https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Open Access Earth System Science Data Discussions

CAMELS-KR: Catchment attributes, meteorology, and reconstructed streamflow for large-sample hydrology in South Korea

Songyun Lee[1], Guchang Jung[1], and Kuk-Hyun Ahn[1]

[1]Department of Civil and Environmental Engineering, Kongju National University, Cheonan, South Korea

Correspondence to: Kuk-Hyun Ahn (ahnkukhyun@gm.kongju.ac.kr)

Abstract. Large-sample hydrology has increasingly relied on harmonized datasets that integrate hydrometeorological observations with catchment attributes across diverse environmental settings. Although the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) initiative has expanded rapidly worldwide, South Korea remains underrepresented despite its distinctive hydroclimatic characteristics, including a monsoon-dominated climate, steep mountainous terrain, rapid runoff generation, and extensive anthropogenic water regulation. Here, we present CAMELS-KR, the first CAMELS-style dataset developed for South Korea. CAMELS-KR provides harmonized hydrometeorological time series, catchment boundaries, and catchment attributes for 282 quality-controlled catchments distributed across the major river basins of the country. The dataset includes daily streamflow and water-level observations, catchment-scale meteorological forcing data from 1981 to 2025, and a comprehensive set of static attributes describing topography, climate, hydrology, land cover, soils, and water infrastructure. CAMELS-KR also includes reconstructed streamflow generated using a regionally trained long short-term memory (LSTM) model and a locally calibrated conceptual HBV model, together with model performance metrics and calibrated parameter sets. By providing open-access, analysis-ready hydrological data from a monsoon-dominated and highly regulated environment, CAMELS-KR fills an important geographic gap within the global CAMELS network. The dataset is expected to support comparative hydrology, prediction in ungauged basins, climate-impact assessments, and the development and benchmarking of next-generation data-driven and process-based hydrological models.

1 Introduction

Large-sample hydrology has become a central paradigm in contemporary hydrological science because it enables the investigation of hydrological processes across diverse climatic, topographic, geological, and anthropogenic settings (Gupta et al., 2014; Kuentz et al., 2017). Recent advances in hydrological modeling, comparative hydrology, prediction in ungauged basins (PUB), and data-driven approaches have increasingly relied on harmonized datasets that integrate hydrometeorological time series with catchment attributes across large numbers of basins (Brunner et al., 2021; Feng et al., 2023). Such datasets provide a foundation for understanding regional hydrological variability, testing hydrological theories, benchmarking models, and accelerating reproducible research.

The development of the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) datasets has played a pivotal role in this transition. Since the introduction of CAMELS-US (Addor et al., 2017), the CAMELS initiative has expanded rapidly to multiple countries and regions, including Great Britain (Coxon et al., 2020), Australia (Fowler et al., 2021), Brazil (Chagas et al., 2020), Chile (Höge et al., 2023), Germany (Loritz et al., 2024), Denmark (Liu et al., 2025), India (Mangukiya et al., 2025), Finland (Seppä et al., 2025), and New Zealand (Bushra et al., 2025). These datasets generally provide harmonized hydrometeorological time series together with static catchment descriptors such as climate, topography, soils, geology, land cover, and anthropogenic influences. Their standardized and analysis-ready structure has substantially reduced the effort required for data acquisition and preprocessing, thereby improving reproducibility and enabling broad international collaboration in hydrological research.

The increasing adoption of machine learning and artificial intelligence (AI) in hydrology has further elevated the importance of large-sample datasets (Nearing et al., 2024; Yoon and Ahn, 2024). Harmonized datasets such as CAMELS provide standardized benchmarks for the development and evaluation of hydrological models, including deep-learning frameworks, transfer-learning approaches, and hybrid process-informed models (Khoshkalam et al., 2023; Sun et al., 2025; Yoon et al., 2026; Yoon and Ahn, 2025). Consequently, the availability of high-quality, large-sample hydrological datasets has become increasingly important for both traditional hydrological investigations and emerging AI-driven applications.

Despite the rapid global expansion of CAMELS datasets, South Korea currently lacks a standardized national-scale dataset specifically designed for large-sample hydrological studies. This gap is particularly important because South Korea represents a distinctive hydroclimatic environment within the global CAMELS community. Unlike many existing CAMELS datasets that are dominated by snow-controlled, temperate, or semi-arid hydrological regimes, Korean catchments are strongly influenced by the East Asian summer monsoon, with approximately two-thirds of annual precipitation occurring between June and September (Noh and Ahn, 2020; Park et al., 2021). In addition, tropical cyclones frequently contribute intense rainfall events that generate severe floods and hydrological extremes (Alcantara and Ahn, 2021; Parmes et al., 2026). Combined with steep mountainous terrain, short river networks, and highly responsive catchments, these climatic conditions produce rapid runoff generation and flashy hydrological behavior (Kim et al., 2016; Lee and Ahn, 2022).

Korean river systems are also characterized by extensive anthropogenic modification. Major dams, agricultural reservoirs, channel engineering, and urban drainage systems play critical roles in water-resource management and flood mitigation (Chang et al., 2023; Hwang et al., 2021; Yoo and Ahn, 2025). Consequently, South Korea provides a valuable natural laboratory for investigating hydrological processes in monsoon-dominated, highly urbanized, and strongly regulated environments. Despite the scientific importance of these characteristics, East Asia remains relatively underrepresented within the global CAMELS network. The development of a Korean CAMELS dataset therefore provides an opportunity to broaden the hydroclimatic

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

diversity represented within the international large-sample hydrology community and to support more comprehensive interregional comparisons of hydrological behavior.

Although extensive hydrological and meteorological observations are available from Korean governmental agencies, these datasets are typically distributed across multiple platforms and formats and are not readily integrated at the catchment scale. Considerable effort is therefore required to collect, preprocess, quality-control, and harmonize the data before they can be used for hydrological analysis, comparative studies, or machine-learning applications. Similar challenges have been reported in many countries prior to the development of national CAMELS datasets (Arias-Rodriguez et al., 2023; Yin et al., 2024). Such fragmentation reduces reproducibility and creates barriers for both domestic and international hydrological research.

