https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Open Access Earth System Science Data Discussions

CAMELS-DE-1h: hourly hydro-meteorological time series, weather forecasts, and attributes for 1611 catchments in Germany

Alexander Dolich[1], Eduardo Acuña Espinoza[1], Uwe Ehret[1], Michael Kraft[2], Jan Bondy[3], Johannes Meuer[4], Ralf Loritz[1]

[1]Karlsruhe Institute of Technology (KIT), Institute for Water and Environment, Karlsruhe, Germany
[2]State Office for the Environment Rhineland-Palatinate, Mainz, Germany
[3]Deutscher Wetterdienst (DWD), Offenbach am Main, Germany
[4]Data Analysis Department, German Climate Computing Center (DKRZ), Hamburg, Germany

Correspondence to: Alexander Dolich (alexander.dolich@kit.edu)

Abstract. CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) datasets have been a major driver for advances in large-sample hydrology, facilitating regional studies and the development of deep learning methods and hydrological models by providing homogenized data across large domains, typically at the national scale. However, investigating highly dynamic events, such as flash floods, requires sub-daily resolution, which is often hindered by the daily time steps of most existing CAMELS datasets. Here, we present CAMELS-DE-1h, providing hourly time series of discharge and meteorology for 1,611 catchments in Germany, spanning the period from 2001 to 2024. This dataset homogenizes the extensive but deeply fragmented high-resolution hydrological gauge data managed independently by the German federal states, combining it with high-resolution meteorological forcing from the German Weather Service (DWD). With a median catchment area of 132.4 km², CAMELS-DE-1h includes many small-to-medium-sized basins where hydrological responses occur primarily on sub-daily scales. Alongside the time series, the dataset includes comprehensive static catchment attributes covering soil characteristics, land cover, hydrogeology, and human influences. A novel addition to CAMELS-DE-1h is a readily processed archive of operational short-term weather forecasts (ICON-D2, 48-hours lead time), which are used by German flood forecasting agencies in their operational settings. Including both deterministic runs and precipitation ensembles, these operational forecast data are available at a large scale for the first time for the period from 2021 to 2024. This allows for the evaluation of historical weather forecast quality, the testing of hydrological models in realistic operational settings, and the use of ensemble data to investigate the coupling of meteorological and hydrological forecast uncertainties. Finally, we provide baseline performance benchmarks using a regionally trained Long Short-Term Memory (LSTM) network and a conceptual HBV (Hydrologiska Byråns Vattenbalansavdelning) model. These models achieve median Nash-Sutcliffe Efficiencies (NSE) of 0.82 (LSTM) and 0.69 (HBV) for 1496 catchments selected based on their data availability for training / calibration and testing. By combining high-resolution observations with operational forecasts, CAMELS-DE-1h provides a consistent basis for the systematic comparison and development of hydrological and hydro-meteorological models under realistic conditions. CAMELS-DE-1h is available at: https://doi.org/10.5880/fidgeo.2026.045 (Dolich et al., 2026).

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

1 Introduction

CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) datasets provide discharge measurements, meteorological time series, and catchment attributes in a harmonized, accessible format, usually on the scale of entire countries, e.g. CAMELS-US (Newman et al., 2015; Addor et al., 2017), CAMELS-GB (Coxon et al., 2020), CAMELS-DE (Loritz et al., 2024), CAMELS-CH (Höge et al., 2023). The availability of such datasets has acted as a primary accelerator for the hydrological sciences - most significantly for large-sample hydrology - and has been a key driver in the rise of deep learning methods for rainfall-runoff modelling (Nearing et al., 2021). The adoption of CAMELS datasets as benchmarks enables direct comparison of model results, fundamentally changing how hydrological models, and specifically data-driven architectures, are developed and evaluated in a reproducible manner. However, large-sample hydrology currently focuses predominantly on the daily scale because most harmonized datasets provide only daily data, with a few exceptions discussed below. While the performance of daily rainfall-runoff models has reached a high standard - often plateauing with "vanilla" LSTMs (Liu et al., 2024, 2025) - this resolution imposes significant limitations on capturing rapid hydrological processes.

The Necessity of Sub-Daily Resolution

While daily time steps may be sufficient for modelling larger river basins, they are often inadequate for smaller catchments, where runoff generation is typically much more rapid. In small catchments, floods and peak discharge events frequently occur within hours following convective precipitation, signals that are smoothed out in daily aggregations or lost completely in terms of magnitude. Although the requirement for high-resolution data has been well established in hydrological research (Pilgrim et al., 1982), its practical implications, especially for rainfall-runoff modelling, are profound. Figure 1 illustrates this impact using a flood event in the "Gampelmühle" catchment in Bavaria (River Ölschnitz, catchment area of 63.8 km²). In this example, the daily data resolution fails to capture the dynamics of a 100-year flood (HQ_{100}) event; the daily aggregate smooths the peak as well as the dynamics of the flood. Conversely, hourly data resolve the event dynamics much better, capturing the small peaks before and after the main event, the rapid rising limb, and the peak magnitude. Analysing flood events in small-to-medium-sized catchments therefore strictly requires sub-daily discharge data; moreover, this higher temporal resolution is essential for training models, such as the depicted hourly LSTM, to accurately represent these characteristics. Despite this clear need, the availability of harmonized datasets at sub-daily resolution remains scarce globally. Only a few datasets, such as the recently published CAMELS-GB v2 (Coxon et al., 2025), CAMELS-NZ (Bushra et al., 2025), and CAMELS-LUX (Nijzink et al., 2025), currently include data in hourly resolution. While first specific deep-learning architectures aiming at high-resolution simulations have already been published, such as the MF-LSTM (Acuña Espinoza et al., 2024) and MTS-LSTM (Gauch et al., 2021) architectures, the rigorous testing and benchmarking of these methods are hindered by a lack of standardized sub-daily datasets. To advance the field, the community requires high-resolution harmonized datasets and modelling benchmarks for research on runoff generation in small catchments and the precise timing of hydrological events. While hourly data is necessary to capture dynamics of many hydrological processes, it

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

should also be noted that uncertainty is typically higher and data quality is often worse in comparison to daily data, e.g. by introducing sensor noise to the discharge data which is smoothed out by the daily aggregation. Furthermore, working with hourly data is computationally more demanding, as the amount of data rises, posing a challenge especially in large-sample modelling tasks.

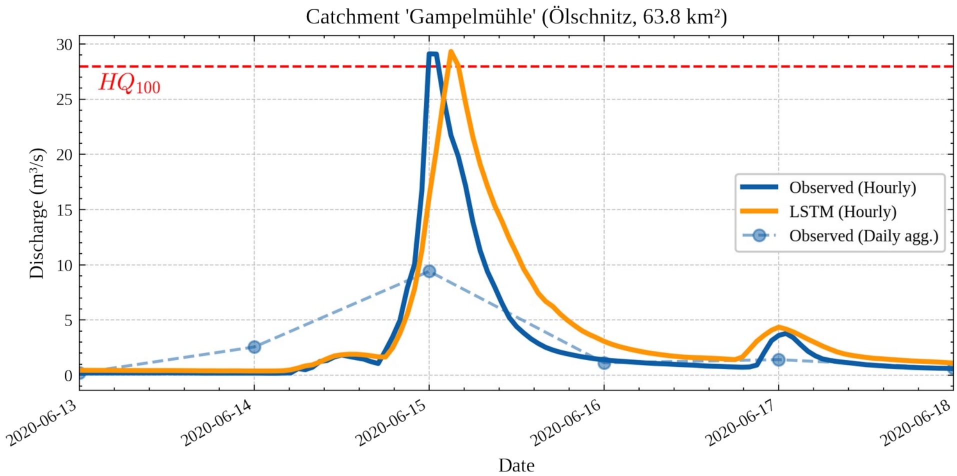

Figure 1: Comparison of hourly discharge observations and LSTM discharge simulations for a 100-year flood event (HQ_{100}) in the Gampelmühle catchment (river Ölschnitz, Bavaria; 63.8 km²). The dashed line shows the hourly observations aggregated to daily means. The hourly data and model resolve the rapid event dynamics, whereas temporal aggregation smooths the hydrograph and significantly underestimates the flood peak.

Bridging the Gap to Operational Forecasting

Currently, hydrological modelling studies predominantly evaluate performance in "simulation mode", relying exclusively on observed meteorological records as input (hindcasting, pseudo-forecasts). In contrast, many real-world applications, especially flood forecasting, require models to generate streamflow predictions from uncertain meteorological forecasts, exposing them to substantially higher input uncertainty and fundamentally different operational constraints than those encountered in hindcast evaluations. Operational systems must provide short-term forecasts with typical lead times of 24 to 72 hours for warning issuance in the context of civil protection or the regulation of reservoir outflows. These workflows usually rely on the usage of Numerical Weather Predictions (NWP) as meteorological model inputs, which differ significantly from meteorological observations due to inherent forecast errors and uncertainties (Konold et al., 2025).

For any hydrological model to transition from simulation mode to operational forecasts, it must prove its robustness within this specific context. This transition requires moving beyond purely observational benchmarks and evaluating models using real-world historical weather forecasts to assess how forecast uncertainty propagates into hydrological predictions. Furthermore, meteorological ensemble forecasts are needed for developing probabilistic forecasting chains, including both classical ensemble prediction systems and modern probabilistic machine learning approaches. To date, the lack of accessible, large-sample datasets containing high-resolution historical ensemble NWP data has been a major barrier to the advancement

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

of academic research in this direction and to the transition of large-sample hydrological methods into operational practice. While the Caravan MULTIMET dataset (Shalev and Kratzert, 2024) recently introduced a global archive of operational weather forecasts, it is limited to a daily resolution and lacks ensemble predictions.

To address the need for hourly large-sample hydrological datasets and to enable benchmarking under real-world forecasting conditions in mesoscale catchments, we introduce CAMELS-DE-1h. This comprehensive hydro-meteorological dataset covers 1,611 catchments in Germany with time series data spanning the period from 2001 to 2024. Designed to be fully compatible with the existing daily CAMELS-DE dataset (Loritz et al., 2024), the identical catchment identifiers and boundaries for catchments that are part of both datasets are utilized, aiming to facilitate seamless analyses across the different time scales. CAMELS-DE-1h makes the following distinct contributions to the hydrological community:

1. **High-Resolution Data**: Harmonized streamflow data and meteorological forcings are provided in hourly resolution, enabling the detailed study of the runoff generation in small-to-medium-sized catchments.
2. **Operational Relevance**: Historical short-term meteorological forecasts (ICON-D2, 48-hour lead time) from 2021 to 2024 (DWD, 2021), both deterministic and ensemble predictions are processed and available for all catchments. This enables the development and evaluation of forecast models for operational settings and facilitates research into the propagation of meteorological uncertainty into hydrological predictions.
3. **Benchmarks**: Simulations from an LSTM (Hochreiter and Schmidhuber, 1997) and the HBV (Hydrologiska Byråns Vattenbalansavdelning) conceptual model (Bergström and Forsman, 1973; Seibert, 2005) provide baselines for hourly discharge simulation in Germany, facilitating future model development and comparison.

2 Data sources and providers

At the core of CAMELS-DE-1h are the hourly time series of observed discharge for 1611 catchments (Sect. 4.1). Collecting these data in Germany poses a significant challenge due to Germany's federal structure, where streamflow monitoring and data management are organized independently by the federal states. Consequently, while high-quality hydrological data is abundant, it is decentralized and fragmented, making nationwide collection and harmonization a highly labor-intensive process. In addition, licensing and publication agreements must be concluded with each federal state if their data are not already publicly available and openly licensed. The discharge ($m^3\ s^{-1}$) and water level (cm) time series data was sourced from 12 different federal state agencies in Germany (Fig. 2a): Landesanstalt für Umwelt Baden-Württemberg (LUBW); Bayerisches Landesamt für Umwelt (LfU-Bayern); Landesamt für Umwelt Brandenburg (LfU-Brandenburg); Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG); Landesamt für Umwelt, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, Landesamt für Natur (NLWKN); Landesamt für Natur, Umwelt und Klima Nordrhein-Westfalen (LANUK NRW); Landesamt für Umwelt Rheinland-Pfalz (LfU-Rheinland-Pfalz); Landesamt für Umwelt- und Arbeitsschutz Saarland (LUA); Landesamt für Umwelt, Landwirtschaft und Geologie Sachsen (LfULG); Landesamt für Umweltschutz Sachsen-Anhalt (LAU); Landesamt für Landwirtschaft, Umwelt und ländliche Räume

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Schleswig-Holstein (LLUR); and Thüringer Landesamt für Umwelt, Bergbau und Naturschutz (TLUBN). In addition to the federal states, several water management associations (“Wasserwirtschaftsverbände”) in North Rhine-Westphalia, which also manage and operate streamflow gauges on a local level, contributed their hydrological gauge data. These associations are Bergisch Rheinischer Wassserverband, Niersverband, Aggerverband, Ruhrverband, and Erftverband. The city states of Berlin, Hamburg and Bremen, which account for only 0.6% of Germany's total area, are not included. Furthermore, the gauges of the federal state of Mecklenburg-Vorpommern are also not included as they do not provide the data with an open license. In addition, most federal waterways are excluded, as they comprise heavily managed canals (e.g., the Mittellandkanal) and major transboundary rivers (e.g., the Rhine, Moselle, and Elbe). These major rivers extend beyond the German borders, where the required harmonized meteorological forcing data are unavailable.

