CAMELS-DE: hydro-meteorological time series and attributes for 1582 catchments in Germany

Alexander Dolich*[1], Eduardo Acuña Espinoza[1,A], Pia Ebeling[2,A], Björn Guse[3,4,A], Jonas Götte[5,6,7,A], Sibylle K. Hassler[8,A], Corina Hauffe[9,A], Ingo Heidbüchel[2,10,A], Jens Kiesel[3,11,A], Mirko Mälicke[1,A], Hannes Müller-Thomy[12,A] , Michael Stölzle[13,A], Larisa Tarasova[14,A], Ralf Loritz[*1]

*equal contribution, A alphabetic order

[1]Karlsruhe Institute of Technology (KIT), Institute for Water and Environment, Karlsruhe, Germany

[2]Helmholtz Centre for Environmental Research - UFZ, Department Hydrogeology, Leipzig, Germany

[3]Kiel University, Hydrology and Water Resources Management, Kiel, Germany

[4]German Research Centre for Geosciences - GFZ Potsdam, Section Hydrology, Potsdam, Germany

[5]WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland

[6]Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf

[7]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

[8]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe, Germany

[9]University of Technology Dresden (TUD), Institute of Hydrology and Meteorology, Dresden, Germany

[10]Bayreuth Centre of Ecology and Environmental Research, University of Bayreuth, Bayreuth, Germany

[11]Stone Environmental, 535 Stone Cutters Way, 05602 Montpelier (VT), USA

[12]Technische Universität Braunschweig, Leichtweiß-Institute for Hydraulic Engineering and Water Resources, Division of Hydrology and River Basin Management, Braunschweig, Germany

[13]Chair of Hydrology, University of Freiburg, Freiburg, Germany, now at: LUBW Landesanstalt für Umwelt (State Agency for Environment), Karlsruhe, Germany

[14]Helmholtz Centre for Environmental Research - UFZ, Department Catchment Hydrology, Germany

Correspondence to: Alexander Dolich (Alexander.Dolich@kit.edu) and Ralf Loritz (Ralf.Loritz@kit.edu)

1 Licence

2 Citation

When using the data please cite:

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

The data are supplementary material to

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 1555 catchments in Germany, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2024-318, in review, 2024.

Please refer to the published paper once its review process is finished.

3 Data Description

CAMELS-DE provides a comprehensive collection of hydro-meteorological time series data and catchment attributes for 1582 streamflow gauges across Germany (Fig. 1a). The time series data are in daily resolution and span up to 70 years, from January 1951 to December 2020. The static catchment attributes include information about topography, soils, land cover, hydrogeology and human influences. Additionally, the dataset includes discharge simulations from a regional Long-Short Term Memory (LSTM) network and a locally trained conceptual HBV model, providing benchmark data for future hydrological modelling studies in Germany. This dataset was developed to address the need for a harmonised, nation-wide hydrological dataset that integrates data from Germany's extensive and decentralised measurement infrastructure. By providing consistent and high-quality data, CAMELS-DE aims to facilitate large-scale hydrological studies and advance research in areas such as quantification of long-term trends, climate change impacts, and data-driven hydrology.