To address these limitations, we introduce CAMELS-KR, the first CAMELS-style hydrological dataset for South Korea. CAMELS-KR provides harmonized hydrometeorological time series and catchment attributes for 282 catchments spanning diverse climatic, physiographic, and anthropogenic conditions across the country. The dataset includes observed streamflow and water-level records, catchment-scale meteorological forcings, and static attributes describing topography, climate, hydrology, land cover, soils, and water infrastructure. Following the conventions established by previous CAMELS datasets, CAMELS-KR was designed to facilitate interoperability and comparative hydrology at both regional and global scales.

The primary objectives of CAMELS-KR are threefold. First, we present the first harmonized CAMELS dataset for South Korea by integrating quality-controlled streamflow observations, catchment-scale meteorological forcing, and comprehensive catchment attributes into a consistent, analysis-ready dataset. This unified resource provides a standardized foundation for large-sample hydrological studies and enhances the reproducibility of hydrological analyses. Second, we characterize the hydroclimatic, physiographic, and anthropogenic diversity by the 282 CAMELS-KR catchments, encompassing the broad range of environmental conditions found in the monsoon-dominated and highly regulated river basins of East Asia. Finally, we provide benchmark hydrological products, including LSTM-based reconstructed streamflow, together with comprehensive catchment attributes to support comparative hydrology, model benchmarking, and the development of AI-driven hydrological applications. The remainder of this paper describes the catchment selection procedures, hydrometeorological datasets, catchment attributes, quality-control procedures, and streamflow reconstruction methods used to develop CAMELS-KR, followed by an overview of the dataset characteristics, potential applications, and future extensions.

2 Study Area and Catchment Selection

2.1 Hydroclimatic Characteristics of South Korea

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Open Access Earth System Science Data Discussions

South Korea is located in East Asia and occupies the southern portion of the Korean Peninsula (Fig. 1). The climate of South Korea is strongly seasonal. During winter, the peninsula is influenced by cold and dry continental air masses associated with the Siberian High, whereas summer conditions are dominated by warm and humid air masses originating from the North Pacific subtropical high-pressure system (Noh and Ahn, 2020). As a result, precipitation is highly concentrated during the summer monsoon season, with nearly two-thirds of the annual precipitation occurring between June and September, while winter precipitation typically accounts for less than 10 % of the annual total (Do and Kim, 2020; Lee et al., 2023). In addition to monsoonal rainfall, typhoons passing near or across the peninsula frequently produce intense precipitation events that substantially contribute to floods and other hydrological extremes (Alcantara and Ahn, 2020; Chang et al., 2023).