Meteorological forcing data were derived from two primary high-resolution gridded products provided by the German Weather Service (Deutscher Wetterdienst, DWD). Precipitation data (Sect. 4.2) was sourced from RADKLIM-YW (Winterrath et al., 2018), a radar-based climatology dataset adjusted with gauge observations in 5-minute resolution. To fill gaps in the RADKLIM-YW time series, an infilled version of RADKLIM-YW was utilized (Meuer et al., 2025). All other meteorological variables (Sect. 4.3) were obtained from HOSTRADA (DWD, 2024; Krähenmann et al., 2018), a high-resolution interpolated dataset derived from station observations in hourly resolution. Short-term weather forecasts (Sect. 5) were processed from ICON-D2 (DWD, 2021), the non-hydrostatic numerical weather prediction (NWP) model currently in operational use by the DWD. DWD makes real-time forecasts available via its Open Data server, but a public archive of historical ICON-D2 forecasts does not exist. Consequently, the historical forecast data included in CAMELS-DE-1h were not sourced directly from the DWD, but were provided by the Flood Forecasting Centre of Rhineland-Palatinate (LfU-Rheinland-Pfalz). The data include both deterministic and ensemble forecasts generated with ICON-D2.

The static catchment attributes (Sect. 6) included in CAMELS-DE-1h correspond to those provided in the daily resolution CAMELS-DE dataset. To ensure consistency and compatibility, the identical data sources and processing workflows were utilized. Topographical attributes were derived from the Copernicus GLO-30 Digital Elevation Model (European Space Agency and Airbus, 2022), while land cover characteristics were obtained from the CORINE Land Cover 2018 dataset (European Environment Agency, 2019). Soil properties were extracted from the global SoilGrids250m dataset (Poggio et al., 2021). For hydrogeological attributes, the "Hydrogeologische Übersichtskarte von Deutschland 1:250.000" provided by the Federal Institute for Geosciences and Natural Resources (BGR, 2019) was used. Finally, indicators of human influence, such as the presence of dams and weirs, were sourced from Speckhann et al. (2021).

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

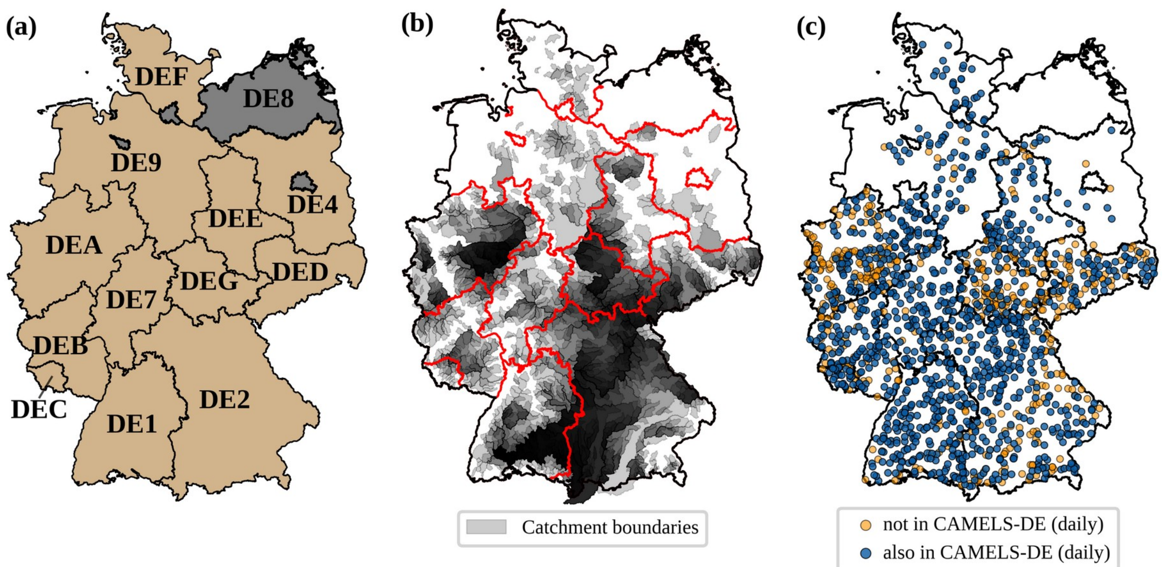

Figure 2: Panel (**a**) shows all German federal states labelled with their NUTS level 1 IDs (see Sect. 6.1), as used in the CAMELS-DE gauge IDs, grey colored states are not included in CAMELS-DE-1h. Panel (**b**) shows the catchment geometries, as the geometries are plotted transparently, a darker colour translates to a higher density of catchments in that area. Panel (**c**) shows all gauge locations coloured based on if the catchment is also part of the daily CAMELS-DE dataset. Borders of Germany: © GeoBasis-DE / BKG (2023). Figure adopted and modified from (Loritz et al., 2024).

3 Catchment selection and boundaries

To ensure compatibility between the hourly and daily versions of CAMELS-DE, the existing catchment geometries and identifiers (IDs) were adopted without modification for all gauging stations already present in the daily CAMELS-DE dataset (Loritz et al., 2024). This allows for the direct reuse of static attributes and facilitates cross-comparison between daily and hourly data and model results. In total, 1133 out of the 1611 catchments in CAMELS-DE-1h were also part of the daily dataset (Fig. 2c), the information whether a catchment is also part of the daily dataset is included in the topographic attributes (*in_camelsde_1d*). For gauges not included in the daily dataset - either due to the relaxed selection criteria described below, or newly available data, catchment boundaries were derived using the identical methodology established in the daily version. The MERIT Hydro dataset (Yamazaki et al., 2017) was utilized in conjunction with the *delineator.py* package (Heberger, 2023) to automatically derive boundaries based on the gauge location. This approach was necessary because official catchment geometries were not available for all gauging stations in Germany and even where boundaries were provided, the derivation methods used by the different federal states are not uniform. By consistently deriving all catchment geometries from the MERIT Hydro dataset with the same methodology, a homogeneous and reproducible set of catchment boundaries was ensured across the entire dataset. As a quality control measure and to minimize the uncertainty inherent in automated delineation, the geometries were validated by comparing the derived area to the official catchment area reported in the gauge metadata. Consistent with the daily dataset, only catchments where the derived area deviated by

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

less than 20 % from the reported metadata were retained. Catchment delineation can represent a major source of uncertainty in the dataset, as both the meteorological forcing and the catchment attributes depend on accurate catchment geometries. To assess this uncertainty, both the catchment area reported in the gauge metadata (*area_metadata*) and the MERIT-Hydro-based area (*area*) are reported in the topographic attributes.

The catchments included in CAMELS-DE-1h were selected based on multiple criteria designed to ensure data quality and coverage. While these selection criteria are similar to those of the daily CAMELS-DE dataset, specific constraints were relaxed to include a higher number of catchments, particularly in the border regions of Germany. This relaxation was also possible as precipitation data in hourly resolution is available beyond the German national borders. Additionally, unlike the daily dataset, no lower or upper bounds were applied to the catchment area. In total, hourly hydrological data and metadata were sourced for 2,014 gauges from the data providers. From this pool, catchments were selected based on the following criteria:

- **Metadata availability**: Stations were required to have a valid gauge name, location, and a reported catchment area (n = 1993).
- **Delineation accuracy**: Consistent with the daily dataset, the derived catchment area was required to deviate by no more than 20% from the officially reported area to ensure spatial accuracy (n = 1835), the deviation is visualized for all catchments in Figure 3c. The median absolute deviation of calculated to reported catchment area is 0.85 % for the catchments included in the final CAMELS-DE-1h dataset.
- **Minimum Data Length**: The minimum requirement for discharge data availability was reduced from 10 years in CAMELS-DE to 1 year (n = 1935). This criterion was relaxed as recent studies have demonstrated that hydrological simulations can benefit significantly even from relatively short observation records (Etter et al., 2018; Pool et al., 2017; Seibert and Beven, 2009; Staudinger et al., 2025). The vast majority of included catchments possess much longer records of discharge data (median length of 23.8 years, see Fig. 3b).
- **Spatial Coverage**: In the daily CAMELS-DE dataset (Loritz et al., 2024), catchments were required to be located 100% within the borders of Germany to strictly avoid uncertainties related to missing meteorological data. For CAMELS-DE-1h, this threshold was relaxed to include catchments with at least 90% of their area within Germany (n = 1813). This decision represents a trade-off between minimizing forcing uncertainty and maximizing the number of headwater catchments - many of which extend slightly across national borders, especially in the foothills of the Alps and the border region between Saxony and the Czech Republic. Precipitation as the primary meteorological forcing for many hydrological studies is derived from radar measurements (RADKLIM-YW). As the radar coverage extends beyond the German national borders, it also provides complete data for the transboundary portions of these catchments (Fig. 5a). However, other meteorological variables sourced from HOSTRADA are limited to the German territory. By allowing a maximum of 10% area outside Germany, a degree of uncertainty in non-precipitation variables is accepted to retain a significant number of valuable headwater catchments. To estimate the uncertainty introduced for these specific catchments, the attribute

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

area_pct_in_germany is provided in the topographic attributes, allowing users to filter the dataset according to specific strictness requirements. Generally, only a small fraction of the catchments included in the final dataset extends notably beyond the national border (catchments with > 98% area inside Germany: n = 1546, Fig. 3a).

Applying these criteria yielded 1,637 catchments. From this selection, 19 catchments were removed because major parts of the precipitation data were missing due to insufficient radar coverage (Sect. 4.2), and a further 7 catchments were excluded following a visual inspection of the discharge time series (Sect. 4.1). This resulted in a final set of 1,611 catchments for CAMELS-DE-1h (Fig. 2). The resulting dataset covers a wide range of spatial scales, with a median catchment area of 132.4 km². The distribution of catchment sizes is illustrated in Fig. 4b, showing a large variety of catchment sizes and a high number of small to medium sized catchments, where the hourly resolution of the dataset is especially valuable.

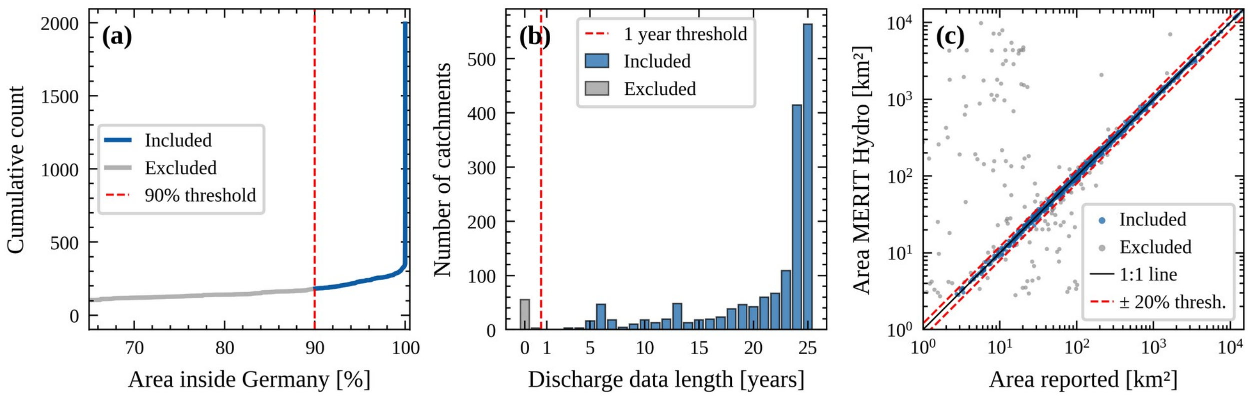

Figure 3: Visualization of the selection criteria and thresholds for catchments to be included in CAMELS-DE-1h. Panel (**a**) shows the cumulative count of catchments against their area percentage located inside Germany. Panel (**b**) shows a histogram of discharge time series length for all catchments, catchments with 0 years of discharge data means that only water level data was provided. Panel (**c**) shows a scatterplot of the deviation between the reported catchment area and the calculated area of the MERIT Hydro catchment. Figure adopted and modified from Loritz et al. (2024).