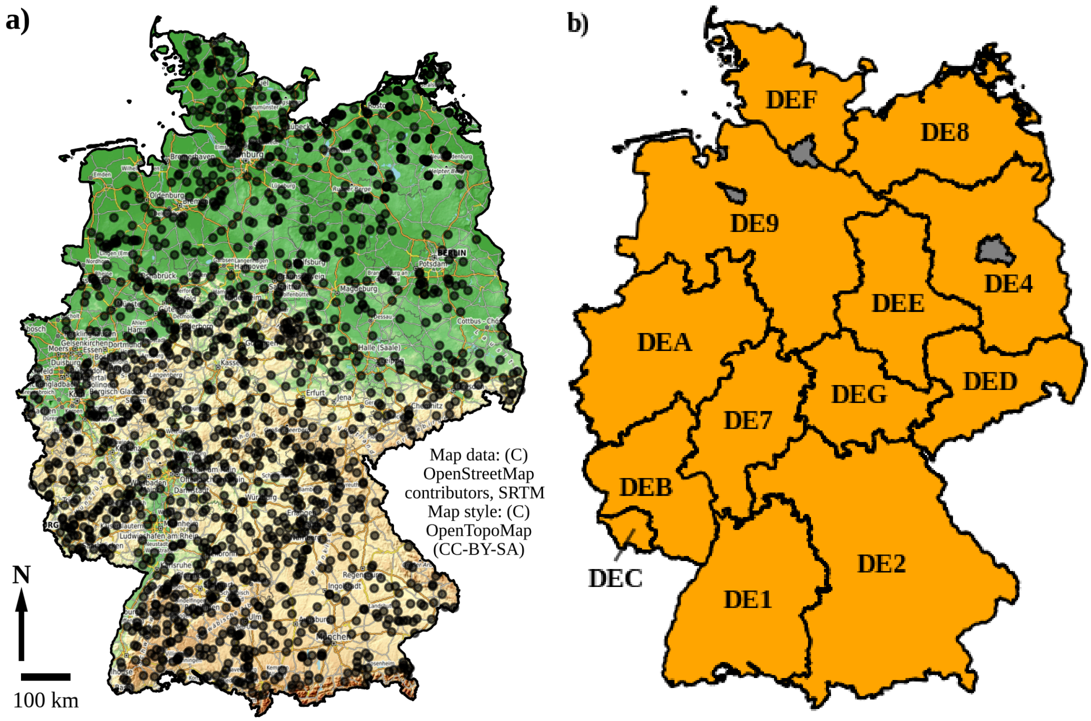

Figure 1: Panel (**a**) shows a topographic map of Germany with all 1582 gauging stations of CAMELS-DE. Panel (**b**) shows a map of Germany with the borders of the federal states and their NUTS Level 1 IDs, which are used to generate IDs for the gauging stations (e.g. DE110000, DE110010, …), borders of Germany: © GeoBasis-DE / BKG (2024)

3.1 Purpose and Method of Data Collection

The CAMELS-DE dataset was collected to support hydrological research and education by offering a reliable, large-sample dataset that spans multiple decades and diverse catchments across Germany. Data collection involved the integration of hydro-meteorological observations and catchment attributes from various federal and state agencies as well as from third party data providers.

Hydrological data, including daily measurements of discharge (m³/s) and water levels (m), as well as gauge metadata were sourced from thirteen German federal state agencies (by directly contacting them), namely the Landesanstalt für Umwelt Baden-Württemberg (LUBW, Nomenclature of Territorial Units for Statistics (NUTS) Level 1: DE1), Bayerisches Landesamt für Umwelt (LfU-Bayern, DE2), Landesamt für Umwelt Brandenburg (LfU-Brandenburg, DE4), Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG, DE7; Data provision by the HLNUG in accordance with the Hessian Environmental Information Act (HUIG)), Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern (LUNG MV, DE8), Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, Landesamt für Natur (NLWKN, DE9), Umwelt und Verbraucherschutz Nordrhein-Westfalen (LANUV NRW, DEA), Landesamt für Umwelt Rheinland-Pfalz (LUA-Rheinland Pfalz, DEB), Landesamt für Umwelt- und Arbeitsschutz

Saarland (LUA, DEC), Landesamt für Umwelt, Landwirtschaft und Geologie Sachsen (LfULG, DED), Landesamt für Umweltschutz Sachsen-Anhalt (LAU, DEE), Landesamt für Landwirtschaft, Umwelt und ländliche Räume Schleswig-Holstein (LLUR, DEF), and Thüringer Landesamt für Umwelt, Bergbau und Naturschutz (TLUBN, DEG). The state agencies do not guarantee the accuracy or completeness of the data. Additionally, all hydrological data may be subject to future revisions, including adjustments to rating curves or corrections of errors. Therefore, it is necessary to obtain the most recent discharge time series directly from the federal state agencies for projects that require water law permits. The regulations of the respective federal state apply, and specific inquiries should be made as needed. It is also important to note that the state agencies explicitly disclaim any guarantees regarding data accuracy or completeness, thus excluding any liability claims against any of the federal states. Hydrological catchment boundaries were derived using the MERIT Hydro dataset (Yamazaki et al., 2019) and controlled by comparing the derived catchment areas with those reported by the state authorities. Meteorological data, specifically precipitation, temperature, relative humidity, and radiation, were obtained from the German Weather Service (DWD) from the HYRAS dataset (Deutscher Wetterdienst, HYRAS, 2024). Spatially aggregated catchment attributes were obtained from various sources. Topographic catchment attributes were derived from the Copernicus GLO-30 DEM (EU-DEM, Copernicus, 2022), the CORINE Land Cover 2018 dataset (CLC, 2018) was used to extract information about land cover. Soil attributes were derived from the global SoilGrids250m dataset generated at the International Soil Reference and Information Centre (ISRIC; Poggio et al., 2021). Hydrogeological catchment attributes were derived from the "Hydrogeologische Übersichtskarte von Deutschland 1:250.000 (HÜK250)" provided by the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) while information about human influences, e.g., dams or weirs, was sourced from Speckhann et al. (2021).