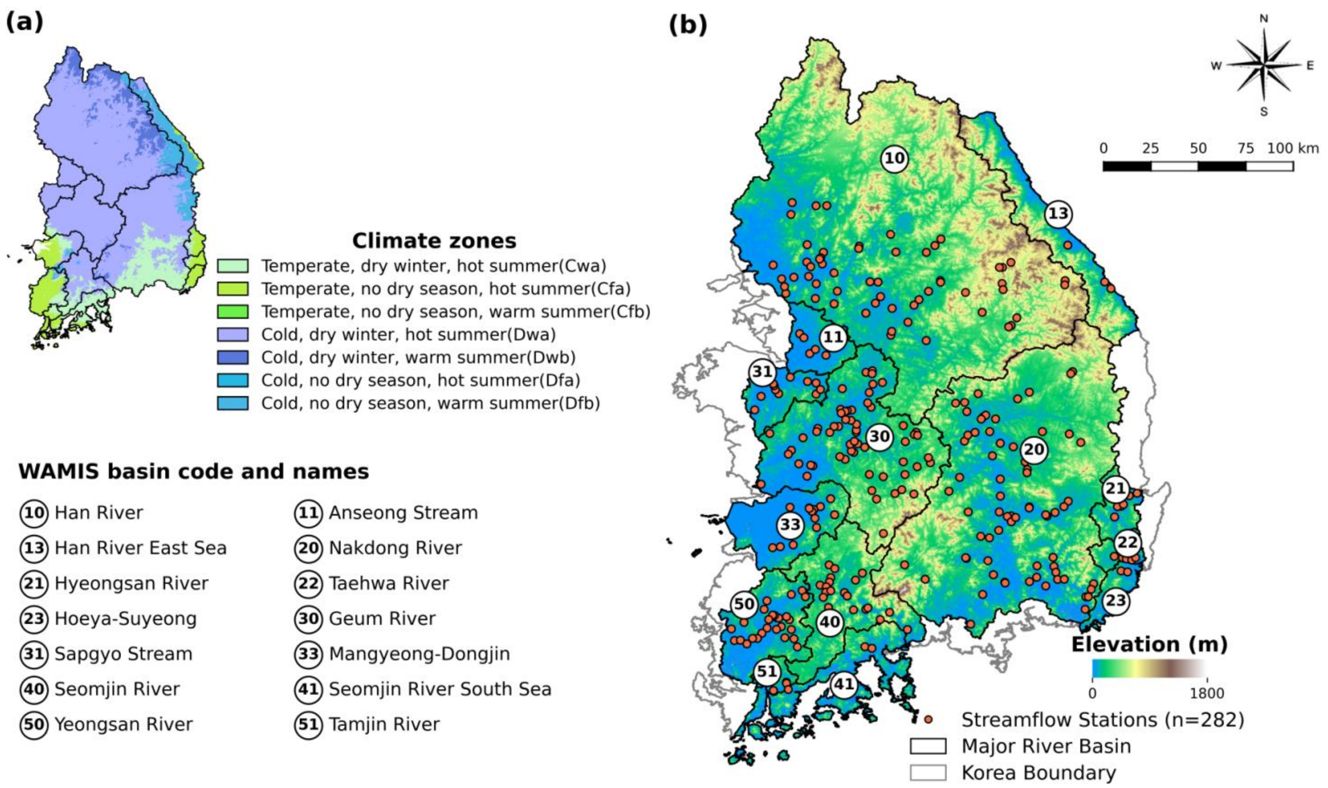


Figure 1 Overview of the study domain and monitoring network. (a) Major river basins and climate zones across South Korea. (b) Spatial distribution of the 282 gauging stations included in CAMELS-KR, shown together with topographic elevation. Figure 4 Spatial distribution of the long-term annual mean meteorological variables included in CAMELS-KR. Histograms show the frequency distributions across the 282 catchments. Panels display (a) precipitation, (b) minimum air temperature, (c) maximum air temperature, (d) mean air temperature, (e) relative humidity, (f) wind speed, (g) maximum wind speed, (h) solar radiation, (i) Penman–Monteith-based potential evapotranspiration (PET), (j) GLEAM-based potential evapotranspiration (PET), and (k) GLEAM-based actual evapotranspiration (AET). Panel (a) additionally illustrates the 0.1° × 0.1° meteorological grid system covering South Korea used for the development of the gridded forcing dataset.

Topography also exerts strong controls on the spatial distribution of hydroclimatic conditions. Mountainous regions dominate the eastern and northern parts of the country, particularly along the Taebaek Mountains, with elevations ranging from coastal lowlands near sea level to high-elevation mountainous regions exceeding 1900 m above mean sea level, whereas relatively low-lying plains and densely urbanized areas are concentrated in the western and southern regions (Tsai et al., 2025; Yun et al., 2025). Approximately 70 % of the country is mountainous, and the terrain generally slopes westward and southward toward broad alluvial plains and coastal regions. Several major river systems originate from the mountainous interior and flow toward the western and southern coastal plains. Consequently, many catchments are characterized by steep slopes, relatively short river networks, and rapid hydrological responses, often producing flashy hydrographs during extreme rainfall events. The principal river basins include the Han River, Nakdong River, Geum River, and Yeongsan River basins. Among them, the Nakdong River is the longest river in South Korea, whereas the Han River basin encompasses the Seoul metropolitan area, the largest urbanized region in the country. These river systems support multiple functions, including municipal water supply, hydropower generation, irrigation, flood control, and ecological conservation.

The combined influences of a monsoonal climate, mountainous terrain, rapid runoff generation, and extensive human regulation make South Korea a unique environment for large-sample hydrological studies. These hydroclimatic characteristics provide valuable opportunities for investigating hydrological variability, flood generation mechanisms, climate change impacts, and AI-based hydrological modeling across a wide range of catchment conditions.

2.2 Catchment Selection Criteria

For CAMELS-KR, we sourced discharge ($m^3\ s^{-1}$), water-level data (m), and metadata for 985 gauging and water level stations from multiple national agencies (see Section 3.1). We subsequently created a subset of the data by selecting only stations that satisfied several consistency criteria. Specifically, stations were required to contain all essential metadata information, including gauge name, location, and catchment area (n = 914 stations); have a catchment area larger than 20 km² (n = 850 stations); provide at least 10 years of discharge observations, although the records were not required to be continuous (n = 314 stations); and satisfy the condition that the delineated catchment area differed by less than 20 % from the value reported by the national agency (n = 310 stations). Lastly, discharge and water-level time series at these stations were manually cross-checked to identify prolonged unrealistic records and inconsistencies between the two variables, such as cases in which discharge continued to vary while the corresponding water level remained constant for more than 30 consecutive days. This screening resulted in the exclusion of 28 stations, leaving 282 stations for further analysis. These criteria were applied to ensure that the resuling dataset was sufficiently reliable for large-sample hydrological analyses. A minimum of 10 years of discharge observations was considered necessary to provide sufficiently long time series for hydrological modeling and the derivation of hydrological signatures. The minimum catchment area threshold of 20 km² was selected to improve compatibility with the 0.1° × 0.1° gridded precipitation product. Very small catchments may not be adequately represented by the spatial resolution of the forcing data and are more sensitive to uncertainties in catchment delineation and spatial averaging. The 20 % discrepancy threshold between delineated and reported catchment areas was selected to identify catchments with potentially unreliable boundaries while retaining adequate spatial coverage across the study domain (Fig. 2b).

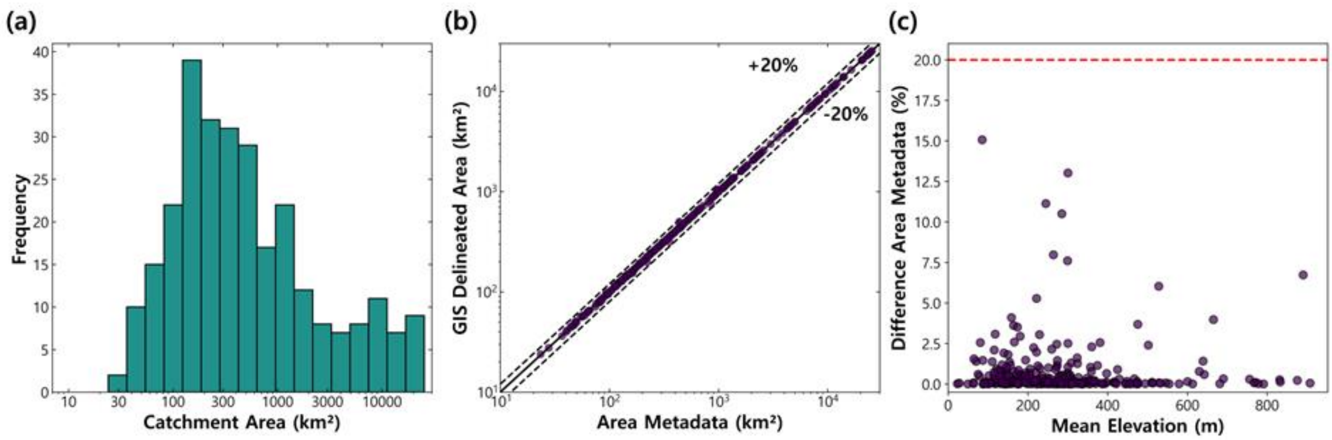


Figure 2 Evaluation of basin delineation accuracy for CAMELS-KR. (a) Distribution of basin areas on a logarithmic scale. (b) Comparison of delineated basin areas with the corresponding areas reported by the national agency for the 282 selected basins. (c) Absolute difference between delineated and reported basin areas plotted against mean basin elevation. The red line indicates the 20% area-difference threshold used to determine basin eligibility for inclusion in CAMELS-KR.