4 Time series

CAMELS-DE-1h provides hourly hydro-meteorological time series, as detailed in Table 1, covering the period from 1 January 2001 to 31 December 2024. The dataset starts in 2001 as RADKLIM-YW, the source for precipitation data starts in 2001. While discharge and water level data were provided by the sources described in Section 2 (primarily the federal state agencies), meteorological data were sourced from the German Weather Service (DWD) as gridded raster products. The precipitation data from radar measurements extends beyond the German national borders, whereas other meteorological variables are limited to the German territory. To ensure consistency across the dataset and to ease integration into global datasets, all time series are provided in Coordinated Universal Time (UTC+0). Meteorological data were already provided in UTC by the DWD. In contrast, hydrological data were frequently provided in Central European Time (CET/MEZ, UTC+1) and were subsequently transformed to UTC by subtracting one hour. The meteorological raster data were clipped to the catchment boundaries and aggregated to hourly time series using the *exactextract* package (Baston, 2025). This method

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

ensures high accuracy by calculating the fraction of each raster cell covered by the catchment polygon and weighting the cell values accordingly when computing spatial statistics such as the mean.

4.1 Discharge and water levels

The discharge and water level time series were sourced from the federal state agencies and water associations in North Rhine-Westphalia. All observed hourly discharge and water level data follow a "right-labeling" convention, where the value reported at a specific timestamp (e.g., 10:00) represents the mean value over the preceding hour (i.e., from 09:00 to 10:00). Data originally supplied at a 15-minute resolution were aggregated to hourly intervals by calculating the mean of the four preceding 15-minute records. This temporal alignment ensures consistency with precipitation and other meteorological variables. It is important to note that while the data format within CAMELS-DE-1h is homogeneous, the raw data originate from various providers (see Section 2). These providers utilize different methods for data collection, storage, and quality control. Consequently, the quality and uncertainty of the discharge and water level data may vary between federal states. The availability of discharge and water level data for all catchments is depicted in Figure 4a, showing that generally, discharge data is available at more gauges than water level data and the number of catchments with data steadily increases from 2001 to around 2022, when the majority of catchments have data, and a drop in data availability afterwards.

Quality Control and Data Reliability

No exhaustive, direct quality control procedures were applied for CAMELS-DE-1h. Instead, reliance was placed primarily on the quality assurance protocols of the individual data providers (federal states). However, as the source data comprise a mix of raw data and quality-controlled records subject to varying methods, a set of basic processing steps and indirect quality assessments were applied. All time series containing negative discharge values (n=54) were visually inspected. Following this review, all negative records were replaced with *NaN*. Unlike the daily dataset, CAMELS-DE-1h does not include tide-influenced catchments where negative flow might be physically valid; thus, all values below zero were treated as erroneous. To further assess data quality, an indirect screening method utilizing the hydrological benchmark models was employed (Sect. 7). Hydrographs were visually inspected for all catchments where both the HBV model and the regionally trained LSTM model only achieved Nash-Sutcliffe Efficiencies (NSE) < 0.4. These benchmark results serve as a proxy for the consistency between precipitation and discharge: the catchment-calibrated HBV model indicates whether a rainfall-runoff relationship exists and the mass balance is closed, whereas the LSTM can partially compensate for mass balance inconsistencies. However, the LSTM still relies on a causal relationship between the meteorological forcing and the target discharge. Therefore, if both models perform worse than simply predicting the long-term mean (NSE < 0.0), this lack of predictive skill strongly indicates either erroneous data or significant anthropogenic influence on river flow. Consequently, seven catchments were removed from the dataset after visual inspection confirmed clear data errors. A conservative approach was adopted regarding data removal. Data were only deleted where hydrographs showed clear sensor errors or artifacts. Hydrographs exhibiting "unnatural" behavior were retained if the decoupling of rainfall and runoff could be

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

attributed to natural features (e.g., Karst systems) or, more commonly, anthropogenic influences. Many catchments with low benchmark performance are heavily managed, containing weirs, locks (*Schleusen*), canals (*Kanäle*), or reservoirs (*Talsperren* / *Stauseen*), often indicated by the gauge or river name. In some cases, regime shifts were visible in the hydrograph, likely pointing to the construction of new water management infrastructure during the observation period.
It is worth noting that hourly discharge data is inherently more prone to noise, such as sudden jumps or short plateaus, than daily data, where such artifacts are often smoothed out by aggregation. Therefore, rather than applying aggressive filters to the raw data, we recommend that users utilize the benchmark results described in Section 7 to assess the suitability of specific catchments for rainfall-runoff modelling. We welcome feedback from the community regarding specific data issues to help improve the quality of CAMELS-DE-1h in future versions.

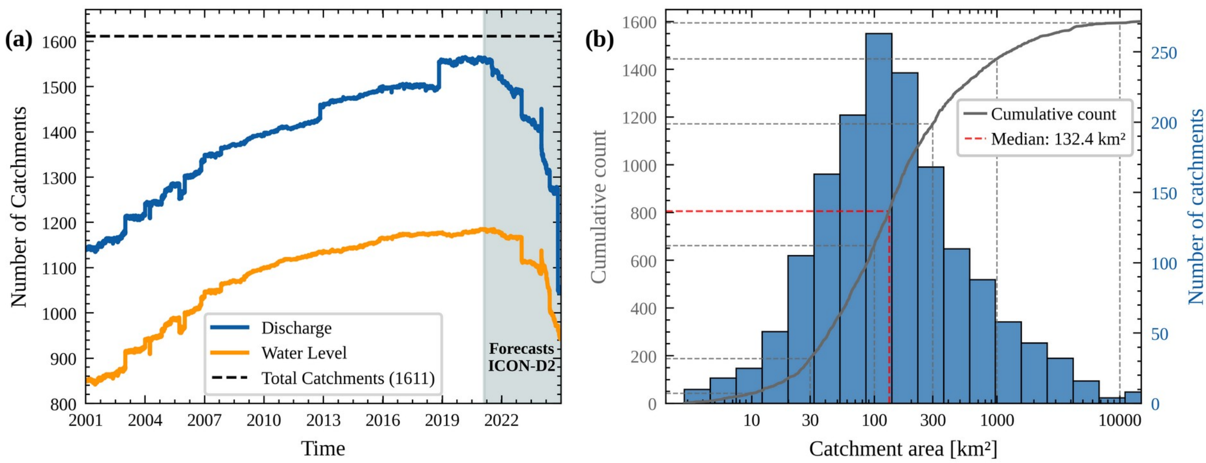

Figure 4: Panel (a) shows the availability of discharge and water level data for all catchments over the time period of CAMELS-DE-1h as well as the availability of ICON-D2 forecast data starting in 2021. Panel (**b**) shows a histogram as well as the cumulative count of catchment areas in CAMELS-DE-1h. Figure adopted and modified from Loritz et al. (2024).

4.2 Precipitation

Hourly precipitation data were derived from the RADKLIM-YW dataset, a high-resolution climatological reference product provided by the DWD. This dataset covers the period from 2001 to 2024 with a spatial resolution of 1 km x 1 km and a temporal resolution of 5 minutes. The RADKLIM products are based on the national weather radar network of 17 C-band Doppler radars, which continuously scan the atmosphere to measure both spatial extent and intensity of precipitation, capturing also short-lived convective storms often missed by the station network. To improve accuracy, the radar reflectivity-derived precipitation fields are adjusted using measurements from quality-controlled rain gauges. In contrast to the operational RADOLAN datasets, which are generated in real-time, RADKLIM is a reprocessed product designed for climatological consistency. It utilizes all available station data records and applies retrospective correction algorithms (e.g., artifact removal) to create a homogeneous dataset (Winterrath et al., 2018). It should be noted that previous studies have shown that RADKLIM data nevertheless can systematically underestimate precipitation (Kreklow et al., 2020). Additionally,

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

the radar network as well as the number of ground stations used to adjust the radar data changed over time, leading to temporal inconsistencies in the dataset (Winterrath et al., 2017).

For CAMELS-DE-1h, the 5-minute RADKLIM-YW dataset was selected over the standard hourly RADKLIM-RW product. While RADKLIM-RW provides fully gauge-adjusted hourly totals, it presents two major limitations. First, the RW product is generated with a timestamp at 50 minutes past each hour (e.g., 00:50, 01:50), causing misalignment with other meteorological variables and streamflow records recorded at the full hour. Second, the RW data are clipped to the German national borders, resulting in incomplete coverage of headwater catchments located in border regions. To address these issues, the RADKLIM-YW product was utilized. YW provides "quasi-adjusted" precipitation rates, meaning that 5-minute radar scans are scaled to match the gauge-adjusted hourly RW totals, and retains the full radar circles extending beyond the national border. To ensure temporal consistency, the 5-minute YW data were aggregated into hourly accumulations by calculating the mean precipitation rate from the available 5-minute frames and scaling it to a full 60-minute duration.

For each catchment, the hourly aggregated YW gridded fields were clipped to the catchment boundaries. The mean, minimum, maximum, and standard deviation of precipitation grid values were calculated for each time step to reflect the spatial variability of the precipitation field within the catchment. However, the RADKLIM-YW time series contain gaps where radar observations were unavailable. While often short, these gaps can be significant in specific regions and periods (Fig. 5). The handling of these data availability issues and the applied gap-filling strategy are detailed in the following subsection.

Data Availability and Gap Filling

While the radar network generally provides robust precipitation coverage across Germany, the time series are subject to interruptions when radars are out of service due to maintenance or technical failure. Although the majority of these gaps are short, typically spanning only a few hours (Fig. 5b), significant data gaps exist in regions covered by a single radar device, where no backup coverage from overlapping radars is available to compensate for longer outages. It is worth mentioning that (short) gaps due to radar maintenance are scheduled and planned, anticipating fair weather in its domain, which should limit the hydro-meteorological importance of this type of data gaps. The gaps are spatially and temporally heterogeneous. As shown in Fig. 5a, regions such as southern Bavaria, Saarland, Thuringia and northern Lower Saxony lack redundant coverage, making them particularly susceptible to extended periods without radar coverage. For instance, catchments in Thuringia exhibit a gap of approximately 280 days in 2011, and catchments in Saarland miss approximately 230 days of data between 2013 and 2014 due to long-term device outages. Notably, the frequency and duration of these gaps decrease significantly after 2015 (Fig. 5c), when the radar network was upgraded to modern dual-polarization Doppler radar systems.

The aggregation of gridded radar data to the catchment scale introduces two distinct types of data availability issues:

1. **Complete Gaps**: No operating radar covers the catchment area during the time step, resulting in NaN values.
2. **Partial Coverage**: Catchments located at the edge of a radar's range are only partially covered during specific time steps. Aggregating precipitation from such incomplete spatial data can introduce high uncertainty.

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

To quantify data availability, the percentage of the catchment area covered by valid radar measurements for each time step was calculated (0 % indicating no data, 100 % indicating full coverage). This "coverage" metric is included in the time series files (*perc_nan_precipitation_original*), allowing users to apply filters of variable strictness (e.g., requiring 100 % coverage) if necessary.

Continuous forcing time series are desirable for many hydrological applications, particularly for calibrating conceptual models and training deep learning architectures. This means that for such applications, the gaps in precipitation should be filled. To ensure the comparability of results and studies and therefore ensure the usability of CAMELS-DE-1h as a benchmark dataset, we put particular effort into filling the precipitation gaps. Standard methods, such as using data from the nearest meteorological ground station or downsampling daily gridded products were discarded. Ground station density is insufficient to capture local dynamics, particularly prior to 2005, and downsampling daily products results in a loss of the sub-daily dynamics which are the core value of this hourly dataset. Consequently, a gap-filled precipitation time series is included, processed from an infilled gridded RADKLIM dataset generated by a memory-assisted deep learning approach (Meuer et al., 2025), which was extended through 2024 for CAMELS-DE-1h. As this method makes use of both spatial and temporal context, it excels at preserving sub-daily dynamics compared to the disaggregation of daily grids. Meuer et al. (2025) demonstrated that the model outperforms standard baselines across complementary metrics of pointwise error, temporal consistency, and spatial structure. Real-outage case studies further showed that the method preserves coherent precipitation patterns across consecutive frames, although extreme accumulations may still be underestimated. All complete gaps as well as instances of partial coverage (see above) in the precipitation time series were filled using this dataset. In the case of partial coverage, the infilled data was only utilized to supplement the unmeasured area; for the rest of the catchment, the original data was retained. The gap-filled data should be used with caution, as it introduces uncertainty that increases with the temporal length of the gap and the percentage of missing data. The vast majority of outages are thereby brief, with 77.1 % of the gaps extending for six hours or less (Fig. 5b). Furthermore, 89.0 % of the infilled precipitation values are smaller than 0.1 mm and therefore correspond to periods of no to very little precipitation, where meteorological conditions are stable and the associated estimation error is expected to be low. The flag *perc_nan_precipitation_original* (see above) enables users to strictly utilize original observations only or to apply custom filters, such as retaining infilled data only for short interruptions (e.g., up to three hours) while treating longer, more uncertain gaps as missing data, or to fill partial gaps based on thresholds of available original measurement.

