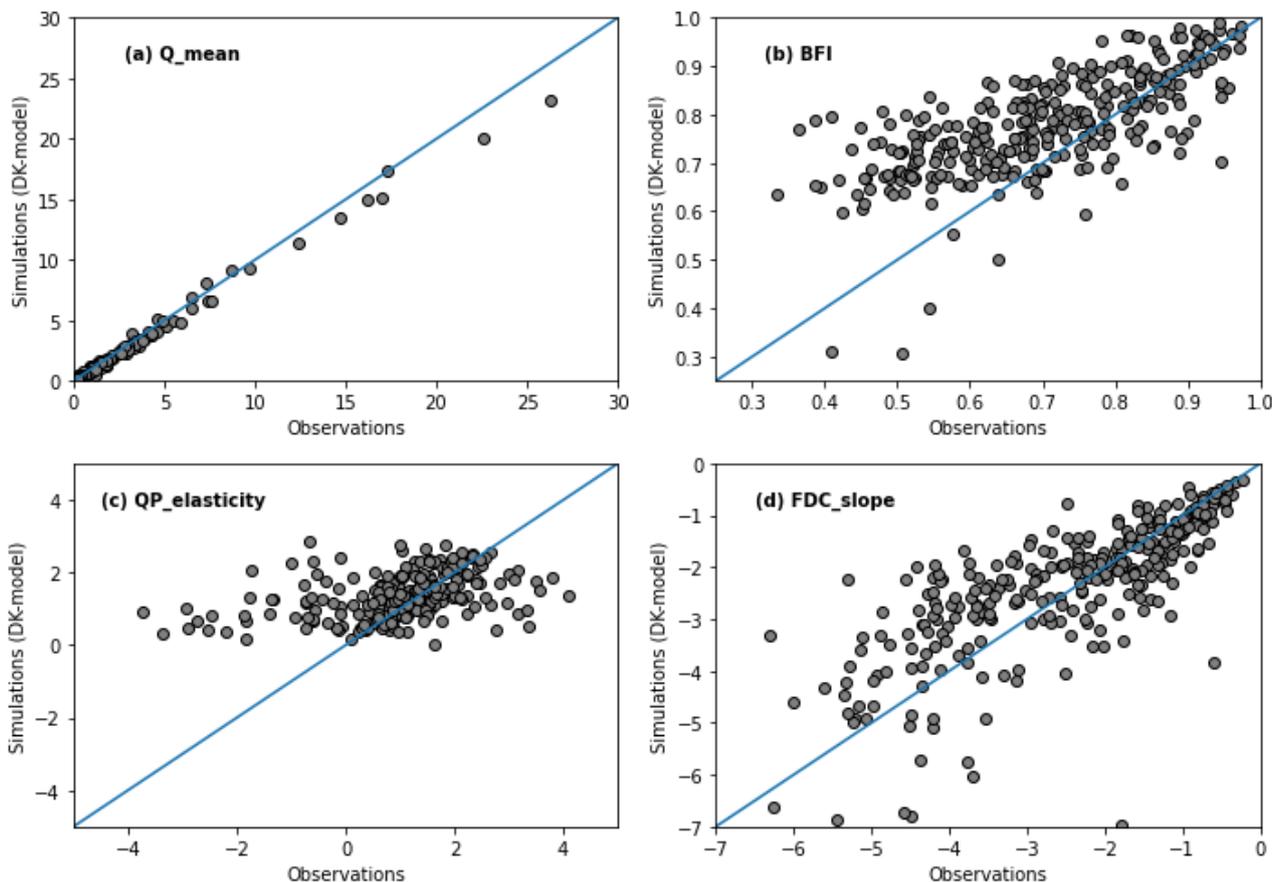
Figure B1. Comparison of observed signatures and DK-model simulated signatures.