3.2 Data Structure

CAMELS-DE data is a series of individual csv files for the hydro-meteorological time series, the catchment attributes as well as the results of the LSTM and the HBV model runs. Shapefiles and geopackage files provide information about the catchment boundaries and the location of streamflow gauges. Additionally, HBV model parameters and the LSTM training epochs are included. The files are combined into a zip folder structured to provide easy access and usability for researchers. The structure and file names are similar to other already published CAMELS datasets.

3.3 Data Quality

To ensure the accuracy and reliability of the dataset, the following measures were implemented:

- Quality controlled discharge and water level data was provided by the federal agencies.
- Integration of high quality, already quality controlled data that build a harmonised data source for the derivation of catchment attributes.

- Data conversion and transformation to harmonise measurements from different sources.
- Manual detection and correction of clear errors and inconsistencies in the data.
- The dataset has been tested with an LSTM (Long short-term memory) network and a conceptual hydrological model in order to identify catchments where the relationship between meteorological forcing and streamflow is difficult to capture, indicating possible strong human influences such as dams or reservoirs, or potential issues with the catchment delineation or the streamflow or meteorological time series. Results of this model run are also part of the CAMELS-DE dataset.

4 Data processing

The processing pipeline for the CAMELS-DE dataset is structured to ensure clarity, reproducibility, and modularity. The initial phase of the pipeline involves processing and harmonising raw discharge and water level data, along with station metadata provided by the federal states. Following this, MERIT-Hydro catchment boundaries are delineated for each station with the delineator.py package (Heberger, 2023), a critical step as all further datasets rely heavily on these boundaries. Meteorological time series data for these catchments are then processed to compute statistics such as area mean and median. Subsequently, catchment attributes including soil properties, hydrogeology, land cover, topography, and human influences are derived for each catchment. In the final stage, all derived data are integrated and formatted according to the established structure of the CAMELS datasets, mirroring the organisational schema of CAMELS-GB (Coxon et al., 2020) or CAMELS-CH (Höge et al., 2023).

The source code repositories for the CAMELS-DE project are accessible through the CAMELS-DE GitHub organisation at https://github.com/camels-DE/. The CAMELS-DE processing pipeline was also published separately with some more details and permalinks to released versions that clearly represent the code state that was used to process CAMELS-DE (Dolich, 2024). The processing pipeline was structured in a modular manner to enhance its clarity and reproducibility. For each component of the CAMELS-DE data, a separate GitHub repository was established. Within each repository, a dedicated Docker container processes specific input datasets (e.g., DWD HYRAS, EU-DEM). Docker is particularly well-suited for the processing pipeline as it ensures that each component of the data processing pipeline runs consistently across different computing environments. This containerization simplifies dependency management, enhances reproducibility, and facilitates the deployment and version control of each processing module. Figure 2 illustrates the architecture of the processing pipeline, where each blue block represents an individual GitHub repository with a Docker container that processes the yellow input data to produce the green output data. All repositories are uniformly structured, and the accompanying documentation provides detailed descriptions of each repository, guidelines for building and running the Docker containers, including the necessary folder mounts, and instructions for accessing the required input data.

CAMELS-DE, in addition to hydro-meteorological observations and catchment attributes, includes rainfall-runoff simulations for each catchment from a regionally trained LSTM network and a locally trained lumped HBV model. Both models were trained over the period from October 1, 1970, to December 31, 1999, validated from October 1, 1965, to September 30, 1970, and tested from January 1, 2000, to December 31, 2020. CAMELS-DE includes the simulated discharges for both models for the entire 70 years as well as model parameters for the purpose of reproducibility. The code of the LSTM model and the HBV model were carefully tested and benchmarked (Acuña Espinoza et al., 2024) and are available here (https://github.com/KIT-HYD/Hy2DL/tree/v1.1).