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Open Access Earth System Science Data Discussions

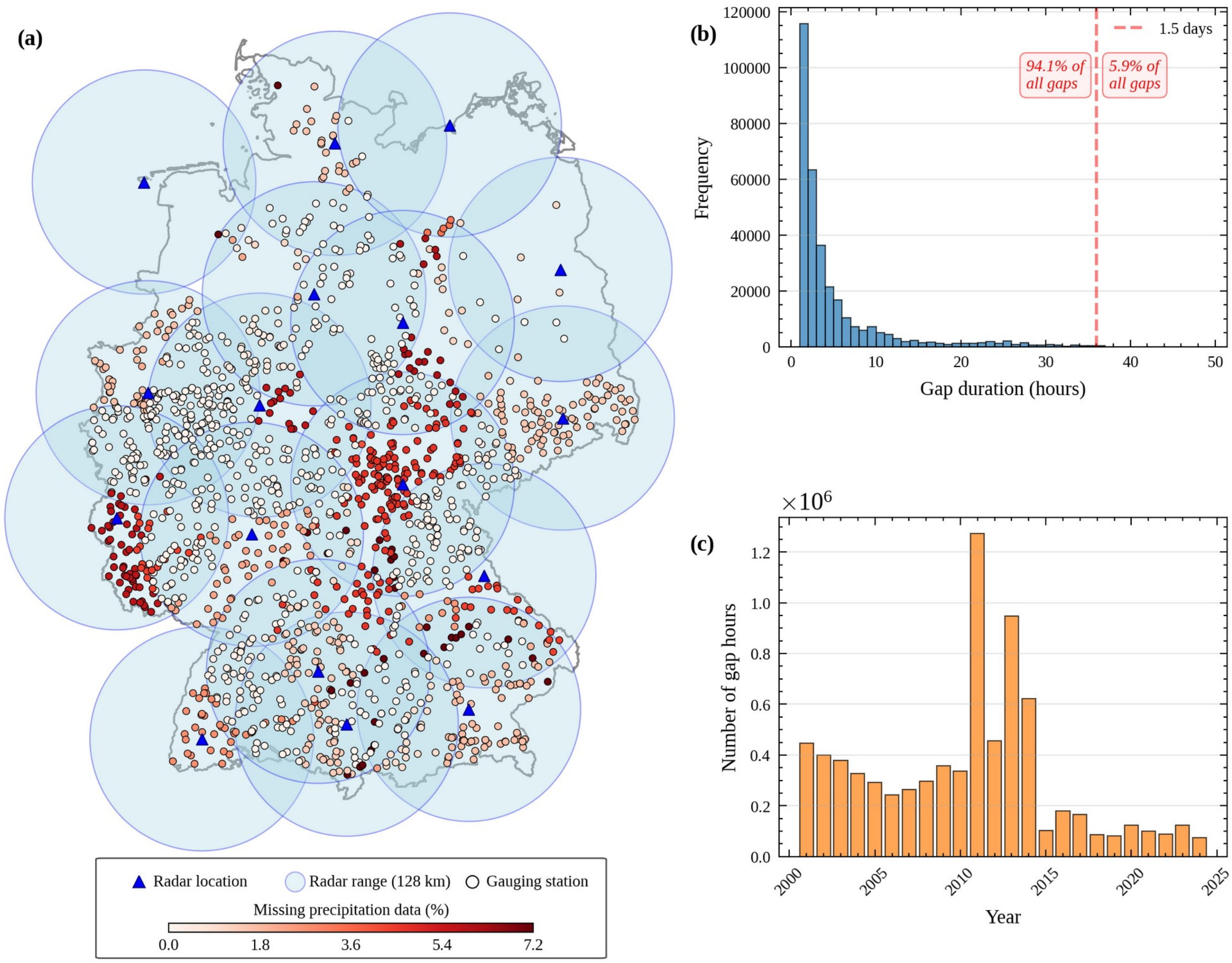

Figure 5: Panel (**a**) shows the locations of the radars that are part of the German weather radar network used to produce RADKLIM as well as their 128 km measuring range. Darker blue colors of the measuring radii translate to areas covered by multiple radars. CAMELS-DE-1h river gauging stations are plotted as circles and are colored based on the percentage of full and partially missing precipitation data; the color scale is capped at the 98th percentile of the missing data distribution, with stations above this threshold shown in the darkest red. Panel (**b**) shows a histogram of the gap duration for all gaps shorter than 50 hours. Panel (**c**) shows a histogram of the number of partial and full gap hours for each year in the time period of CAMELS-DE-1h.

4.3 Other meteorological variables

CAMELS-DE incorporates hourly meteorological variables from the HOSTRADA dataset, a climatological reference product for Germany (DWD, 2024). This dataset, available since 1995, provides interpolated meteorological parameters at a spatial resolution of 1 km x 1 km and a temporal resolution of one hour. The HOSTRADA data is based on the interpolation of hourly station data from the official German meteorological network, complemented by satellite and climate model data to ensure a consistent and high-quality product. The meteorological variables sourced for CAMELS-DE-1h include 2-meter air temperature, dew point temperature, cloud cover, wind speed and direction at 10-meter height, relative humidity, water vapor mixing ratio, air pressure at station height and sea level, and global shortwave radiation. The provided data represents

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

instantaneous values at the start of each hour, with the exception of radiation, which is the sum over the last hour. The gridded data was aggregated to a time series for each catchment by clipping it to the catchment boundaries and calculating the spatial mean for each variable and timestep. It should be noted that the HOSTRADA dataset does not extend beyond Germany's borders, which is why catchments with a portion of their area outside Germany are not fully covered by HOSTRADA data. This portion amounts to a maximum of 10 percent (Sect. 3) and can be determined using the topographical attributes (Sect. 6).

Table 1: Catchment-specific hydro-meteorological variables available as hourly time series in CAMELS-DE-1h.

Time series class	Time series name	Description	Unit	Data source
Hydrologic time series (1 Jan 2001 01:00 - 31 Dec 2024 23:00)	discharge_vol_obs	Catchment discharge, usually calculated from the water level and gauge geometry	$m^3 s^{-1}$	Federal state agencies and associations (Sect. 2)
	discharge_spec_obs	Observed catchment-specific discharge (converted to millimetres per hour using the MERIT Hydro catchment areas described in Section 3)	$mm\ hr^{-1}$	
	water_level_obs	Observed hourly water level	cm	
Meteorologic time series (1 Jan 2001 01:00 - 31 Dec 2024 23:00)	precipitation_mean, precipitation_min, precipitation_max, precipitation_stdev	Radar-observed and gauge adjusted spatial mean, minimum, maximum and standard deviation of hourly precipitation sums over the last hour (RADKLIM-YW, 1x1 km2)	$mm\ hr^{-1}$	RADKLIM-YW (Winterrath et al., 2018), Meuer et al. (2025), and HOSTRADA (DWD, 2024)
	precipitation_mean_gapfilled, precipitation_min_gapfilled, precipitation_max_gapfilled, precipitation_stdev_gapfilled	Spatial mean, minimum, maximum and standard deviation of hourly precipitation sums calculated from a spatially gap-filled version of RADKLIM-YW (Meuer et al., 2025)	$mm\ hr^{-1}$	
	perc_nan_precipitation_original	Percentage of the catchment area that is not covered by the original RADKLIM-YW precipitation data	%	
	air_temperature_mean air_temperature_min air_temperature_max	Spatial mean, minimum and maximum of mean air temperature at 2 m height (instantaneous values at the start of the hour)	°C	
	relativ_humidity_mean	Spatial mean of relative humidity at 2 m height (instantaneous values at the start of the hour)	%	
	water_vapor_mixing_ratio_mean	Spatial mean of water vapor mixing ratio at 2 m height (instantaneous values at the start of the hour)	$g\ kg^{-1}$	
	global_radiation_mean	Spatial mean of global short wave radiation (sum over the last hour)	$W\ m^{-2}$	
	air_pressure_sea_level_mean	Spatial mean of air pressure at sea level height (instantaneous values at the start of the hour)	hPa	
	air_pressure_surface_mean	Spatial mean of air pressure at the surface (instantaneous values at the start of the hour)	hPa	
	cloud_cover_mean	Spatial mean of cloud coverage (instantaneous values at the start of the hour)	%	
	dew_point_temperature_mean	Spatial mean of dew point temperature at 2 m height (instantaneous values at the start of the hour)	°C	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

	wind_speed_eastward_mean	Spatial mean of eastward wind speed at 10 m height (instantaneous values at the start of the hour)	m s^{-1}	
	wind_speed_northward_mean	Spatial mean of northward wind speed at 10 m height (instantaneous values at the start of the hour)	m s^{-1}	
	wind_speed_mean	Mean catchment wind speed at 10 m height, calculated from eastward and northward wind components (instantaneous values at the start of the hour)	m s^{-1}	
	wind_direction_mean	Mean catchment wind direction at 10 m height, calculated from eastward and northward wind components, representing the direction of the mean wind vector (instantaneous values at the start of the hour)	°	
Simulated hydrologic time series (testing period: 1 Jan 2020 01:00 – 31 Dec 2024 23:00)	discharge_spec_obs / y_obs	Observed catchment-specific discharge	mm hr^{-1}	Regional LSTM, conceptual HBV (Sect. 7)
	discharge_spec_sim_ensemble / y_sim_ensemble	Catchment-specific discharge simulated with the LSTM and the HBV (Sect. 7). Ensemble median time series which is also used for the calculation of benchmark NSEs (Table 3)	mm hr^{-1}	
	discharge_spec_sim_seed_N / y_sim_seed_100 y_sim_seed_110 y_sim_seed_120	Catchment-specific discharge simulated with the LSTM and the HBV for all ensemble members (Sect. 7).	mm hr^{-1}	

5 Weather Forecasts

While most large-sample hydro-meteorological datasets, with the exception of the global Caravan-MultiMet dataset (Shalev and Kratzert, 2024), primarily focus on historical meteorological observations, CAMELS-DE-1h additionally introduces historical numerical weather prediction (NWP) data. The inclusion of meteorological forecasts allows users to move beyond calibration and validation of rainfall-runoff models on observed data towards assessing the performance of hydrological models in a realistic operational forecasting setting. This is crucial for applications such as flood prediction and warnings as well as real-time reservoir management. By providing both the deterministic run and the ensemble members of the state-of-the-art NWP system ICON-D2 (DWD, 2021), CAMELS-DE-1h enables research about the propagation of meteorological uncertainty into hydrological predictions, the development of deep learning models for operational settings, and the development of probabilistic flood forecasting frameworks.

5.1 ICON-D2 Forecasting System

The meteorological forecast data in CAMELS-DE-1h is derived from ICON-D2, the current operational, limited-area numerical weather prediction (NWP) system of the DWD. The ICON (ICOsahedral Nonhydrostatic) modelling framework (DWD, 2021; Zängl et al., 2022) replaced the former COSMO-D2 system on 10 February 2021. The ICON-D2 model domain covers Germany and neighboring countries (including Austria, Switzerland, and parts of the Netherlands, Belgium, France, Luxembourg, Poland, and the Czech Republic) with its native triangular computational grid corresponding to an effective horizontal grid resolution of around 2.2 km and 65 vertical levels. Due to its high spatial resolution, ICON-D2 is a

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

"convection-permitting" model. Unlike strongly parametrized coarser models, ICON-D2 is able to better account for convective processes. This capability improves precipitation forecast performance, particularly for short-term, local, high-impact summer convective events. Furthermore, the model better captures topographically induced phenomena relevant to catchment hydrology, such as orographic lift, foehn winds, and cold air pools (Reinert et al., 2026). The data presented here represents an "operational archive"; unlike a reanalysis or a reforecast set, using a fixed model version, the operational ICON-D2 system undergoes regular updates and changes to its model physics and data assimilation schemes. Therefore, the dataset reflects the available forecast at the time of issuance rather than a methodologically homogenous time series.

5.2 ICON-D2 deterministic

The deterministic run of ICON-D2 provides high-resolution, single-trajectory meteorological forecasts, representing an unperturbed model run, differing from the ICON-D2 ensemble in the use of higher resolved ICON-EU deterministic boundary conditions. It serves as a reference run and often as a "best estimate" for users who are not interested in probabilistic information. The historical forecasts were processed for all catchments in the CAMELS-DE-1h dataset, covering the period from February 2021 to December 2024. A subset of variables was selected based on their relevance for rainfall-runoff modelling and to ensure consistency with the observational datasets (RADKLIM-YW and HOSTRADA). These variables include precipitation, air temperature, global radiation, air pressure, snow water equivalent, wind speed and wind direction. A detailed description of the variables is contained in Table 2.