These selection procedures resulted in a final subset of 282 gauges for the CAMELS-KR dataset, providing a representative and reliable characterization of hydrological processes across South Korea (Fig. 1). We note that the majority of the selected catchments correspond to operational flood forecasting and warning points currently adopted within the Korean flood forecasting system. The selected catchments span broad ranges of drainage area, elevation, slope, precipitation, and urbanization intensity (further discussed in Section 4). In addition, some catchments are strongly influenced by anthropogenic regulation through dams, reservoirs, and urban drainage infrastructure, whereas others preserve relatively natural hydrological conditions. This diversity enables CAMELS-KR to support investigations of both natural and human-modified hydrological processes.

3 Hydrometeorological Time Series

3.1 Streamflow Data

Streamflow data and water levels were obtained from two national agencies: the Water Resources Management Information System (WAMIS; https://www.wamis.go.kr/) and the Han River Flood Control Office (HRFCO; https://www.hrfco.go.kr/). Figure 3 illustrates the spatial and temporal distribution of data availability across the hydrological monitoring stations. The main panel presents the spatial distribution of gauging stations, with colors indicating the length of available discharge records.

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Overall, 94.3 % (266 gauges) and 95.4 % (269 gauges) of the stations provide at least 15 years of discharge and water-level observations, respectively, whereas 27.7 % (78 gauges) and 60.6 % (171 gauges) contain records longer than 25 years for discharge and water level, respectively. Figures 3b and c show the temporal availability of streamflow data, indicating that the number of stations with available observations for both discharge and water level increases toward more recent years, suggesting a gradual expansion of the monitoring network and enhanced data availability over time.

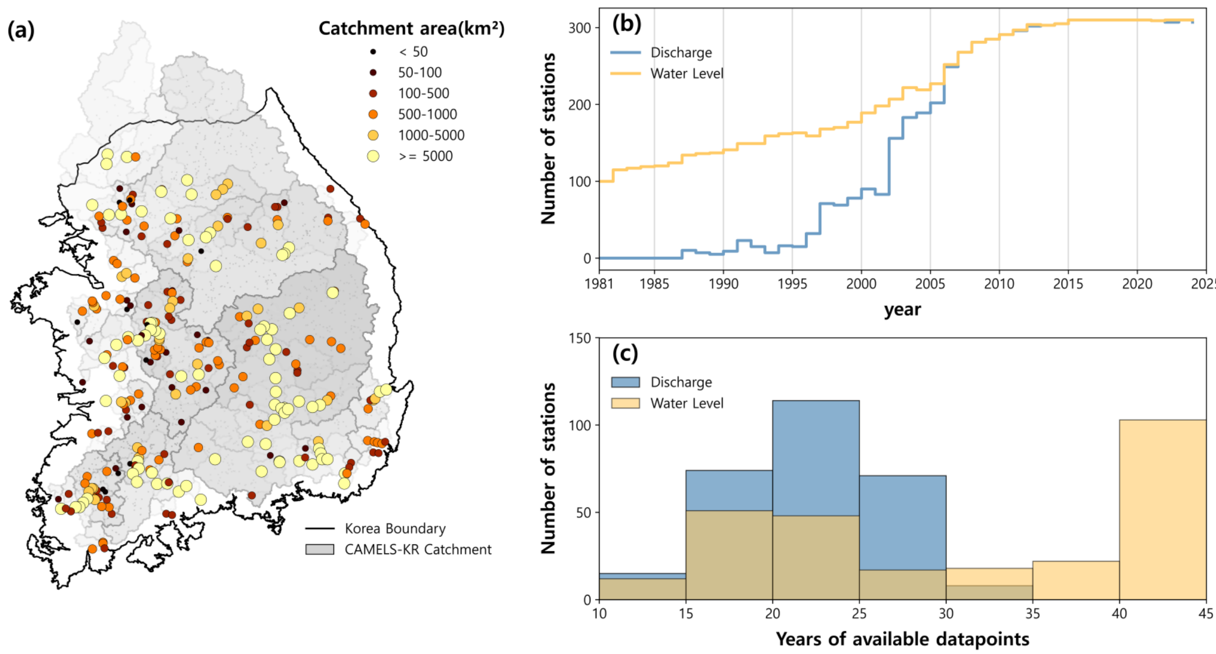


Figure 3 Overview of streamflow and water-level monitoring stations included in CAMELS-KR. (a) Spatial distribution of the 282 gauging stations and their associated catchment areas. (b) Annual number of stations with available discharge and water-level observations. (c) Frequency distribution of record lengths for discharge and water-level observations.

Comprehensive quality-control procedures were applied to all discharge and water level time series. For both discharge and water level records, characteristic low-, normal-, and high-condition statistics were cross-validated against corresponding values reported in the Korea Annual Hydrological Report (Han River Flood Control Office, 2025, 2024). Specifically, low water level (LWL), normal water level (NWL), and high water level (HWL), together with their discharge counterparts, were used to assess the consistency of the compiled observations. These evaluations provide an effective diagnostic tool because flow characteristics generally reflect long-term basin and channel conditions. Substantial deviations from the reported values may therefore indicate potential processing errors, metadata inconsistencies, or observational issues.