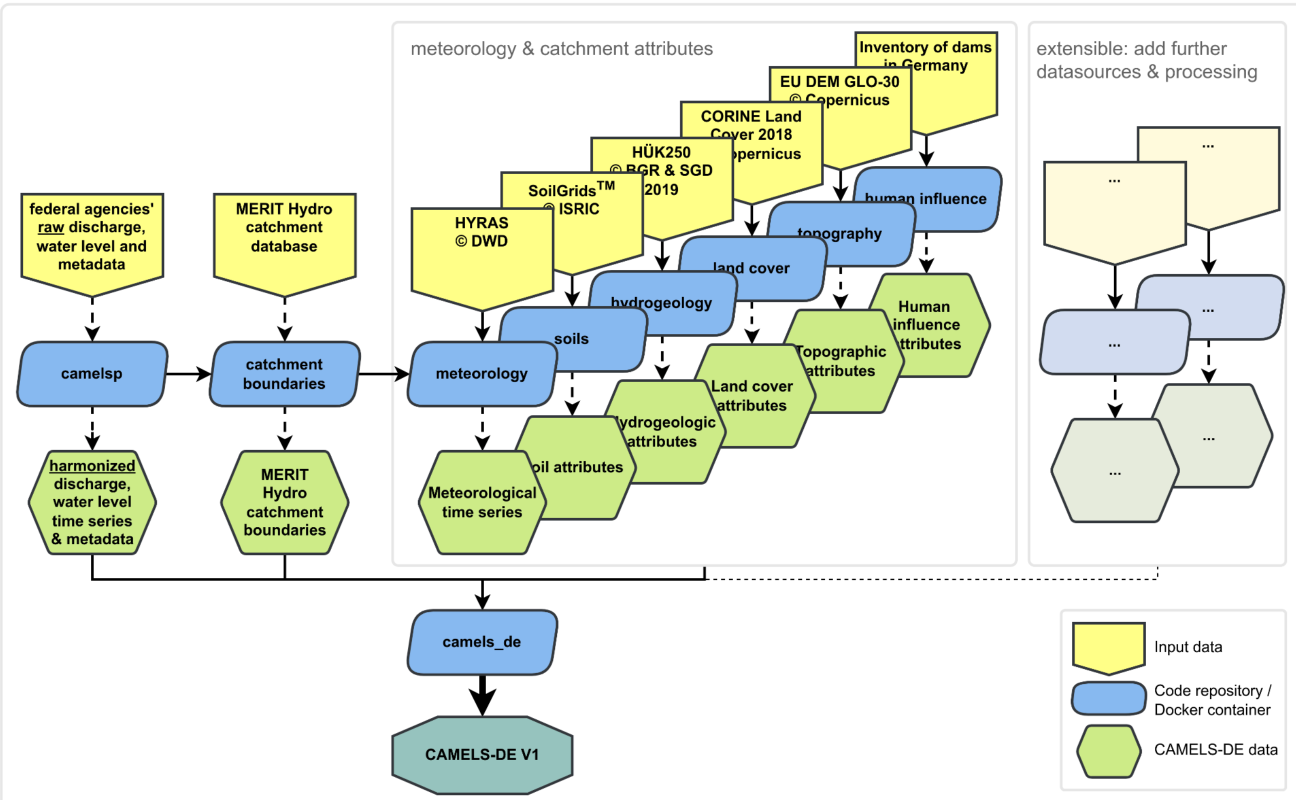

Figure 2: Diagram of the CAMELS-DE data processing pipeline. Starting with raw discharge and metadata harmonisation using the camelsp package, it proceeds to derive MERIT-Hydro catchment boundaries. Subsequent processing includes meteorological data and extraction of various catchment attributes, all integrated into the camels_de repository, culminating in the structured CAMELS-DE dataset consistent with CAMELS-GB or CAMELS-CH.

5 Data Format

5.1 Hydro-meteorological time series files

Hydro-meteorological time series are provided in .csv format and are contained in a folder timeseries, CAMELS-DE includes data for 1582 catchments. All time series files range from 1951-01-01 until 2020-12-3l, with potentially missing values in the discharge and water level time series; the meteorological time series do not have missing values.
The naming convention of the files include the generated gauge ID, which was generated for CAMELS-DE based on the NUTS classification (Nomenclature of territorial units for statistics, Fig. 1b):

CAMELS_DE_hydromet_timeseries_<gauge_id>.csv

Refer to Table 1 for information about the actual structure of the .csv files, variable names and units and the data sources.