The ICON-D2 data consists of gridded hourly forecasts initialized every 3 hours (00:00, 03:00 ... 21:00 UTC). Prior to 23 June 2021 (09:00 UTC), the maximum lead time was 27 hours; from this date onwards, it was extended to 48 hours. To preserve a homogenous data structure, all variable values beyond a lead time of 27 hours were set to NaN for dates prior to the extension. To generate catchment-specific time series for each initialization time, the gridded data were spatially aggregated the same way as the observed time series data, using the *exactextract* (Baston, 2025) method, which calculates the fraction of each grid cell covered by the catchment polygon to ensure precise area-weighted statistics. For precipitation, the spatial minimum, maximum, and standard deviation were additionally computed to capture within-catchment variability. For the remaining variables, the spatial mean was calculated for each timestep.

Figure 6 provides an initial analysis of the deterministic ICON-D2 precipitation forecasts included in CAMELS-DE-1h, comparing forecast precipitation against RADKLIM observations across all catchments. To focus on rainy conditions, all analyses are restricted to time steps with observed precipitation exceeding 0.1 mm h^{-1}. The error distributions across lead times (Fig. 6a) show a systematic shift from overestimation at very short lead times, likely related to data assimilation influences, towards underestimation at higher lead times: at lead time +1 h, forecast precipitation is on average higher than observed, while from lead time +2 h onward there is a growing tendency to underestimate precipitation with increasing forecast horizon, likely related to an underestimation of convective heavy rainfall events. The seasonal spatial maps of relative bias (Fig. 6b–i) reveal distinct seasonal patterns: summer precipitation (JJA) is systematically underestimated across many catchments at both +3 h and +12 h lead times, likely reflecting the convective nature of summer rainfall events, which

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

are inherently difficult to capture with deterministic NWP forecasts. In contrast, winter precipitation (DJF) tends to be overestimated at short lead times. Beyond these seasonal differences, consistent regional patterns of under- and overestimation are evident across seasons (Fig. 6b–i), suggesting that systematic errors are at least partly tied to orography or regional weather regimes rather than random noise. The dependence of forecast bias on catchment area (Fig. 6j–k) reveals a clear size dependency that strengthens with lead time: small catchments (< 50 km²) exhibit the strongest underestimation of forecast precipitation. A likely explanation is that precipitation estimates in small catchments are more sensitive to spatial displacement and intensity errors, because these are averaged over a much smaller area and therefore benefit less from spatial error compensation than larger catchments. As catchment area increases, the underestimation decreases, and the largest catchments (> 1000 km²) show a tendency toward overestimation for short lead times. The increasing forecast errors with lead time (Fig. 6a), the observed spatial and seasonal patterns (Fig. 6b-i) and the systematic underestimation in small catchments (Fig. 6j-k) are consistent with known limitations of convection-permitting NWP models (Shrestha et al., 2013; Yano et al., 2018).

Table 2: Catchment-specific forecast variables available in CAMELS-DE-1h, forecasts are initialized every 3 hours and have a lead time of 48 hours. The processed deterministic run data contains precipitation as well as all other meteorological variables listed, while the ensemble run (EPS) contains precipitation variables from 20 different ensemble members.

Variable	Description	Unit	Data source
precipitation_mean precipitation_min precipitation_max precipitation_stdev	Spatial mean, minimum, maximum and standard deviation of hourly precipitation. Values represent the precipitation amount in the last (lead time) hour (ICON-D2 and ICON-D2 EPS, ~ 2.2 x 2.2 km²)	mm	ICON-D2 and ICON-D2 EPS (DWD, 2021)
air_temperature_mean	Spatial mean of air temperature measured at 2 m height (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	°C	
global_radiation_mean	Spatial mean of global radiation (instantaneous values at the start of the lead time hour), only available in ICON-D2 deterministic run	W m^{-2}	
air_pressure_sea_level_mean	Spatial mean of air pressure at sea level height (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	hPa	
snow_depth_water_eq_mean	Spatial mean of snow depth water equivalent (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	kg m^{-2}	
wind_speed_mean	Spatial mean of wind speed measured in 10 m height (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	m s^{-1}	
wind_speed_eastward_mean	Spatial mean of eastward wind speed measured in 10 m height (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	m s^{-1}	
wind_speed_northward_mean	Spatial mean of northward wind speed measured in 10 m height (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	m s^{-1}	
wind_direction_mean	Mean wind direction at 10 m height, calculated from spatially averaged wind components (eastward and northward), representing the direction of the mean wind vector (instantaneous values at the start of the lead time hour), only available in the ICON-D2 deterministic run	°	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

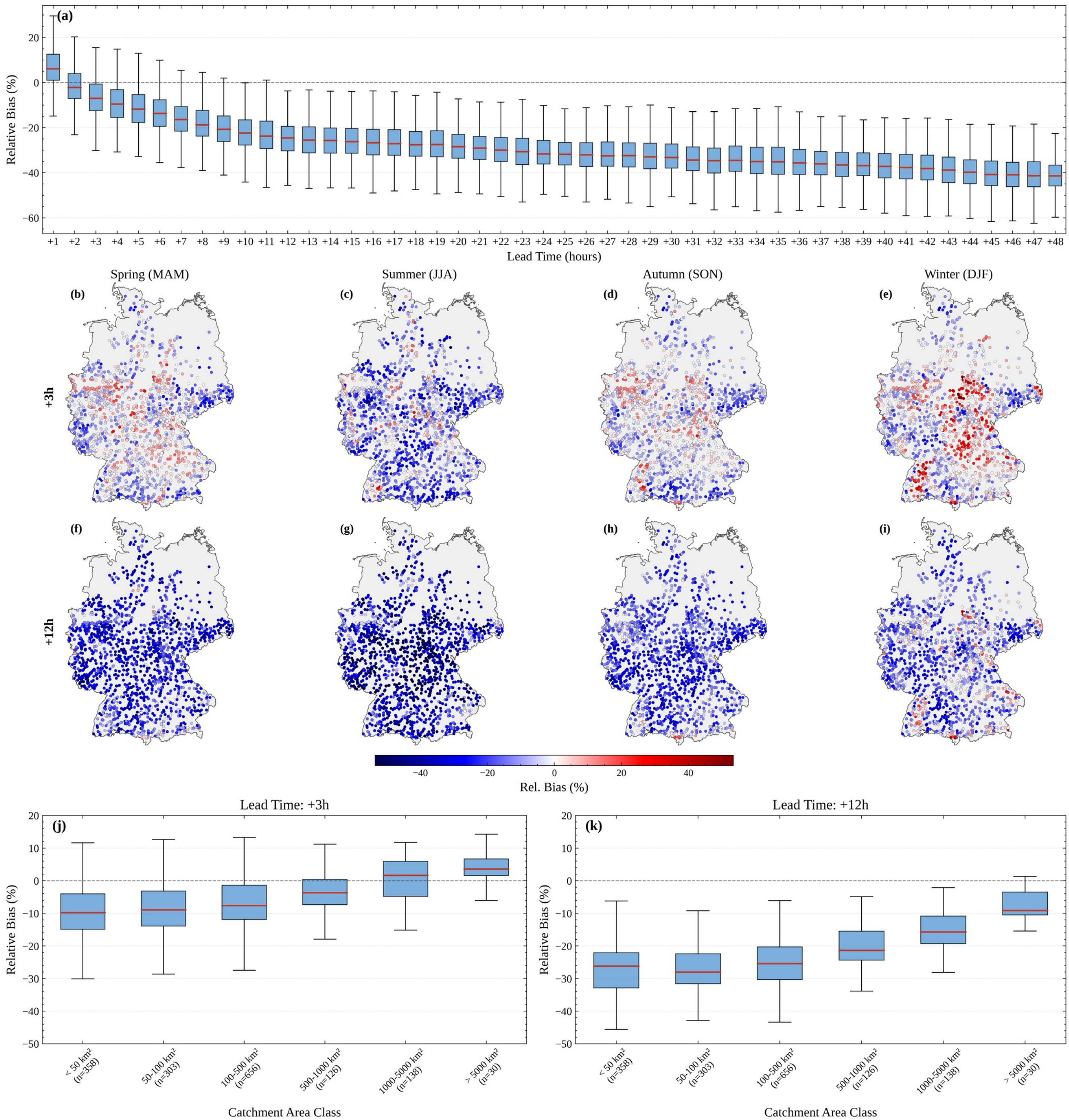

Figure 6: Relative bias of ICON-D2 deterministic precipitation forecasts evaluated against RADKLIM observations for CAMELS-DE-1h catchments (filtered to observed precipitation > 0.1 mm h^{-1}). Panel (**a**) shows the distribution of relative bias across all catchments as a
435 function of forecast lead time. Panels (**b–e**) and (**f–i**) show the spatial distribution of mean relative bias for lead times of +3 h and +12 h, respectively, stratified by season (Spring: MAM, Summer: JJA, Autumn: SON, Winter: DJF). Panels (**j–k**) show the distribution of relative bias for different catchment area classes at lead times of +3 h and +12 h, respectively. Borders of Germany: © GeoBasis-DE / BKG (2023).

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

5.3 ICON-D2 EPS (ensemble)

To allow the quantification of meteorological forecast uncertainty, the dataset includes forecasts from the ICON-D2 Ensemble Prediction System (EPS). The ensemble system runs on the same domain and convection-permitting grid (around 2.2 km resolution) as the deterministic run, forced by the coarser boundary conditions of ICON-EU EPS. The ensemble consists of 20 members designed to capture the uncertainty in the development of the atmospheric state. The spread across members is generated by the variation or perturbation of three components of the modelling system:

1. **Boundary Conditions**: Variations in the lateral boundary conditions derived from the driving ensemble (ICON-EU-EPS).
2. **Initial Conditions**: Perturbations to the initial atmospheric state.
3. **Model Physics**: Random variations in parameterizations (e.g., in microphysics).

Consistent with the deterministic run, ensemble forecasts are initialized every 3 hours (00:00, 03:00 ... 21:00 UTC). The maximum forecast horizon follows the same operational history: it was restricted to 27 hours prior to 23 June 2021 (09:00 UTC) and extended to 48 hours thereafter. Values beyond the 27-hour limit for dates prior to this extension are set to NaN. Due to data availability constraints and to manage data volume, the processed ensemble data only includes precipitation. Given that precipitation uncertainty is the primary driver of hydrological response uncertainty for small-scale flood events, the values for other meteorological variables (temperature, radiation, etc.) can typically be used from the deterministic run for most rainfall-runoff modelling applications. Unlike the deterministic run, the *exactextract* method was not employed for the ensemble forecast due to the high computational cost of processing 20 members per timestep. Instead, spatial statistics (mean, minimum, maximum, and standard deviation) were calculated using all grid cells that intersect the catchment geometry. As an exemplary illustration, Figure 7 shows the forecast at initialization time 20 October 2021 09:00 for the catchment Bad Mergentheim over the lead time of 48 hours. The figure shows the deterministic ICON-D2 forecast as well as the 20 ensemble members of the ensemble forecast together with the 10th to 90th percentile ensemble spread compared to the observed precipitation from RADKLIM-YW, highlighting the wide spread across ensemble members.

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

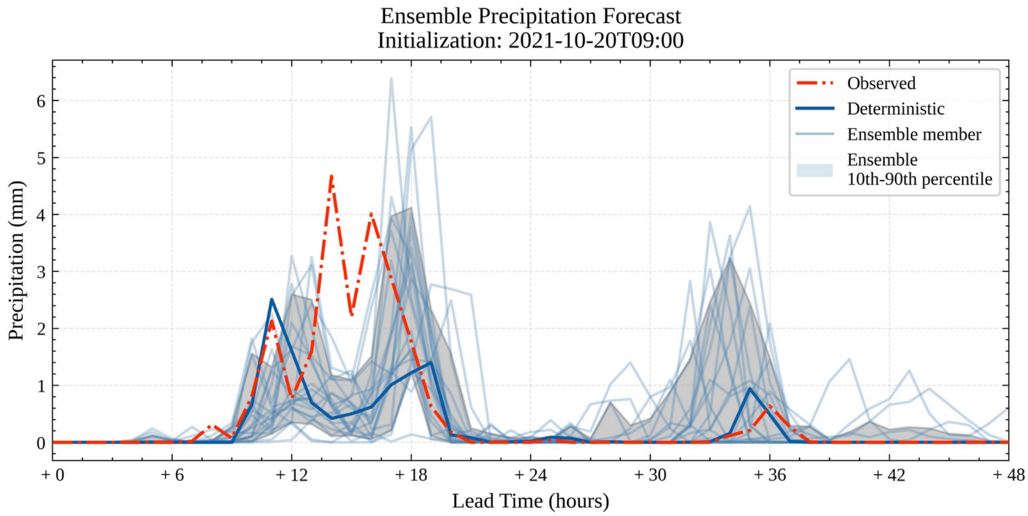

Figure 7: Exemplary visualization of meteorological forecast data at initialization time 20 October 2021 18:00 (48 hour ICON-D2 deterministic forecast and ICON-D2 20 member ensemble forecast, visualized individually and as the 10th to 90th percentile spread) as well as the observed precipitation data in CAMELS-DE-1h for the catchment DE110200 (Bad Mergentheim, river Tauber, 1017 km²).