The CAMELS-KR dataset provides three hydrological time series for each gauging station (see Table 1): discharge volume (discharge_vol; $m^3\ s^{-1}$), specific discharge normalized by upstream drainage area (discharge_spec; $mm\ d^{-1}$), and observed

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

water level (waterlevel; m). Providing both discharge volume and specific discharge facilitates inter-basin comparisons across catchments of different sizes, while water level records offer complementary information on stage variability.

Table 1 Summary of daily time series variables in CAMELS-KR

Time series class	Abbreviation	Description	Unit	Data source
Hydrological time series	discharge_vol	Observed catchment discharge	$m^3\ s^{-1}$	Water Resources Management Information System and Han River Flood Control Office
	discharge_spec	Observed catchment-specific discharge converted to millimeters per day	$mm\ d^{-1}$	
	water_level	Observed daily water height measured relative to the station-specific zero datum	m	
Meteorological time series	prec	Observed daily summed precipitation	$mm\ d^{-1}$	Korea Meteorological Administration (2026)
	temp_min	Observed daily averaged minimum temperature	°C	
	temp_max	Observed daily averaged maximum temperature	°C	
	temp_avg	Observed daily averaged temperature	°C	
	rel_hum	Observed daily averaged relative humidity	%	
	wind_speed	Observed daily averaged wind speed measured at a standard height of 10 m above ground level.	$m\ s^{-1}$	
	wind_speed_max	Observed maximum instantaneous wind speed (wind gust)	$m\ s^{-1}$	
	solar_radiation	Total daily incoming shortwave solar energy[2]	$MJ\ m^{-2}$	
	pet	Estimated daily averaged potential evapotranspiration using FAO Penman-Monteith equation	$mm\ d^{-1}$	
	pet_gleam	Estimated daily averaged potential evapotranspiration	$mm\ d^{-1}$	GLEAM (Miralles et al., 2025)
	aet_gleam	Estimated daily averaged actual evapotranspiration	$mm\ d^{-1}$	

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Table 1 Long table continued.

Time series class	Abbreviation	Description	Unit	Data source
Simulated hydrological time series	discharge_vol_sim_lstm	Volumetric discharge reconstructed from the LSTM-simulated specific discharge and catchment area	$m^3 s^{-1}$	Regional LSTM model and HBV model (see Section 5)
	discharge_spec_sim_lstm	Catchment-specific discharge simulated by the LSTM model	$mm d^{-1}$	
	discharge_vol_sim_hbv	Volumetric discharge reconstructed from the HBV-simulated specific discharge and catchment area	$m^3 s^{-1}$	
	discharge_spec_sim_hbv	Catchment-specific discharge simulated by the HBV model	$mm d^{-1}$	
	sim_period	Flag indicating whether the daily value belongs to the training or evaluation period	-	

3.2 Meteorological Forcing Data

Meteorological forcing data for CAMELS-KR were developed following the gridded climate data construction framework proposed by Noh and Ahn (2020), which was designed to generate spatially continuous meteorological data for the Korean Peninsula. To be specific, after missing values in the station records were infilled, ordinary kriging was applied to interpolate the station observations onto a 0.1° grid. The resulting daily gridded data provide consistent hydroclimatic forcing information for all catchments included in CAMELS-KR. The dataset covers the period from 1981 to 2025 and contains 11 variables: precipitation; minimum, maximum, and mean air temperature; relative humidity; mean and maximum wind speed; solar radiation; potential evapotranspiration; and GLEAM-based potential and actual evapotranspiration.

To construct the forcing dataset, daily observations from the Automated Synoptic Observing System (ASOS) and Automatic Weather Station (AWS) networks operated by the Korea Meteorological Administration (KMA) were first collected. These station observations were then interpolated to a regular 0.1° latitude–longitude grid covering South Korea (Fig. 4a). The resulting gridded products provide spatially continuous representations of meteorological conditions while preserving the temporal variability captured by the underlying station observations. Mean air temperature was calculated as the average of daily minimum and maximum temperatures, while potential evapotranspiration was estimated using the FAO Penman–Monteith method (Sentelhas et al., 2010) based on the gridded meteorological variables. In addition to the KMA-based forcing variables, actual and potential evapotranspiration datasets were obtained from the Global Land Evaporation Amsterdam Model version 4 (GLEAM4; Miralles et al., 2025). GLEAM provides satellite-driven estimates of terrestrial evaporation processes at a spatial resolution of 0.1°, matching the resolution adopted in CAMELS-KR. Lastly, catchment-scale daily meteorological time series were calculated by averaging all 0.1° grid cells located within each catchment. The same procedure was applied consistently to all forcing variables.

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Figure 4 illustrates the spatial distribution of the long-term mean values of the eleven forcing variables, calculated from the annual averages during 1981–2025. Distinct hydroclimatic gradients are evident across South Korea. Precipitation generally increases in mountainous and southern coastal regions, whereas temperature exhibits strong elevation and latitudinal dependencies. Evapotranspiration-related variables show pronounced spatial variability associated with regional differences in energy availability and atmospheric demand (see Figs. 4i and 4k). These patterns collectively highlight the diverse climatic environments represented by the 282 CAMELS-KR catchments and demonstrate the ability of the forcing dataset to capture major hydroclimatic gradients across the country.