5.2 Simulated hydrological time series files

Simulated hydrological time series are provided as .csv files and are located in a folder timeseries_simulated. The simulated hydrological time series range from 1951-01-01 until 2020-12-3l and are based on a regionally trained LSTM network (trained on all catchments at the same time) and a locally trained lumped HBV model (trained at each individual catchment; Bergström and Forsman, 1973, Seibert, 2005, Feng et al., 2022). For the HBV model, a time series of the potential evapotranspiration was also calculated using the Hargreaves equation, which is included here alongside the observed and simulated specific discharge time series and the volumetric discharge time series calculated from the simulated specific discharge and the catchment area. An additional column indicates the simulation period in which the daily value is located (training: 1970-10-01 – 1999-12-31, validation: 1965-10-01 – 1970-09-30, testing: 2001-10-01 – 2020-12-31). If the results are to serve as a benchmark for other hydrological modelling, only the training period should be used. The entire time series can be used, for example, to fill gaps in the observed discharge values, although the model performance (Nash-Sutcliffe Efficiency, NSE) for the respective gauging station should be taken into account.
The naming convention is similar to the naming of the hydro-meteorological time series files:

CAMELS_DE_discharge_sim_<gauge_id>.csv

Table 2 gives more information about the structure of the .csv files, variable names and the units.

Table 1: Catchment-specific hydro-meteorological variables available as daily time series in CAMELS-DE (CAMELS_DE_hydromet_timeseries_<gauge_id>.csv)

Column header	Description	Unit	Data source
discharge_vol	Catchment discharge calculated from the water level and gauge geometry	$m^3 s^{-1}$	Federal state agencies (see section 3.1)
discharge_spec	Observed catchment-specific discharge (converted to millimetres per day using catchment areas described in section. 3.1)	$mm d^{-1}$	
water_level	Observed daily water level	m	
precipitation_mean, precipitation_median, precipitation_min, precipitation_max, precipitation_stdev	Observed interpolated spatial mean, median, minimum, maximum and standard deviation of the daily precipitation	$mm d^{-1}$	German Weather Service HYRAS (DWD-HYRAS, 2024)
temperature_min	Observed interpolated spatial mean daily minimum temperatures	°C	
temperature_mean	Observed interpolated spatial mean daily mean temperatures	°C	
temperature_max	Observed interpolated spatial mean daily maximum temperatures	°C	
humidity_mean, humidity_median, humidity_min, humidity_max, humidity_stdev	Observed interpolated spatial mean, median, minimum, maximum and standard deviation of the daily humidity	%	
radiation_global_mean, radiation_global_median, radiation_global_min, radiation_global_max, radiation_global_stdev	Observed interpolated spatial mean, median, minimum, maximum and standard deviation of the global radiation	$W m^2$	

Table 2: Catchment-specific simulated hydro-meteorological variables available as daily time series in CAMELS-DE (CAMELS_DE_discharge_sim_<gauge_id>.csv)

Column header	Description	Unit	Data source
pet_hargreaves	Daily mean of potential evapotranspiration calculated using the Hargreaves equation	mm d^{-1}	Regional LSTM model, HBV model and Hargreaves equation for potential evapotranspiration (see section 6, https://github.com/KIT-HYD/Hy2DL/tree/v1.1, last access: 24 July 2024)
discharge_vol_obs	Observed volumetric discharge	m^3 s^{-1}	
discharge_spec_obs	Observed catchment-specific discharge	mm d^{-1}	
discharge_vol_sim_lstm	Volumetric discharge calculated from discharge_spec_sim_lstm and the catchment area	m^3 s^{-1}	
discharge_spec_sim_lstm	Catchment-specific discharge simulated with the LSTM (see section 6)	mm d^{-1}	
discharge_vol_sim_hbv	Volumetric discharge calculated from discharge_spec_sim_hbv and the catchment area	m^3 s^{-1}	
discharge_spec_sim_hbv	Catchment-specific discharge simulated with the HBV model (see section 6)	mm d^{-1}	
simulation_period (training, validation, testing)	Flag indicating the simulation period in which the daily value is contained (training, validation, testing)	–	

5.2 Catchment attributes

CAMELS-DE includes one .csv file for each catchment attribute class, in total seven files (topographic, climatic, hydrologic, soils, land cover, hydrogeology, human influences) and one more .csv file containing information about the LSTM and HBV model results.

CAMELS_DE_topographic_attributes.csv

CAMELS_DE_climatic_attributes.csv

CAMELS_DE_hydrologic_attributes.csv

CAMELS_DE_soil_attributes.csv

CAMELS_DE_landcover_attributes.csv

CAMELS_DE_hydrogeology_attributes.csv

CAMELS_DE_humaninfluence_attributes.csv

CAMELS_DE_simulation_benchmark.csv

All .csv files contain the CAMELS-DE gauge ID as the first column, followed by statistics derived from the different catchment attribute data sources. Tables 3-10 give information about contents of the individual catchment attribute .csv files.