6 Catchment attributes

To ensure full compatibility and facilitate comparative studies between the daily and hourly datasets, CAMELS-DE-1h provides the same comprehensive set of static attributes as CAMELS-DE (Loritz et al., 2024). For catchments already present in the original CAMELS-DE dataset, the attribute values were transferred directly to preserve consistency. For catchments unique to CAMELS-DE-1h, attributes were derived using the same source data and processing pipeline described in the original CAMELS-DE publication (Loritz et al., 2024; Dolich, 2024). This workflow is fully reproducible using the open-source tools provided with the dataset, allowing researchers to derive consistent attributes for additional user-specific catchments if necessary. Furthermore, maintaining a unified set of attributes across both CAMELS-DE and CAMELS-DE-1h streamlines long-term dataset maintenance and simplifies potential future updates for both datasets. While geophysical and human influence attributes (topography, land cover, soil, hydrogeology, human influences) stem from exactly the same source data, climatic and hydrological signatures were recalculated based on the high-resolution time series data in CAMELS-DE-1h.

6.1 Geophysical and human influence attributes

This category encompasses time-invariant properties describing the physical landscape and anthropogenic modifications of the catchments. All catchment attributes together with further information are listed in Table 3.

- **Location and Topography**: To address the inconsistency of official gauge IDs across different federal states, the harmonized ID system developed for CAMELS-DE is used. These IDs are based on the NUTS classification (Nomenclature of Territorial Units for Statistics), where the first level (NUTS 1) represents the federal state (e.g., DE7 for Hessen, DED for Saxony, see Fig. 2a). Each gauge ID begins with this NUTS 1 code, followed by a numerical sequence starting at 10000 and increasing in steps of 10 (e.g., DE710000, DE710010). This system ensures unique, consistent identifiers across Germany while directly encoding the federal state into the gauge ID. While this harmonized ID is the primary key for the dataset, the official IDs provided by state agencies are preserved within the topographic attributes table. Gauge name, river / waterbody name, gauge elevation and catchment area were provided as gauge metadata from the federal agencies. In addition, information whether the catchment is also part of the daily CAMELS-DE dataset and its area share outside of Germany is provided. Furthermore, catchment areas were also calculated from the catchment shapes, and the GLO-30 Digital Elevation Model (European Space Agency and Airbus, 2022) was used to derive the gauge point elevation (m) and minimum, mean, median, 5th and 95th percentiles and maximum of the catchment elevation (m). The locations of all gauging stations and their catchment boundaries are provided as both Shapefiles and GeoPackage files.

- **Land Cover**: Land cover fractions are derived from the CORINE Land Cover (CLC) 2018 dataset (European Environment Agency, 2019) with a spatial resolution of 100 m. Attributes include the percentage coverage of artificial surfaces, agricultural areas, forests, wetlands, and water bodies of each catchment.
- **Soil**: Soil properties are calculated based on the SoilGrids250m dataset (Poggio et al., 2021). Spatially aggregated values for depths of 0 - 30, 30 - 100, and 100 - 200 cm are provided for soil bulk density, organic carbon content, and texture classes (sand, silt, clay), along with the volumetric percentage of coarse fragments.
- **Hydrogeology**: Hydrogeological attributes are derived from the Hydrogeological Overview Map of Germany (BGR, 2019). These attributes quantify the areal percentage of aquifer characteristics, including permeability, porosity, rock type, and geochemical classification.
- **Human Influence**: Anthropogenic influence is primarily characterized using the "Inventory of Dams in Germany" (Speckhann et al., 2021), which provides detailed metadata for 530 dams across the country. Attributes include the total number of dams, their names, the associated rivers, the operational years of the oldest and newest dams, the total surface area and storage volume of all reservoirs at full capacity, and their primary purposes (e.g., flood control, water supply). It is important to note that this inventory is not exhaustive; the absence of recorded dams does not guarantee that there is no dam in the catchment, as smaller weirs or reservoirs may not be documented. The proportion of artificial and agricultural land cover (see Land Cover above) can also be used as indications for the extent of anthropogenic alteration in the catchment.

6.2 Climatic and hydrological signatures

The hydrological signatures for CAMELS-DE-1h were recalculated based on the hydro-meteorological hourly time series data covering the time period 2001 - 2024. The original precipitation data was used to calculate the signatures, rather than the gap-filled data. All climatic and hydrological signatures were calculated based solely on complete hydrological years (defined as 1 October to 30 September of the following year), with a maximum tolerance of 5 % missing values per

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

hydrological year, following the methodology applied in other CAMELS datasets (e.g. CAMELS-DE, CAMELS-CH, CAMELS-UK). Derived from the hourly discharge time series, the hydrological attributes include the mean specific discharge, the runoff ratio, the 5th and 95th percentiles of specific discharge, and the slope of the flow duration curve between the log-transformed 33rd and 66th percentiles. The dataset also provides indicators regarding data availability and timing, specifically the start and end dates of available discharge, the percentage of days with valid data, and the day of the year when the cumulative discharge (since 1 October) reaches half of the annual total. The occurrence of extreme events is characterized by the frequency of high-flow, low-flow, and zero-flow days, along with the average duration of high-flow and low-flow events. Climatic attributes were calculated based on the meteorological time series and include the mean precipitation, the seasonality of precipitation, and the fraction of precipitation falling as snow. To capture climatic extremes and variability, the dataset includes the frequency of high- and low-precipitation days, the average duration of high-precipitation events and dry periods, as well as the season during which the majority of high- and low-precipitation days occur. All climatic and hydrological signatures are also listed in Table 3.

Table 3: Catchment attributes available in CAMELS-DE-1h, adopted from Loritz et al. (2024). Updated some descriptions and climatic and hydrologic attributes for CAMELS-DE-1h.

Attribute class	Attribute name	Description	Unit	Data source
Location and topography	gauge_id	catchment identifier based on the NUTS classification as described in Section 6.1 e.g. DE110000, DE110010, …	–	Federal state agencies and water associations (Sect. 2)
	provider_id	official gauging station ID assigned by the federal states and water associations	–	
	gauge_name	gauging station name	–	
	water_body_name	water body name	–	
	federal_state	federal state in which the gauging station is located	–	
	area_pct_in_germany	percentage of the catchment area located outside the German border, indicating the percentage of the catchment area not covered by HOSTRADA meteorological data, maximum of 10 percent (Sect. 3)	–	
	in_camelsde_1d	binary flag indicating whether the catchment is also part of the daily CAMELS-DE dataset (Loritz et al., 2024)	%	
	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 data providers	m.a.s.l.	
	area_metadata	catchment area as given by the data providers	km^2	
	gauge_elev	gauging station elevation derived from the GLO-30 DEM	m a.s.l.	Copernicus GLO-30 DEM (European Space Agency and Airbus, 2022)
	area	catchment area derived from the MERIT Hydro catchment	km²	
	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.	
Climate	p_mean	mean hourly precipitation	mm hr^{-1}	RADKLIM-YW (Winterrath et al., 2018) and HOSTRADA (DWD, 2024)
	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	–	
	precipitation_perc_complete	percentage of hours for which precipitation is available from Jan 2001–Dec 2024, in case of missing precipitation data the gap-filled data can be used	%	
	high_prec_freq	frequency of high-precipitation hours (≥ 5 times mean	hr yr^{-1}	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Open Access Earth System Science Data Discussions

		hourly precipitation)		
	high_prec_dur	mean duration of high-precipitation events (number of consecutive hours ≥ 5 times mean hourly precipitation)	hr	
	high_prec_timing	season during which most high-precipitation hours 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 hours (< 1 mm d-1)	hr yr^{-1}	
	low_prec_dur	mean duration of dry periods (number of consecutive hours < 1 mm d-1 mean hourly precipitation)	hr	
	low_prec_timing	season during which most dry season hours occur, e.g. 'son' for autumn. If two seasons register the same number of events a value of NA is given.	season	
Hydrology	q_mean	mean hourly specific discharge	mm hr^{-1}	Federal state agencies (see Section 2) and RADKLIM-YW (Winterrath et al., 2018)
	runoff_ratio	runoff ratio (ratio of mean hourly discharge to mean hourly precipitation)	–	
	flow_period_start	first date and time for which hourly streamflow data is available	–	
	flow_period_end	last date and time for which hourly streamflow data is available	–	
	flow_perc_complete	percentage of hours for which streamflow data is available from Jan 2001–Dec 2024	%	
	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 discharge reaches half of the annual discharge)	d	
	Q5	5 % flow quantile (low flow)	mm hr^{-1}	
	Q95	95 % flow quantile (high flow)	mm hr^{-1}	
	high_q_freq	frequency of high-flow hours (> 9 times the median daily flow)	hr yr^{-1}	
	high_q_dur	mean duration of high-flow events (number of consecutive hours > 9 times the median hourly flow)	hr	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Open Access Earth System Science Data Discussions

	low_q_freq	frequency of low-flow days (< 0.2 times the mean daily flow)	hr yr^{-1}	
	low_q_dur	mean duration of low-flow events (number of consecutive hours < 0.2 times the mean hourly flow)	hr	
	zero_q_freq	fraction of hours with zero stream flow	–	
Land cover	artificial_surfaces_perc	areal coverage of artificial surfaces	%	CORINE Land Cover 2018 (European Environment Agency, 2019)
	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	%	
Soil	clay_0_30cm_mean clay_30_100cm_mean clay_100_200cm_mean	weight percent of clay particles (< 0.002 mm) in the fine earth fraction at depths 0 - 30 cm, 30 - 100 cm and 100 - 200 cm	wt. %	SoilGrids250 m (Poggio et al., 2021)
	silt_0_30cm_mean silt_30_100cm_mean silt_100_200cm_mean	weight percent 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	weight percent 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}	
Hydrogeology	aquitard_perc aquifer_perc aquifer_aquitard_mixed_perc	areal coverage of aquifer media type classes	%	HGM250 (BGR, 2019)
	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})	areal coverage of permeability classes	%	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

	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_perc (<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	%	
Human influence	dams_names	names of all dams located in the catchment	–	Inventory of dams in Germany (Speckhann et al., 2021)
	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	–	

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

	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	–	
Benchmark Simulations	training_perc_complete	percentage of observed specific discharge values in the training period (2004-01-01 01:00 – 2019-12-31 23:00) that are not NaN	%	Regional LSTM model, HBV model (Sect. 7)
	validation_perc_complete	percentage of observed specific discharge values in the validation period (2001-01-01 01:00 – 2003-12-31 23:00) that are not NaN	%	
	testing_perc_complete	percentage of observed specific discharge values in the testing period (2020-01-01 01:00 – 2024-12-31 23:00) that are not NaN	%	
	NSE_lstm	Nash-Sutcliffe model efficiency coefficient of the LSTM in the testing period, calculated from the median time series of the 3 ensemble members	–	
	NSE_hbv	Nash-Sutcliffe model efficiency coefficient of the conceptual model in the testing period, calculated from the median time series of the 3 ensemble members	–	

7 Benchmark models

CAMELS-DE-1h aims to assist in extending the field of large-sample hydrology to the sub-daily resolution through the inclusion of benchmark discharge simulation model results that can serve as a baseline for further model development. These models include a regionally trained Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) and a catchment-wise calibrated conceptual HBV (Hydrologiska Byråns Vattenbalansavdelning) model (Bergström and Forsman, 1973; Seibert, 2005), operating on an hourly timescale. Both models consist of an ensemble of three members. The models serve multiple purposes within the dataset: (a) they act as benchmarks for future model development, testing, and inter-comparison studies; (b) they function as quality control measures, helping to distinguish catchments with physically plausible behavior from those exhibiting data inconsistencies or non-causal rainfall-runoff relationships (Sect. 4.1); and (c) they provide a baseline for filling potential gaps in observed discharge time series. The provided benchmark model results cover hindcasts exclusively.

7.1 MF-LSTM setup

Training a vanilla LSTM exclusively with hourly data presents significant computational challenges. The sequence length for one year of data increases from 365 time steps (daily resolution) to 8,760 time steps (hourly resolution). This dramatically increases computational costs and can hinder model learning due to the difficulty of propagating gradients over

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

such long sequences (vanishing gradients). To circumvent this, the Multi-Frequency LSTM (MF-LSTM) architecture proposed by Acuña Espinoza et al. (2025) was developed and applied for the benchmark in CAMELS-DE-1h. The MF-LSTM processes the input sequence at multiple temporal resolutions, maintaining high resolution (hourly) for the immediate past, while aggregating older data into coarser resolutions (daily and weekly). This approach is intended to preserve critical sub-daily information for the immediate response while capturing long-term dependencies more efficiently.

The MF-LSTM for CAMELS-DE-1h was trained regionally (one model for all catchments) using specific discharge as the target variable. As dynamic inputs the model utilizes (gap-filled) mean precipitation, standard deviation of precipitation, mean air temperature, global shortwave radiation, surface air pressure, relative humidity, wind speed, and cloud cover. These dynamic inputs are aggregated to multiple frequencies as follows:

- Hourly: Applied to the last 96 steps (4 days).
- Daily: Applied to the preceding 193 steps (approx. 6 months).
- Weekly: Applied to the preceding 24 steps (approx. 6 months).