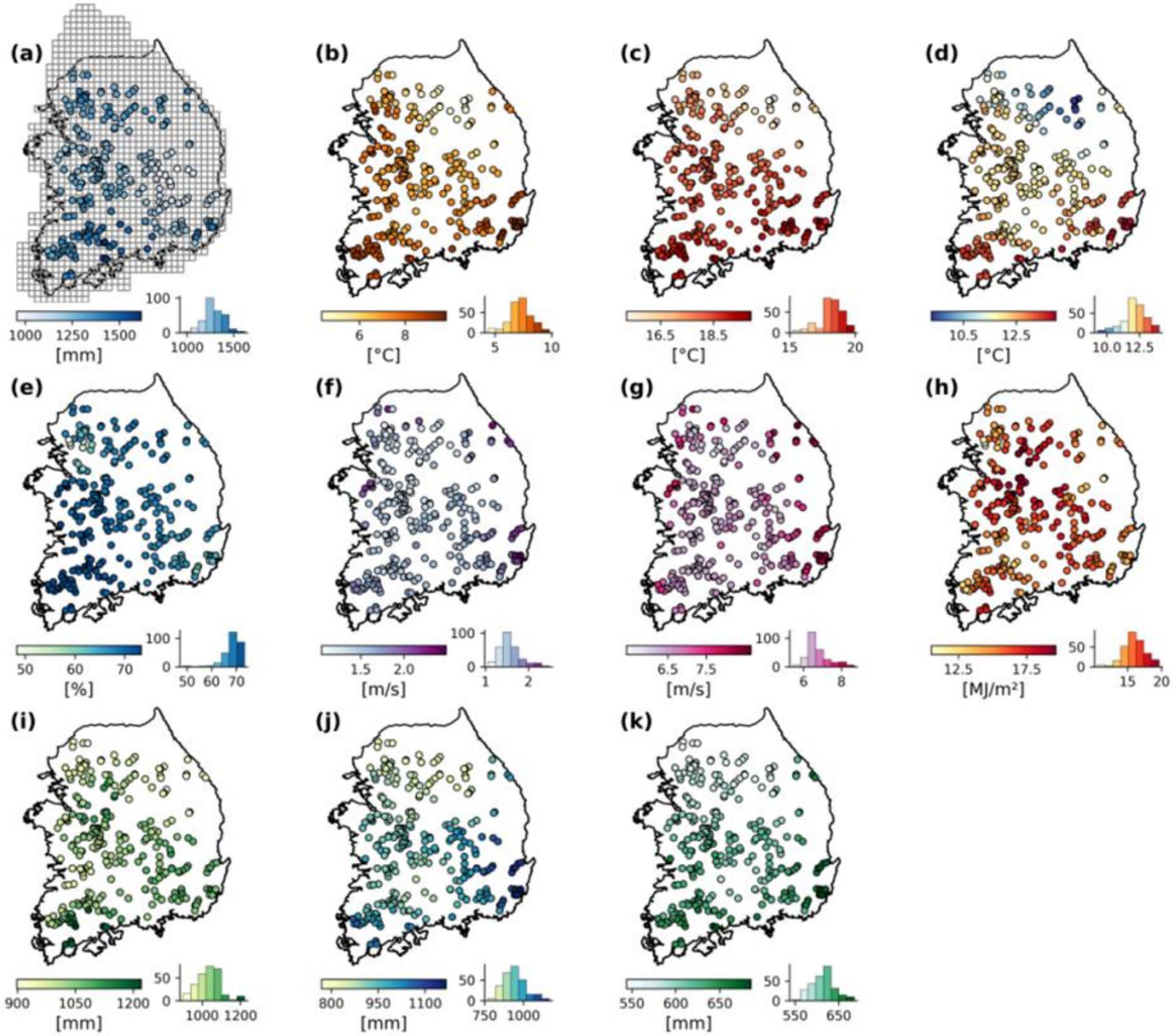


Figure 4 Spatial distribution of the long-term annual mean meteorological variables included in CAMELS-KR. Histograms show the frequency distributions across the 282 catchments. Panels display (a) precipitation, (b) minimum air temperature, (c) maximum air temperature, (d) mean air temperature, (e) relative humidity, (f) wind speed, (g) maximum wind speed, (h) solar radiation, (i) Penman–Monteith-based potential evapotranspiration (PET), (j) GLEAM-based potential evapotranspiration (PET), and (k) GLEAM-based actual evapotranspiration (AET). Panel (a) additionally illustrates the 0.1° × 0.1° meteorological grid system covering South Korea used for the development of the gridded forcing dataset.

4 Catchment Attributes

4.1 Location and Topographic Attributes

Each catchment in CAMELS-KR is assigned a unique station identifier (gauge_id) and catchment identifier (basin_id), following the official identification systems maintained by WAMIS and HRFCO. The corresponding station name (gauge_name) is also provided to facilitate linkage with the original monitoring network. Gauging station locations are reported as outlet coordinates in the World Geodetic System 1984 (WGS84; EPSG:4326), including the latitude and longitude of each catchment outlet. Catchment area (basin_area) was calculated from the basin boundaries derived following the procedures described in Section 2. Topographic attributes were derived from the 30-m Advanced Land Observing Satellite (ALOS) World 3D digital elevation model (AW3D30; Takaku et al., 2020). For each catchment, a suite of elevation statistics was computed from all DEM grid cells within the delineated catchment boundary, including minimum elevation (elev_min), 5th percentile elevation (elev_5), median elevation (elev_50), mean elevation (elev_mean), 95th percentile elevation (elev_95), and maximum elevation (elev_max). The elevation at each gauging station (gauge_elev) was also extracted from the same DEM.

In addition to the tabular attributes, the locations of all gauging stations and delineated catchment boundaries are provided as GIS-compatible shapefiles to support spatial analyses and visualization. A summary of all location and topographic attributes included in CAMELS-KR is provided in Table 2.