Table 3: Topographic attributes in CAMELS-DE (CAMELS_DE_topographic_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	Federal state agencies (see section 3.1)
provider_id	official gauging station ID assigned by the federal states	–	
gauge_name	gauging station name		
water_body_name	water body name	–	
federal_state	federal state in which the measuring station is located		
gauge_lon	gauging station longitude (EPSG:4326)	°	
gauge_lat	gauging station latitude (EPSG:4326)	°	
gauge_easting	gauging station easting (EPSG:3035)	m	
gauge_northing	gauging station northing (EPSG:3035)	m	
gauge_elev_metadata	gauging station elevation as given by the federal states	m.a.s.l.	
area_metadata	catchment area as given by the federal states	km^2	
gauge_elev	gauging station elevation derived from the EU-DEM	m a.s.l.	Copernicus GLO-30 DEM (EU-DEM, Copernicus, 2022)
area	catchment area derived from the MERIT Hydro catchment	km^2	
elev_mean	mean elevation in the catchment based on the MERIT Hydro geometry	m a.s.l.	
elev_min	minimum elevation within catchment	m a.s.l.	
elev_5	5th percentile elevation within catchment	m a.s.l.	
elev_50	median elevation within catchment	m a.s.l.	
elev_95	95th percentile elevation within catchment	m a.s.l.	
elev_max	maximum elevation within catchment	m a.s.l.	

Table 4: Climatic attributes in CAMELS-DE (CAMELS_DE_climatic_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	German Weather Service HYRAS (DWD-HYRAS, 2024)
p_mean	long-term mean of daily precipitation from 1951 to 2020	mm d^{-1}	
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).	–	
frac_snow	fraction of precipitation falling as snow, i.e. while mean air temperature is < 0° C	–	
high_prec_freq	frequency of high-precipitation days (≥ 5 times mean daily precipitation)	d yr^{-1}	
high_prec_dur	mean 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, e.g. 'jja' for summer. If two seasons register the same number of events a value of NA is given.	season	
low_prec_freq	frequency of dry days (< 1 mm d^{-1})	d yr^{-1}	
low_prec_dur	mean duration of dry periods (number of consecutive days < 1 mm d^{-1} mean daily precipitation)	d	
low_prec_timing	season during which most dry season days occur, e.g. 'son' for autumn. If two seasons register the same number of events a value of NA is given.	season	

Table 5: Hydrologic attributes in CAMELS-DE (CAMELS_DE_hydrologic_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	Federal state agencies (see section 3.1) and German Weather Service HYRAS (DWD-HYRAS, 2024)
q_mean	mean daily specific discharge	mm d^{-1}	
runoff_ratio	runoff ratio (ratio of mean daily discharge to mean daily precipitation)	–	
flow_period_start	first date for which daily streamflow data is available	–	
flow_period_end	last day for which daily streamflow data is available	–	
flow_perc_complete	percentage of days for which streamflow data is available from Jan 1951–31 Dec 2020	%	

slope_fdc	slope of the flow duration curve (between the log-transformed 33rd and 66th stream flow percentiles, see Coxon et al. (2020)	–	
hfd_mean	mean half-flow date (number of days since 1. Oct at which the cumulative dis charge reaches half of the annual discharge)	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 yr^{-1}	
high_q_dur	mean 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 yr^{-1}	
low_q_dur	mean duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	d	
zero_q_freq	fraction of days with zero stream flow	–	

Table 6: Soil attributes in CAMELS-DE (CAMELS_DE_soil_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	SoilGrids250m (Poggio et al., 2021)
clay_0_30cm_mean clay_30_100cm_mean clay_100_200cm_mean	Proportion of clay particles (< 0.002 mm) in the fine earth fraction at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	wt. %	
silt_0_30cm_mean silt_30_100cm_mean silt_100_200cm_mean	Proportion of silt particles (≥ 0.002 mm and ≤ 0.05/0.063 mm) in the fine earth fraction at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	wt. %	
sand_0_30cm_mean sand_30_100cm_mean sand_100_200cm_mean	Proportion of sand particles (> 0.05/0.063 mm) at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	wt. %	
coarse_fragments_0_30cm_mean coarse_fragments_30_100cm_mean coarse_fragments_100_200cm_mean	Volumetric fraction of coarse fragments (> 2 mm) at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	vol %	
soil_organic_carbon_0_30cm_mean soil_organic_carbon_30_100cm_mean soil_organic_carbon_100_200cm_mean	Soil organic carbon content in the fine earth fraction at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	g kg^{-1}	

bulk_density_0_30cm_mean bulk_density_30_100cm_mean bulk_density_100_200cm_mean	Bulk density of the fine earth fraction at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	kg dm^{-3}	