Data availability

All the data provided in this study are freely accessible. CAMELS-DK is available at GESU Dataverse https://doi.org/10.22008/FK2/AZXSYP (Koch et al., 2024). The ID15 catchment shapefile is provided with this dataset, previous version, and the newest version (ID15v3) of the shapefile can be requested by writing to id15@ecos.au.dk. High resolution DEM data is available here https://dataforsyningen.dk/data/928. The gridded climate data is created by the Danish Meteorological Institute (DMI) and downloaded from DMI Data API (https://opendatadocs.dmi.govcloud.dk/DMIOpenData). Precipitation has been corrected based on the approach described by Stisen et al. (2011), original data without correction or DMI corrections of precipitation are unavailable in this dataset. The observed streamflow data can be downloaded from the Danish Environmental Protection Agency (https://mst.dk/erhverv). Users should contact for streamflow data from additional stations. Simulation results of surface water and groundwater dynamics from The National Hydrological model of Denmark are available at (https://dennationalehydrologiskemodel.dk/).

Earth System Science Data Discussions

Open Access

Competing interests

485   The authors declare that they have no conflict of interest.

Author contribution

JK, RJMS, and JL conceptualized and formulated the overarching research ideas and collected data from different sources. JL processed the data, conducted the formal analysis, and prepared the original draft under the supervision of RJMS and JK.

490   MFTH assisted with data processing. LT, ALH, and HT provided the ID15 catchment, validated the data sources, and reviewed the manuscript.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.: An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling

495   system, J. Hydrol., 87, 45–59, https://doi.org/https://doi.org/10.1016/0022-1694(86)90114-9, 1986.

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.

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, Hydrol. Sci. J., 65, 712–725,

500   https://doi.org/10.1080/02626667.2019.1683182, 2020.

Alvarez-Garreton, C., Mendoza, P. A., Pablo Boisier, J., 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.

505   Andersson, J. C. M., Pechlivanidis, I. G., Gustafsson, D., Donnelly, C., and Arheimer, B.: Key factors for improving large-scale hydrological model performance, Eur. Water, 49, 77–88, 2015.

Bathelemy, R., Brigode, P., Andréassian, V., Perrin, C., Moron, V., Gaucherel, C., Tric, E., and Boisson, D.: Simbi: historical hydro-meteorological time series and signatures for 24 catchments in Haiti, 1–34, 2023.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., and Soukup, T.: The CORINE land cover 2000 project, EARSeL

510   eProceedings, 3, 331–346, 2004.

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

Clerc-schwarzenbach, F. M., Selleri, G., Neri, M., Toth, E., Meerveld, I. Van, and Seibert, J.: HESS Opinions : A few camels

515   or a whole caravan ?, 1–29, 2024.

Danapour, M., Fienen, M. N., Højberg, A. L., Jensen, K. H., and Stisen, S.: Multi-Constrained Catchment Scale Optimization of Groundwater Abstraction Using Linear Programming., Ground Water, 59, 503–516, https://doi.org/10.1111/gwat.13083, 2021.

Demir, I., Xiang, Z., Demiray, B., and Sit, M.: WaterBench-Iowa: a large-scale benchmark dataset for data-driven streamflow
520 forecasting, Earth Syst. Sci. Data, 14, 5605–5616, https://doi.org/10.5194/essd-14-5605-2022, 2022.

DHI: MIKE SHE User Guide and Reference Manual, 2020.

Duque, C., Nilsson, B., and Engesgaard, P.: Groundwater–surface water interaction in Denmark, Wiley Interdiscip. Rev. Water, 10, 1–23, https://doi.org/10.1002/wat2.1664, 2023.

Espinoza, E. A., Loritz, R., Chaves, M. Á., Bäuerle, N., and Ehret, U.: To Bucket or not to Bucket ? Analyzing the performance
525 and interpretability of hybrid hydrological models with dynamic parameterization, 1–22, 2023.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Frame, J., Kratzert, F., Raney, A., Rahman, M., Salas, F., and Nearing, G.: Post-Processing the National Water Model with
530 Long Short-Term Memory Networks for Streamflow Predictions and Model Diagnostics, JAWRA J. Am. Water Resour. Assoc., 57, 885–905, https://doi.org/10.1111/1752-1688.12964, 2021.

Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J. K., and McMillan, H. K.: TOSSH: A toolbox for streamflow signatures in hydrology, Environ. Model. Softw., 138, 104983, 2021.

Hansen, M. and Pjetursson, B.: Free, online Danish shallow geological data, Geus Bull., 23, 53–56, 2011.

535 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 Syst. Sci. Data, 13, 5591–5616, https://doi.org/10.5194/essd-13-5591-2021, 2021.

Helgason, H. B. and Nijssen, B.: LamaH-Ice: LArge-SaMple Data for Hydrology and Environmental Sciences for Iceland, Earth Syst. Sci. Data, 1–38, 2023.

Henriksen, H. J., Troldborg, L., Højberg, A. L., and Refsgaard, J. C.: Assessment of exploitable groundwater resources of
540 Denmark by use of ensemble resource indicators and a numerical groundwater–surface water model, J. Hydrol., 348, 224–240, https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.09.056, 2008.

Henriksen, H. J., Voutchkova, D., Troldborg, L., Ondracek, M., Schullehner, J., and Hansen, B.: National Vandressource Model National Vandressource Model Beregning af udnyttelsesgrader , afsænkning, 2019.

Henriksen, H. J., Kragh, S. J., Gotfredsen, J., Ondracek, M., van Til, M., Jakobsen, A., Schneider, R. J. M., Koch, J., Troldborg,
545 L., Rasmussen, P., Pasten-Zapata, E., and Stisen, S.: Udvikling af landsdækkende modelberegninger af terrænnære hydrologiske forhold i 100m grid ved anvendelse af DK-modellen: Dokumentationsrapport vedr. modelleverancer til Hydrologisk Informations- og Prognosesystem. Udarbejdet som en del af Den Fællesoffen, GEUS, https://doi.org/10.22008/gpub/38113, 2021.

Henriksen, H. J., Ondracek, M., and Troldborg, L.: Vandressourceopgørelse – datarapport. Baggrundsrapport til

550 Miljøstyrelsens samlede afrapportering omkring forvaltning af fremtidens drikkevandsressource. Metode, resultater, usikkerheder og forventede klimapåvirkninger, GEUS, https://doi.org/10.22008/gpub/34675, 2023.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth Syst. Sci. Data, 15, 5755–5784, 555 https://doi.org/10.5194/essd-15-5755-2023, 2023.

Højberg, A. L., Troldborg, L., Stisen, S., Christensen, B. B. S., and Henriksen, H. J.: Stakeholder driven update and improvement of a national water resources model, Environ. Model. Softw., 40, 202–213, https://doi.org/10.1016/j.envsoft.2012.09.010, 2013.

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T.: Using hydrological and climatic catchment clusters to explore 560 drivers of catchment behavior, Hydrol. Earth Syst. Sci., 24, 1081–1100, https://doi.org/10.5194/hess-24-1081-2020, 2020.

Kibler, K. M., Biswas, R. K., and Lucas, A. M. J.: Hydrologic data as a human right? Equitable access to information as a resource for disaster risk reduction in transboundary river basins, Water Policy, 16, 36–58, https://doi.org/10.2166/wp.2014.307, 2014.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences 565 for Central Europe, Earth Syst. Sci. Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

Koch, J., Gotfredsen, J., Schneider, R., Troldborg, L., Stisen, S., and Henriksen, H. J.: High Resolution Water Table Modeling of the Shallow Groundwater Using a Knowledge-Guided Gradient Boosting Decision Tree Model, Front. Water, 3, 1–14, https://doi.org/10.3389/frwa.2021.701726, 2021.

Koch, J., Liu, J., Stisen, S., Troldborg, L., Højberg, A. L., Thodsen, H., Hansen, M. F. T., and Schneider, R. J. M.: CAMELS-570 DK: Hydrometeorological Time Series and Landscape Attributes for 3330 Catchments in Denmark, https://doi.org/doi:10.22008/FK2/AZXSYP.

Konapala, G., Kao, S. C., Painter, S. L., and Lu, D.: Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US, Environ. Res. Lett., 15, https://doi.org/10.1088/1748-9326/aba927, 2020.