As static inputs the model uses catchment area, mean elevation, soil properties (sand, silt, clay), land cover percentages (artificial surfaces, agriculture, forest, wetlands, water bodies), and climatic indices (mean precipitation, seasonality, snow fraction, and frequency / duration of high and low precipitation events). The basin-averaged Nash-Sutcliffe efficiency (NSE) proposed by Kratzert et al. (2019) is used as the loss function. Hyperparameters are set to a hidden size of 128, a number of 20 epochs, a dropout rate of 0.4, and a batch size of 256. The learning rate was scheduled to decay from 0.0005 to 5.0e-05 over 8 epochs. The total sequence length covered one year (8760 hours), processed via the MF-LSTM custom sequence configuration described above. For the dynamic inputs, an embedding layer with a hidden size of 8 was used. The dataset was split into a training period from 1 January 2004 to 31 December 2019, a validation period from 1 January 2001 to 31 December 2003 and a testing period from 1 January 2020 to 31 December 2024. In total, an ensemble of three LSTMs using different random seeds was trained to account for uncertainty in the random initialization of initial model weights and biases.

7.2 HBV setup

The conceptual benchmark utilizes the lumped HBV model, maintaining the general model structure used for the daily HBV benchmark in CAMELS-DE. To accommodate the higher temporal resolution, several adaptations to the parameterization and process descriptions were necessary. Typical parameter ranges used for daily calibration were adapted to the hourly time step to ensure physical plausibility:

- Recession Coefficients (K_0, K_1, K_2): These parameters required geometric scaling. The hourly coefficients (K_{hr}) were derived from daily priors (K_{day}) using the relationship:

$$K_{hr} = 1 - \left(1 - K_{day}\right)^{1/24}$$

- Flux Rates (*PERC*, *CFMAX*): As these are defined in mm/step, daily values were linearly scaled by dividing by 24.

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

- Routing Scale (*beta*): The triangular weighting parameter, representing units of time, was linearly scaled by multiplying by 24 (allowing a maximum lag of 156 hours).

As HBV requires potential evapotranspiration (PET) as an input, this variable had to be estimated from the other meteorological variables included in CAMELS-DE-1h. While the daily CAMELS-DE setup used a modified Hargreaves formula based on daily extreme temperatures (Clerc-Schwarzenbach et al., 2024; Droogers and Allen, 2002), the hourly implementation must account for sub-daily variations in solar energy. Specifically, the equation was adjusted to account for local solar time based on catchment longitude and latitude, ensuring that PET is constrained (e.g., near zero at night) rather than averaged over 24 hours. The HBV model was calibrated using the DiffeRential Evolution Adaptive Metropolis (DREAM; Vrugt, 2016) algorithm as implemented in the SPOTPY library (Houska et al., 2015), consistent with the daily CAMELS-DE approach. The DREAM algorithm was employed with 7 chains and 5000 repetitions for each catchment. As in the case of the LSTM, an ensemble of three parameter sets was calibrated for the HBV by employing the DREAM algorithm with different random seeds. This reflects the issue of equifinality in hydrological modelling (Beven, 2006), where multiple parameter combinations can yield similarly good performance, and thus provides a pragmatic way to sample plausible solutions and represent parameter uncertainty. The model was calibrated individually for each catchment using the Nash-Sutcliffe Efficiency (NSE) as the objective function. For calibration, the period from 1 January 2004 to 31 December 2019 was used, while the models were tested in the period from 1 January 2020 to 31 December 2024. As the same periods are used for the LSTM for training and testing, the results are directly comparable. It should be noted that the HBV model requires a continuous time series of features and target, which is why the gap-filled precipitation time series was used (Sect. 4.2). As the target variable (specific discharge) may still contain gaps, the effective calibration period can be shorter for individual catchments, depending on the availability of continuous observations.

7.3 Results

The following section presents the simulation results of the LSTM and HBV benchmark models. The analysis evaluates the median ensemble discharge time series, derived by computing the median across the three individual ensemble members at each time step in the testing period. These median simulations were used to calculate the NSE values, which are presented here and included as benchmark statistics in the dataset (Table 3). CAMELS-DE-1h provides the simulated discharge time series for both the individual ensemble members and the ensemble median (Table 1). Additionally, the dataset includes the weights and biases of the trained LSTM alongside the HBV model parameters, enabling users to execute both models using the publicly available code (Sect. 8). To ensure sufficient data for model training and the robust calculation of catchment-specific NSE values, only catchments with at least 20 % data availability during the training period (2004 - 2019) and at least 10 % completeness during the test period (2020 - 2024) are evaluated. This filtering results in a subset of 1,496 out of the 1,611 available catchments. In this benchmark, the LSTM achieves a median NSE of 0.82, whereas the HBV model reaches a median NSE of 0.69 (Fig. 8a). Negative NSE values were produced for 35 catchments (2.3 %) by the LSTM and 134

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

catchments (9.0 %) by the HBV model. These negative NSE scores may point to underlying data quality issues (e.g., rating curve errors, missing data) or strong anthropogenic influences such as reservoir operations, interbasin transfers, or irrigation. Figure 8b illustrates model performance across the German federal states, revealing considerable spatial variability that reflects differences in data quality, as well as in hydrological and landscape settings. Both the LSTM and the HBV model exhibit their weakest performance in Brandenburg, consistent with findings from CAMELS-DE (Loritz et al., 2024). Brandenburg is characterized by low topographic relief and extensive human water management through weirs, dams, canals, and cross-connections between streams, which collectively reduce the predictability of precipitation-driven runoff. Similarly, both models struggle in Saxony-Anhalt. In Saxony, however, the LSTM achieves acceptable results despite poor HBV performance. These contrasting performances likely reflect the ability of the LSTM to learn complex, non-linear relationships within partially human-influenced systems and to compensate for localized data issues via regional training. For instance, the conceptual HBV model is sensitive to mass balance errors while the LSTM can compensate for these. In the remaining federal states, both models perform better in comparison, with a consistent pattern of the LSTM outperforming the HBV benchmark. The spatial patterns of NSE for the LSTM (Fig. 8c) and HBV (Fig. 8d) are broadly similar: higher performance is concentrated in the mountainous catchments of southern and central Germany, while lower performance is observed in the northern and eastern lowlands. Notably, an area in northern Thuringia lacks evaluated catchments despite their inclusion in CAMELS-DE-1h (Fig. 2c). This is because the supplied discharge time series for these gauges only begin in October 2018, thus failing to meet the 20 % training data minimum. Figure 8e highlights catchments across Germany with differences in performance between the HBV and the LSTM. Red dots indicate locations where the LSTM substantially outperforms the HBV model, illustrating the added value of jointly applying data-driven and conceptual approaches to better diagnose data quality limitations and characterize predictive uncertainty.

Overall, both benchmark models demonstrate that the CAMELS-DE-1h dataset is suitable for large-sample hourly rainfall-runoff modelling across Germany. The median NSE values, which are close to the NSE values of the daily simulations in CAMELS-DE, confirm the dataset's utility and high data quality. These benchmark results establish a robust baseline for further model development and future studies, such as simulating flood dynamics in small catchments or improving the representation of human-influenced systems.

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

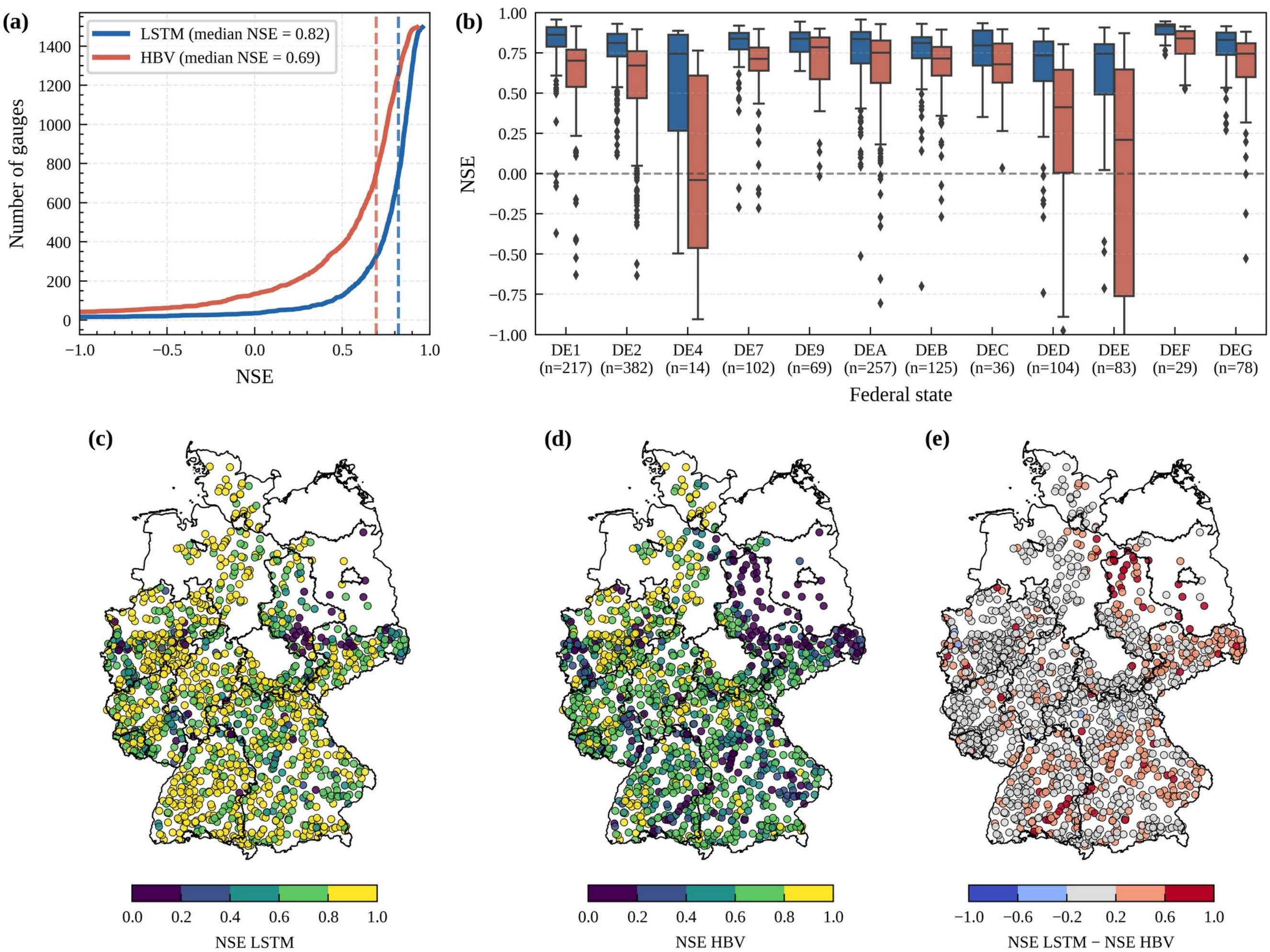

Figure 8: Nash-Sutcliffe efficiency (NSE) of the LSTM and HBV benchmark models evaluated on the test period (2020–2024) for the CAMELS-DE-1h catchments that meet the data completeness thresholds (≥20 % during training, ≥10 % during testing). Panel (**a**) shows the cumulative distribution of NSE for LSTM and HBV; vertical dashed lines mark the respective medians. Panel (**b**) shows the NSE distributions by federal state, the number of catchments per state is given in parentheses. Panels (**c**, **d**) show the spatial distribution of NSE for the LSTM and HBV models, respectively. Panel (**e**) shows the difference in NSE between LSTM and HBV (LSTM - HBV); negative values (red) indicate catchments where LSTM outperforms HBV; for the difference calculation, NSE values below zero are clipped to zero. Borders of Germany: © GeoBasis-DE / BKG (2023).

8 Code availability

The code used to process the CAMELS-DE-1h dataset, including the harmonization of raw hydrological data, metadata processing, calculation of hydrologic and climatic attributes, and compilation of the final dataset, is available via GitHub: https://github.com/CAMELS-DE/camels_de1h (last access: 15 April 2026). To efficiently process multiple terabytes of meteorological data in parallel on an HPC cluster, the Python package *stgrid2area* was developed and is available at: https://github.com/AlexDo1/stgrid2area (last access: 15 April 2026). The specific scripts used to run this package for the

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

various meteorological input datasets are published separately: https://github.com/CAMELS-DE/stgrid2area-workflows/tree/main/hpc (last access: 15 April 2026).