Table 7: Land cover attributes in CAMELS-DE (CAMELS_DE_landcover_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	CORINE Land Cover 2018 (CLC, 2018)
artificial_surfaces_perc	areal coverage of artificial surfaces	%	
agricultural_areas_perc	areal coverage of agricultural areas	%	
forests_and_seminatural_areas_perc	areal coverage of forests and semi-natural areas	%	
wetlands_perc	areal coverage of wetlands	%	
water_bodies_perc	areal coverage of water bodies	%	

Table 8: Hydrogeology attributes in CAMELS-DE (CAMELS_DE_hydrogeology_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	HÜK250 © BGR & SGD 2019
aquitard_perc aquifer_perc aquifer_aquitard_mixed_perc	areal coverage of aquifer media type classes	%	
kf_very_high_perc (>1E-2 m s^{-1}) kf_high_perc (>1E-3 – 1E-2 m s^{-1}) kf_medium_perc (>1E-4 – 1E-3 m s^{-1}) kf_moderate_perc ((>1E-5 – 1E-4 m s^{-1}) kf_low_perc (>1E-7 – 1E-5 m s^{-1}) kf_very_low_perc (>1E-9 - 1E-7 m s^{-1}) kf_extremely_low_perc (<1E-9 m s^{-1}) kf_very_high_to_high_perc (>1E-3 m s^{-1}) kf_medium_to_moderate_perc (>1E-5 – 1E-3 m s^{-1}) kf_low_to_extremely_low_per	areal coverage of permeability classes	%	

c (<1E-5 m s^{-1}) kf_highly_variable_perc kf_moderate_to_low_perc (>1E-6 – 1E-4 m s^{-1})			
cavity_fissure_perc cavity_pores_perc cavity_fissure_karst_perc cavity_fissure_pores_perc	areal coverage of cavity type classes	%	
consolidation_solid_rock_perc consolidation_unconsolidated_rock_perc	areal coverage of consolidation classes	%	
rocktype_sediment_perc rocktype_metamorphite_perc rocktype_magmatite_perc	areal coverage of rock type classes	%	
geochemical_rocktype_silicate_perc geochemical_rocktype_silicate_carbonatic_perc geochemical_rocktype_carbonatic_perc geochemical_rocktype_sulfatic_perc geochemical_rocktype_silicate_organic_components_perc geochemical_rocktype_anthropogenically_modified_through_filling_perc geochemical_rocktype_sulfatic_halitic_perc geochemical_rocktype_halitic_perc	areal coverage of geochemical rock type classes	%	
waterbody_perc	areal coverage of water body areas according to hydrogeological map	%	
no_data_perc	percentage of areas with missing data	%	

Table 9: Human influence attributes in CAMELS-DE (CAMELS_DE_humaninfluence_attributes.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	Inventory of dams in Germany (Speckhann et al., 2021)
dams_names	names of all dams located in the catchment	–	
dams_river_names	names of the rivers where the dams are located	–	
dams_num	number of dams located in the catchment	–	
dams_year_first	year when the first dam entered operation	–	

dams_year_last	year when the last dam entered operation	–	
dams_total_lake_area	total area of all dam lakes at full capacity	km^2	
dams_total_lake_volume	total volume of all dam lakes at full capacity	Mio m^3	
dams_purposes	purposes of all the dams in the catchment	–	

Table 10: Simulation benchmark results in CAMELS-DE (CAMELS_DE_simulation_benchmark.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	Regional LSTM model, HBV model (see CAMELS-DE paper, https://github.com/KIT-HYD/Hy2DL/tree/v1.1, last access: 24 July 2024)
training_perc_complete	percentage of observed specific discharge values in the training period (1970-10-01 – 1999-12-31) that are not NaN	%	
validation_perc_complete	percentage of observed specific discharge values in the validation period (1965-10-01 – 1970-09-30) that are not NaN	%	
testing_perc_complete	percentage of observed specific discharge values in the testing period (2001-10-01 – 2020-12-31) that are not NaN	%	
NSE_lstm	Nash-Sutcliffe model efficiency coefficient of the LSTM in the testing period	–	
NSE_hbv	Nash-Sutcliffe model efficiency coefficient of the hbv model in the testing period	–	