575 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 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 Resour. Res., 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019.

580 Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall-runoff modeling, Hydrol. Earth Syst. Sci., 25, 2685–2703, https://doi.org/10.5194/hess-25-2685-2021, 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, Sci. Data, 10, 1–11,
585 https://doi.org/10.1038/s41597-023-01975-w, 2023.

Levin, G.: Basemap04: Documentation of the data and method for the elaboration of a land use and land cover map for Denmark, 2022.

Liu, J., Koch, J., Stisen, S., Troldborg, L., and Schneider, R. J. M.: A National Scale Hybrid Model for Enhanced Streamflow Estimation–Consolidating a Physically Based Hydrological Model with Long Short-term Memory Networks, Hydrol. Earth
590 Syst. Sci. Discuss., 2023, 1–34, 2023a.

Liu, J., Bian, Y., and Shen, C.: Probing the limit of hydrologic predictability with the Transformer network, ArXiv, abs/2306.1, null, https://doi.org/10.48550/arXiv.2306.12384, 2023b.

Van Loon, A. F.: Hydrological drought explained, Wiley Interdiscip. Rev. Water, 2, 359–392, https://doi.org/10.1002/WAT2.1085, 2015.

595 Ma, K., Feng, D., Lawson, K., Tsai, W., Liang, C., Huang, X., Sharma, A., and Shen, C.: Transferring Hydrologic Data Across Continents – Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions, Water Resour. Res., 57, null, https://doi.org/10.1029/2020WR028600, 2021.

Mahmood, H., Schneider, R. J. M., Frederiksen, R. R., Christiansen, A. V., and Stisen, S.: Using jointly calibrated fine-scale drain models across Denmark to assess the influence of physical variables on spatial drain flow patterns, J. Hydrol. Reg. Stud.,
600 46, 101353, https://doi.org/10.1016/j.ejrh.2023.101353, 2023.

Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., Fortin, V., Fry, L. M., Gauch, M., Klotz, D., Kratzert, F., O'Brien, N., Princz, D. G., Rasiya Koya, S., Roy, T., Seglenieks, F., Shrestha, N. K., Temgoua, A. G. T., Vionnet, V., and Waddell, J. W.: The Great Lakes Runoff Intercomparison Project Phase 4: The Great Lakes (GRIP-GL), Hydrol. Earth Syst. Sci., 26, 3537–3572, https://doi.org/10.5194/hess-26-3537-2022, 2022.

605 Martinsen, G., Bessiere, H., Caballero, Y., Koch, J., Collados-Lara, A. J., Mansour, M., Sallasmaa, O., Pulido-Velazquez, D., Williams, N. H., Zaadnoordijk, W. J., and Stisen, S.: Developing a pan-European high-resolution groundwater recharge map – Combining satellite data and national survey data using machine learning, Sci. Total Environ., 822, 153464, https://doi.org/10.1016/j.scitotenv.2022.153464, 2022.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: A review, Hydrol. Process., 34, 1393–1409, 2020.

610 McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, Wiley Interdiscip. Rev. Water, 5, 1–14, https://doi.org/10.1002/WAT2.1319, 2018.

Meyer Oliveira, A., van Meerveld, H. J., Vis, M., and Seibert, J.: Assessment of the Value of Remotely Sensed Surface Water Extent Data for the Calibration of a Lumped Hydrological Model, Water Resour. Res., 59, 1–19, https://doi.org/10.1029/2023WR034875, 2023.

615 Møller, A. B., Beucher, A., Iversen, B. V, and Greve, M. H.: Predicting artificially drained areas by means of a selective model ensemble, Geoderma, 320, 30–42, 2018.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo,

S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., and Matias, Y.: Global prediction of extreme floods in ungauged watersheds, Nature, 627, 559–563, https://doi.org/10.1038/s41586-024-07145-1, 2024.