The LSTM benchmark model was trained and evaluated using the *Hy2DL* package, which was optimized for large volumes of hourly data: https://github.com/eduardoAcunaEspinoza/Hy2DL/tree/dev_scaling (last access: 15 April 2026). The HBV benchmark model was calibrated and tested using the *HyMC* package, which includes an adapted hourly version of the HBV model. The code utilized to calculate potential evapotranspiration using the Hargreaves formula is also provided within this repository: https://github.com/eduardoAcunaEspinoza/HyMC/tree/main (last access: 15 April 2026).

9 Data availability

CAMELS-DE-1h is available under a CC BY 4.0 license at https://doi.org/10.5880/fidgeo.2026.045 (Dolich et al., 2026).

10 Conclusion

CAMELS-DE-1h represents a large compilation of homogeneous, high-resolution hydro-meteorological catchment data for Germany. Encompassing 1,611 catchments (median area 132.4 km²), it provides the sub-daily resolution required to capture the rapid hydrological responses of small- to medium-sized basins. Furthermore, the dataset includes multiple years of catchment-processed, historical short-term (48 h) deterministic and ensemble weather forecast data. This opens new opportunities to systematically evaluate hydrological models and forecasting tools under realistic operational conditions, including the use of historical deterministic and ensemble weather forecasts.

The accompanying hourly LSTM and HBV benchmarks provide consistent reference baselines that can support future model development and enable systematic investigations of catchment behavior across Germany.

Ultimately, CAMELS-DE-1h complements the existing daily CAMELS-DE dataset, facilitating centralized access to Germany's vast hydrological data, which is otherwise heavily scattered across federal states. While CAMELS-DE provides a 70-year continuous record in daily resolution, ideal for water balance modeling and long-term trend analyses, we specifically aim to transition large-sample hydrology toward high-resolution timescales with the publication of CAMELS-DE-1h. Together, the CAMELS-DE datasets form a comprehensive, ready-to-use data foundation across multiple temporal resolutions, enabling a broad spectrum of hydro-meteorological research and substantially accelerating data acquisition for large-sample hydrological studies in Germany.

Author contributions

AD prepared and processed the data, trained and calibrated the benchmark models, created the figures, and wrote most of the article. EAE optimized the Hy2DL library to work with hourly data. AD, RL, EAE, JB and MK collected the hydrological

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Open Access Earth System Science Data Discussions

data. JM extended and provided the infilled precipitation data. All authors suggested improvements and made additions to the article, and provided expertise for specific topics.

Acknowledgements

We thank the various institutions that provided the foundational discharge data for this work; in particular, we would like to thank the personnel at these agencies for their support and insights. The authors acknowledge support by the state of Baden-Württemberg through bwHPC. This work is supported by the Helmholtz Association Initiative and Networking Fund on the HAICORE@KIT partition.

Financial Support

The results were funded by the Federal Ministry of Research, Technology and Space (BMFTR) as part of the project "AI-based flood prediction in small river basins in Germany" (KI-HopE-De, 16IS24088A). We are grateful to the Volkswagen Foundation for funding the project "Invigorating Hydrological Science and Teaching: Merging Key Legacies with New Concepts and Paradigms" (ViTamins, 0069685).

References

Acuña Espinoza, E., Kratzert, F., Klotz, D., Gauch, M., Álvarez Chaves, M., Loritz, R., and Ehret, U.: Technical note: An approach for handling multiple temporal frequencies with different input dimensions using a single LSTM cell, Hydrol. Earth Syst. Sci., 29, 1749–1758, https://doi.org/10.5194/hess-29-1749-2025, 2025.

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.

Baston, D.: exactextract, https://github.com/isciences/exactextract (last access: 12 February 2026), 2025.

Bergström, S. and Forsman, A.: Development Of A Conceptual Deterministic Rainfall-Runoff Model, Hydrol. Res., 4, 147–170, https://doi.org/10.2166/nh.1973.0012, 1973.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007, 2006.

BGR: HGM250: Hydrogeological Map of Germany 1:250,000 (HÜK250), Federal Institute for Geosciences and Natural Resources (BGR), https://gdk.gdi-de.org/geonetwork/srv/api/records/61ac4628-6b62-48c6-89b8-46270819f0d6 (last access: 24 July 2024), 2019.

BKG: Verwaltungsgebiete 1:250 000 - Stand 01.01., Federal Agency for Cartography and Geodesy (BKG), https://gdk.gdi-de.org/geonetwork/srv/api/records/93a98c5c-cf03-4a95-bf0a-54001fbf3949 (last access: 12 February 2026), 2023.

Bushra, S., Shakya, J., Cattoën, C., Fischer, S., and Pahlow, M.: CAMELS-NZ: hydrometeorological time series and landscape attributes for New Zealand, Earth Syst. Sci. Data, 17, 5745–5760, https://doi.org/10.5194/essd-17-5745-2025, 2025.

Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, Hydrol. Earth Syst. Sci., 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, 2024.

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

Coxon, G., Zheng, Y., Barbedo, R., Cooper, H., Fileni, F., Fowler, H. J., Fry, M., Green, A., Gribbin, T., Harfoot, H., Lewis, E., Neto, G. G. R., Qiu, X., Salwey, S., and Wendt, D. E.: CAMELS-GB v2: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, https://doi.org/10.5194/essd-2025-608, 27 November 2025.

Dolich, A.: CAMELS-DE Processing Pipeline, https://doi.org/10.5281/ZENODO.13842287, 2024.

Dolich, A., Acuna Espinoza, E., Ehret, U., Kraft, M., Bondy, J., Meuer, J., and Loritz, R.: CAMELS-DE-1h: hourly hydro-meteorological time series, weather forecasts, and attributes for 1611 catchments in Germany, https://doi.org/10.5880/FIDGEO.2026.045, 2026.

Droogers, P. and Allen, R. G.: Estimating Reference Evapotranspiration Under Inaccurate Data Conditions, Irrig. Drain. Syst., 16, 33–45, https://doi.org/10.1023/A:1015508322413, 2002.

DWD: ICON-D2 - Short-range numerical weather prediction model, German Weather Service (DWD), https://dwd-geoportal.de/products/G_E6D/ (last access: 12 February 2026), 2021.

DWD: HOSTRADA - High-resolution grids of hourly variables for Germany, Version 1.0, German Weather Service (DWD), https://opendata.dwd.de/climate_environment/CDC/grids_germany/hourly/hostrada/ (last access: 12 February 2026) (1.0), 2024.

Etter, S., Strobl, B., Seibert, J., and Van Meerveld, H. J. I.: Value of uncertain streamflow observations for hydrological modelling, Hydrol. Earth Syst. Sci., 22, 5243–5257, https://doi.org/10.5194/hess-22-5243-2018, 2018.

European Environment Agency: CORINE Land Cover 2018 (raster 100 m), Europe, 6-yearly - version 2020_20u1, May 2020 (20.01), https://doi.org/10.2909/960998C1-1870-4E82-8051-6485205EBBAC, 2019.

European Space Agency and Airbus: Copernicus DEM, https://doi.org/10.5270/ESA-c5d3d65, 2022.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., and Hochreiter, S.: Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network, Hydrol. Earth Syst. Sci., 25, 2045–2062, https://doi.org/10.5194/hess-25-2045-2021, 2021.

Heberger, M.: delineator.py: Fast, accurate watershed delineation using hybrid vector- and raster-based methods and data from MERIT-Hydro, , https://doi.org/10.5281/ZENODO.10143149, 2023.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

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-

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth Syst. Sci. Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

Houska, T., Kraft, P., Chamorro-Chavez, A., and Breuer, L.: SPOTting Model Parameters Using a Ready-Made Python Package, PLOS ONE, 10, e0145180, https://doi.org/10.1371/journal.pone.0145180, 2015.

Konold, O., Feigl, M., Podest, P., Klingler, C., and Schulz, K.: BiasCast: Learning and adjusting real time biases from meteorological forecasts to enhance runoff predictions, https://doi.org/10.5194/egusphere-2025-4978, 27 November 2025.

Krähenmann, S., Walter, A., Brienen, S., Imbery, F., and Matzarakis, A.: High-resolution grids of hourly meteorological variables for Germany, Theor. Appl. Climatol., 131, 899–926, https://doi.org/10.1007/s00704-016-2003-7, 2018.

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

Kreklow, J., Tetzlaff, B., Burkhard, B., and Kuhnt, G.: Radar-Based Precipitation Climatology in Germany—Developments, Uncertainties and Potentials, Atmosphere, 11, 217, https://doi.org/10.3390/atmos11020217, 2020.

Liu, J., Bian, Y., Lawson, K., and Shen, C.: Probing the limit of hydrologic predictability with the Transformer network, J. Hydrol., 637, 131389, https://doi.org/10.1016/j.jhydrol.2024.131389, 2024.

Liu, J., Shen, C., O'Donncha, F., Song, Y., Zhi, W., Beck, H. E., Bindas, T., Kraabel, N., and Lawson, K.: From RNNs to Transformers: benchmarking deep learning architectures for hydrologic prediction, Hydrol. Earth Syst. Sci., 29, 6811–6828, https://doi.org/10.5194/hess-29-6811-2025, 2025.

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

Meuer, J., Bouwer, L. M., Kaspar, F., Lehmann, R., Karl, W., Ludwig, T., and Kadow, C.: Infilling of missing rainfall radar data with a memory-assisted deep learning approach, Hydrol. Earth Syst. Sci., 29, 3687–3701, https://doi.org/10.5194/hess-29-3687-2025, 2025.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resour. Res., 57, e2020WR028091, https://doi.org/10.1029/2020WR028091, 2021.

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

Nijzink, J., Loritz, R., Gourdol, L., Zoccatelli, D., Iffly, J. F., and Pfister, L.: CAMELS-LUX: Highly Resolved Hydro-Meteorological and Atmospheric Data for Physiographically Characterized Catchments around Luxembourg, https://doi.org/10.5194/essd-2024-482, 26 May 2025.

Pilgrim, D. H., Cordery, I., and Baron, B. C.: Effects of catchment size on runoff relationships, J. Hydrol., 58, 205–221, https://doi.org/10.1016/0022-1694(82)90035-X, 1982.

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.

Pool, S., Viviroli, D., and Seibert, J.: Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration?, J. Hydrol., 554, 613–622, https://doi.org/10.1016/j.jhydrol.2017.09.037, 2017.

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C., Gebhardt, C., Marsigli, C., Förstner, J., Zängl, G., Schlemmer, L., Blahak, U., and Welzbacher, C.: DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System, Version 2.5.3, Res. Dev. DWD, 2026.

Seibert, J.: HBV Light Version 2, User's Manual, Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, 2005.

Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, Hydrol. Earth Syst. Sci., 13, 883–892, https://doi.org/10.5194/hess-13-883-2009, 2009.

Shalev, G. and Kratzert, F.: Caravan MultiMet: Extending Caravan with Multiple Weather Nowcasts and Forecasts, https://doi.org/10.48550/ARXIV.2411.09459, 2024.

Shrestha, D. L., Robertson, D. E., Wang, Q. J., Pagano, T. C., and Hapuarachchi, H. A. P.: Evaluation of numerical weather prediction model precipitation forecasts for short-term streamflow forecasting purpose, Hydrol. Earth Syst. Sci., 17, 1913–1931, https://doi.org/10.5194/hess-17-1913-2013, 2013.

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

Staudinger, M., Herzog, A., Loritz, R., Houska, T., Pool, S., Spieler, D., Wagner, P. D., Mai, J., Kiesel, J., Thober, S., Guse, B., and Ehret, U.: How well do hydrological models learn from limited discharge data? A comparison of process- and data-driven models, https://doi.org/10.5194/egusphere-2025-1076, 14 March 2025.

Vrugt, J. A.: Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, Environ. Model. Softw., 75, 273–316, https://doi.org/10.1016/j.envsoft.2015.08.013, 2016.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Walawender, E., Weigl, E., and Becker, A.: Erstellung einer radargestützten Niederschlagsklimatologie: ein Projekt der Strategischen Behördenallianz „Anpassung an den Klimawandel", Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, 2017.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigl, E., and Becker, A.: Radar climatology (RADKLIM) version 2017.002; gridded precipitation data for Germany: Radar-based quasi gauge-adjusted five-minute precipitation rate (YW) (1), https://doi.org/10.5676/DWD/RADKLIM_YW_V2017.002, 2018.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, Geophys. Res. Lett., 44, 5844–5853, https://doi.org/10.1002/2017GL072874, 2017.

Yano, J.-I., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., and Wyszogrodzki, A. A.:

https://doi.org/10.5194/essd-2026-289
Preprint. Discussion started: 30 June 2026

Scientific Challenges of Convective-Scale Numerical Weather Prediction, Bull. Am. Meteorol. Soc., 99, 699–710, https://doi.org/10.1175/BAMS-D-17-0125.1, 2018.

Zängl, G., Reinert, D., and Prill, F.: Grid refinement in ICON v2.6.4, Geosci. Model Dev., 15, 7153–7176, https://doi.org/10.5194/gmd-15-7153-2022, 2022.