4.3 Catchment boundaries and station locations

CAMELS-DE also provides geospatial data with detailed information about the catchment boundaries that were used for extracting the meteorological data and catchment attributes and which were derived from MERIT Hydro (Yamazaki et al., 2019, Heberger, 2023). The gauge locations are also available as geospatial data. Geospatial file format for catchment boundaries and gauge locations are shapefiles and geopackages:

CAMELS_DE_catchments.shp

CAMELS_DE_catchments.gpkg

CAMELS_DE_gauging_stations.shp

CAMELS_DE_gauging_stations.gpkg

The projection of the files is ETRS89-extended / LAEA Europe (EPSG:3035).

4.4 LSTM and HBV model parameters

In addition to the model results, CAMELS-DE also provides the HBV model parameters for each catchment as well as the model training epochs of the regional LSTM. The HBV model parameters are provided as a .csv file, the contents of which are described in Table 11. The 20 LSTM training epochs are included in the zip file CAMELS_DE_epochs_training_lstm.zip, the epochs can be loaded with pytorch or the LSTM network used for the simulations in CAMELS-DE (https://github.com/KIT-HYD/Hy2DL/tree/v1.1, last access: 24 July 2024).

Table 11: Model parameters of the conceptual HBV model in CAMELS-DE (CAMELS_DE_parameters_hbv.csv)

Column header	Description	Unit	Data source
gauge_id	catchment identifier based on the NUTS classification, e.g. DE110000, DE110010, …	–	HBV model (see CAMELS-DE paper, https://github.com/KIT-HYD/Hy2DL/tree/v1.1, last access: 24 July 2024)
BETA	parameter that determines the relative contribution to runoff from rain or snowmelt	–	
FC	maximum soil moisture storage	mm	
K0, K1, K2	storage (or recession) coefficients	day^{-1}	
LP	soil moisture value above which ET_{act} reaches ET_{pot}	mm	
PERC	max. percolation to lower zone	mm day^{-1}	
UZL	threshold parameter response routine	mm	
TT	threshold temperature snow routing	°C	
CFMAX	degree-day factor	mm $°C^{-1}$ day^{-1}	
CFR	refreezing coefficient	–	
CWH	maximum melt	–	
alpha	Unit hydrograph routing alpha gamma function	–	
beta	Unit hydrograph routing beta gamma function	–	

References

Acuña Espinoza, E., Loritz, R., Álvarez Chaves, M., Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, Hydrology and Earth System Sciences, 28, 2705–2719, https://doi.org/10.5194/hess-28-2705-2024, 2024.

Bergström, S. and Forsman, A.: Development of a Conceptual Deterministic Rainfall-runoff Model, Hydrology Research, 4, 147–170, https://doi.org/10.2166/nh.1973.0012, 1973.

CLC: Corine Land Cover https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac (last access: 24 July 2024), 2018.

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

Dolich, A.: CAMELS-DE Processing Pipeline, Zenodo, https://doi.org/10.5281/zenodo.12760336, 2024.

DWD-HYRAS: https://www.dwd.de/DE/leistungen/hyras/hyras.html (last access: 25 March 2024), 2024.

EU-DEM: Copernicus GLO-30 DEM, https://doi.org/10.5270/esa-c5d3d65 (last access: 24 July 2024), 2022.

Feng, D., Liu, J., Lawson, K., & Shen, C.: Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. Water Resources Research, 58, e2022WR032404. https://doi.org/10.1029/2022WR032404, 2022.

Heberger, M.: delineator.py: Fast, accurate watershed delineation using hybrid vector- and raster-based methods and data from MERIT-Hydro, Zenodo, https://doi.org/10.5281/zenodo.10143149, 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 System Science Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Seibert, J., HBV Light Version 2. User's Manual. Department of Physical Geography and Quaternary Geology, Stockholm

University, Stockholm, https://www.geo.uzh.ch/dam/jcr:c8afa73c-ac90-478e-a8c7-929eed7b1b62/HBV_manual_2005.pdf (last access: 19. September 2024), 2005

Speckhann, G. A., Kreibich, H., and Merz, B.: Inventory of dams in Germany, Earth System Science Data, 13, 731–740, https://doi.org/10.5194/essd-13-731-2021, 2021.