620 Pedersen, S. A. S., Hermansen, B., Nathan, C., and Tougaard, L.: Digitalt kort over Danmarks jordarter 1:200.000, version 2. Geologisk kort over de overfladenære jordarter i Danmark, GEUS, https://doi.org/10.22008/gpub/28464, 2011.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

625 Scharling, M.: Klimagrid Danmark - Nedbør, lufttemperatur og potentiel fordampning 20X20 & 40x40 km - Metodebeskrivelse, Danish Meteorol. Inst., 1999a.

Scharling, M.: Klimagrid Danmark Nedbør 10x10 km (ver. 2) - Metodebeskrivelse, Danish Meteorol. Inst., 15–17, 1999b.

Schneider, R., Koch, J., Troldborg, L., Henriksen, H. J., and Stisen, S.: Machine-learning-based downscaling of modelled climate change impacts on groundwater table depth, Hydrol. Earth Syst. Sci., 26, 5859–5877, https://doi.org/10.5194/hess-26-5859-2022, 2022.

630 Seidenfaden, I. K., Sonnenborg, T. O., Stisen, S., and Kidmose, J.: Quantification of climate change sensitivity of shallow and deep groundwater in Denmark, J. Hydrol. Reg. Stud., 41, 101100, https://doi.org/10.1016/j.ejrh.2022.101100, 2022.

Singh, R., Archfield, S. A., and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach, J. Hydrol., 517, 985–996, https://doi.org/10.1016/j.jhydrol.2014.06.030, 2014.

635 Soltani, M., Bjerre, E., Koch, J., and Stisen, S.: Integrating remote sensing data in optimization of a national water resources model to improve the spatial pattern performance of evapotranspiration, J. Hydrol., 603, 127026, https://doi.org/10.1016/j.jhydrol.2021.127026, 2021.

Stisen, S., Sonnenborg, T. O., Højberg, A. L., Troldborg, L., and Refsgaard, J. C.: Evaluation of Climate Input Biases and 640 Water Balance Issues Using a Coupled Surface-Subsurface Model, Vadose Zo. J., 10, 37–53, https://doi.org/10.2136/vzj2010.0001, 2011.

Stisen, S., Ondracek, M., Troldborg, L., Schneider, R. J. M., and Til, M. J. van: National Vandressource Model. Modelopstilling og kalibrering af DK-model 2019, GEUS, Copenhagen, https://doi.org/10.22008/gpub/32631, 2020.

Tang, S., Sun, F., Liu, W., Wang, H., Feng, Y., and Li, Z.: Optimal Postprocessing Strategies With LSTM for Global 645 Streamflow Prediction in Ungauged Basins, Water Resour. Res., 59, 1–16, https://doi.org/10.1029/2022WR034352, 2023.

Tegegn, Z., Abebe, A., and Agide, Z.: Understanding Catchments' Hydrologic Response Similarity of Upper Blue Nile (Abay) basin through catchment classification, Model. Earth Syst. Environ., 8, 3305–3323, https://doi.org/10.1007/s40808-021-01298-y, 2022.

Tshimanga, R. M., Bola, G. B., Kabuya, P. M., Nkaba, L., Neal, J., Hawker, L., Trigg, M. A., Bates, P. D., Hughes, D. A., and 650 Laraque, A.: Towards a framework of catchment classification for hydrologic predictions and water resources management in the ungauged basin of the Congo River: An a priori approach, Congo Basin Hydrol. Clim. Biogeochem. A Found. Futur., 469–

498, 2022.

Wilbrand, K., Taormina, R., ten Veldhuis, M.-C., Visser, M., Hrachowitz, M., Nuttall, J., and Dahm, R.: Predicting streamflow with LSTM networks using global datasets, Front. Water, 5, 1166124, 2023.

655   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,

660   J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: Comment: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 1–9, https://doi.org/10.1038/sdata.2016.18, 2016.

Yin, H., Guo, Z., Zhang, X., Chen, J., and Zhang, Y.: RR-Former: Rainfall-runoff modeling based on Transformer, J. Hydrol., 609, 127781, https://doi.org/10.1016/j.jhydrol.2022.127781, 2022.

665   CLC: CORINE land cover, https://land.copernicus.eu/en/products/ corine-land-cover, last access: 20 February 2024.