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Table 2 Summary of basin attributes used in CAMELS-KR

Attribute class	Abbreviation	Description	Unit	Data source
Location	gauge_id	Official gauging station ID assigned by the national agency	-	Water Resources Management Information System and Han River Flood Control Office
	basin_id	Official catchment ID assigned by the national agency	-	
	gauge_name	Gauging station name	-	
	gauge_lon	Gauging station longitude (EPSG:4326)	°	
	gauge_lat	Gauging station latitude (EPSG:4326)	°	
	basin_area	Area of the catchment	km^2	AW3D30 DEM (Takaku et al., 2020)
Topography	elev_mean	Mean elevation within catchment	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_95	95th percentile elevation within catchment	m.a.s.l.	
	elev_max	Maximum elevation within catchment	m.a.s.l.	
	gauge_elev	Gauging station elevation	m.a.s.l.	
Climate	prec_mean	Long-term mean of daily precipitation during 1981–2020.	mm d^{-1}	Daily observations from the Automated Synoptic Observing System (ASOS) and Automatic Weather Station (AWS) networks
	prec_seasonality	Amplitude of the fitted annual precipitation cycle	-	
	temp_mean	Long-term mean of daily average temperature during 1981–2020.	°C	
	pet_mean	Long-term mean of daily potential evapotranspiration during 1981–2020 using FAO Penman-Monteith equation	mm d^{-1}	
	aridity	Ratio of long-term mean potential evapotranspiration to long-term mean precipitation during 1981–2020	-	
	frac_snow	Fraction of total precipitation occurring on days with mean temperature below 0 °C during 1981–2020	-	
	high_prec_freq	Frequency of high-precipitation days, defined as days with precipitation greater than five times the long-term mean daily precipitation	d yr^{-1}	

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Table 2 Long table continued.

Attribute class	Abbreviation	Description	Unit	Data source
	high_prec_dur	Mean duration of consecutive high-precipitation events	d	
	high_prec_timing	Season during which high-precipitation days occur most frequently	season	
	low_prec_freq	Frequency of low-precipitation days, defined as days with precipitation below 1 mm	d yr^{-1}	
	low_prec_dur	Mean duration of consecutive low-precipitation events	d	
	low_prec_timing	Season during which low-precipitation days occur most frequently	season	
Hydrology	q_mean	Mean daily specific discharge	mm d^{-1}	Water Resources Management Information System and Han River Flood Control Office
	flow_record	Number of days with available discharge observations	d	
	slope_fdc	Slope of the flow duration curve between the 33rd and 66th exceedance percentiles of log-transformed daily discharge	-	
	high_q_freq	Frequency of high-flow days with discharge exceeding nine times the median daily discharge	d yr^{-1}	
	high_q_dur	Mean duration of consecutive high-flow events	d	
	low_q_freq	Frequency of low-flow days, defined as days with discharge lower than 20% of the mean daily discharge	d yr^{-1}	
	low_q_dur	Mean duration of consecutive low-flow events	d	
Land cover	forest_perc	Percentage of catchment area classified as forest land	%	Environmental Geographic Information Service
	agri_perc	Percentage of catchment area classified as agricultural land	%	
	urban_perc	Percentage of catchment area classified as urbanized land	%	

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Table 2 Long table continued.

Attribute class	Abbreviation	Description	Unit	Data source
	grass_perc	Percentage of catchment area classified as grassland land	%	
	wetland_perc	Percentage of catchment area classified as wetlands	%	
	bare_perc	Percentage of catchment area classified as bare land	%	
	water_perc	Percentage of catchment area classified as water areas	%	
	dom_land_cover	Dominant land-cover class occupying the largest proportion within catchment	%	
Soil	clay_0_30cm_mean clay_30_100cm_mean clay_100_200cm_mean	Mean clay particles at depths of 0–30, 30–100, and 100–200 cm	wt%	SoilGrids 2.0 dataset (Poggio et al., 2021)
	silt_0_30cm_mean silt_30_100cm_mean silt_100_200cm_mean	Mean silt particles at depths of 0–30, 30–100, and 100–200 cm	wt%	
	sand_0_30cm_mean sand_30_100cm_mean sand_100_200cm_mean	Mean sand particles at depths of 0–30, 30–100, and 100–200 cm	wt%	
	coarse_fragments_0_30cm_mean coarse_fragments_30_100cm_mean coarse_fragments_100_200cm_mean	Mean volumetric fraction of coarse fragments at depths of 0–30, 30–100, and 100–200 cm	vol%	
	bulk_density_0_30cm_mean bulk_density_30_100cm_mean bulk_density_100_200cm_mean	Mean soil bulk density at depths of 0–30, 30–100, and 100–200 cm	kg m^{-3}	
	soil_organic_carbon_0_30cm_mean soil_organic_carbon_30_100cm_mean soil_organic_carbon_100_200cm_mean	Mean soil organic carbon content at depths of 0–30, 30–100, and 100–200 cm	g kg^{-1}	

275

Table 2 Long table continued.

Attribute class	Abbreviation	Description	Unit	Data source
Human influence	dam_reserv_count	Number of dams and reservoirs within catchment	-	Han River Flood Control Office and Rural Agricultural Water Resources Information System
	dam_names	Names of dams	-	
	reserv_names	Names of reservoirs	-	
	dam_storage_capacity	Total designed storage capacity of all dams within the catchment	$\times 10^6$ m^3	
	dam_effective_storage	Total effective storage capacity of all dams within the catchment available for water supply and operational management	$\times 10^6$ m^3	
	dam_flood_control_storage	Total flood-control storage capacity of all dams within the catchment	$\times 10^6$ m^3	
	major_basin	Name of major river basin in which the catchment is located	-	
	intermediate_basin	Name of the intermediate river basin in which the catchment is located	-	
	sub_basin	Name of the sub-basin in which the catchment is located	-	
	reserv_area_total	Total surface area of all reservoirs located within the catchment	km^2	
	reserv_volume_total	Total storage volume of all reservoirs located within the catchment	$\times 10^6$ m^3	
	inhabitants_2000	population of the catchment in 2000	persons	Statistics Korea (2024)
	dens_2000	population density of the catchment in 2000	persons km^{-2}	
	inhabitants_2010	population of the catchment in 2010	persons	
	dens_2010	population density of the catchment in 2010	persons km^{-2}	
	inhabitants_2020	population of the catchment in 2020	persons	
	dens_2020	population density of the catchment in 2020	persons km^{-2}	
	inhabitants_2024	population of the catchment in 2024	persons	
	dens_2024	population density of the catchment in 2024	persons km^{-2}	

4.2 Climatic and Hydrologic Attributes

Long-term climatic and hydrologic attributes were calculated to characterize the hydroclimatic conditions and streamflow behavior of each CAMELS-KR catchment. To ensure consistency with previous CAMELS datasets, the majority of climatic and hydrologic attributes were derived using the original CAMELS attribute framework developed by Addor et al. (2017) and

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

its subsequent implementations in CAMELS-CH, CAMELS-DE, and CAMELS-FI. The calculations were performed using daily basin-averaged meteorological forcing data and observed streamflow records described in Section 3.

The climatic attributes describe the long-term characteristics of precipitation, temperature, and atmospheric water demand. These include mean daily precipitation (prec_mean), precipitation seasonality (prec_seasonality), mean daily air temperature (temp_mean), mean daily potential evapotranspiration (pet_mean), and the aridity index (aridity). Snow-related hydroclimatic conditions were quantified using the fraction of precipitation falling as snow (frac_snow). In addition, several precipitation indices characterize the magnitude, seasonality, persistence, and timing of wet and dry conditions within each catchment. Specifically, these indices include the frequency of high-precipitation days (high_prec_freq), average duration of high-precipitation events (high_prec_dur), season of maximum high-precipitation occurrence (high_prec_timing), frequency of low-precipitation days (low_prec_freq), average duration of low-precipitation periods (low_prec_dur), and season of maximum low-precipitation occurrence (low_prec_timing).

Hydrologic attributes were derived from the daily observed streamflow records and summarize the long-term flow regime of each catchment. The selected signatures include mean daily specific discharge (q_mean) and the total length of available discharge observations (flow_record). Flow variability was characterized using the slope of the flow duration curve (slope_fdc). High-flow and low-flow behavior were quantified using the frequency and average duration of high-flow events (high_q_freq and high_q_dur) and low-flow events (low_q_freq and low_q_dur), respectively. Together, these metrics provide information on flow variability and the occurrence and persistence of hydrological extremes.

Definitions and units of all climatic and hydrologic attributes are provided in Table 2. The resulting attribute set enables direct comparison between CAMELS-KR and other CAMELS datasets while providing a comprehensive characterization of the diverse hydroclimatic regimes found across South Korea.

4.3 Land Cover and Soil Attributes

Land-cover attributes were derived from the 2019 national Land Cover Map provided by the Environmental Geographic Information Service (EGIS; https://aid.mcee.go.kr/). The dataset classifies land cover into seven major categories: urban areas, agricultural land, forests, grassland, wetlands, barren land, and water bodies. The proportional coverage of each land-cover class was extracted within the delineated catchment boundaries. The resulting attributes include the fractions of forest (forest_perc), agricultural land (agri_perc), urban area (urban_perc), grassland (grass_perc), wetlands (wetland_perc), barren land (bare_perc), and water bodies (water_perc). In addition, the dominant land cover class (dom_land_cover) was determined as the category occupying the largest proportion of each catchment.

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Open Access Earth System Science Data Discussions

Soil attributes were derived from the SoilGrids 2.0 dataset (Poggio et al., 2021), which provides globally consistent estimates of soil properties at a spatial resolution of 250 m and six standard depth intervals. SoilGrids is generated using machine-learning models trained on a large collection of global soil observations combined with environmental covariates describing climate, topography, vegetation, geology, and hydrological conditions. For each catchment, mean values of soil texture fractions (clay, silt, and sand), coarse fragment content, bulk density, and soil organic carbon were extracted from the SoilGrids layers. Following the approach adopted in recent CAMELS datasets, soil properties were aggregated into three representative depth ranges (0–30 cm, 30–100 cm, and 100–200 cm) using depth-weighted averages of the original SoilGrids layers. These soil attributes characterize the physical and hydraulic properties of catchments for understanding differences in water storage, infiltration capacity, and runoff response across South Korea.

4.4 Human Influence Attributes

Water infrastructure attributes were developed to characterize anthropogenic water storage and flow regulation within each catchment. Information on major dams was obtained from the national dam inventory maintained by the HRFCO, while agricultural reservoir information was collected from the Rural Agricultural Water Resources Information System (https://rims.ekr.or.kr/) operated by the Korea Rural Community Corporation (KRC). Together, these datasets represent the major engineered water-storage facilities operating in South Korea.

Spatial overlay analysis was used to associate dams and agricultural reservoirs with each CAMELS-KR catchment. Catchment-scale summary statistics were then derived to quantify the extent of hydraulic infrastructure on hydrological processes. The resulting attributes include the number of water-storage facilities (dam_reserv_count), names of major dams and reservoirs within the catchment (dam_reserv_names), total dam storage volume (dam_storage_capacity), total effective dam storage (dam_effective_storage), and total flood-control storage volume for dams (dam_flood_control_storage). In addition, the major, intermediate, and sub-basin classifications (major_basin, intermediate_basin, and sub_basin) were included to provide hydrological and administrative context. In our dataset, these attributes serve as indicators of anthropogenic modification of the natural flow regime and water-resource management within each catchment. In addition, catchment population and population density are provided for four reference years (2000, 2010, 2020, and 2024), enabling assessments of temporal changes in human influence across the catchments. Together, these attributes provide a comprehensive characterization of anthropogenic influences on catchment hydrology, facilitating comparative analyses of human impacts and water-resource management across South Korea.

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

5 Streamflow Reconstruction

5.1 Motivation

In addition to observed hydrometeorological data and catchment attributes, CAMELS-KR provides simulated streamflow generated using a regionally trained long short-term memory (LSTM) model (Hochreiter and Schmidhuber, 1997) and a locally calibrated conceptual HBV (Bergström, 1995). These two complementary approaches represent data-driven and process-based rainfall–runoff modeling frameworks and provide additional hydrological products for all CAMELS-KR catchments. Among several conceptual rainfall–runoff models evaluated during model development, including GR4J (Perrin et al., 2003) and HyMod model (Boyle et al., 2000), the HBV model was ultimately selected because it consistently achieved superior predictive performance across the study catchments. Moreover, its explicit representation of soil moisture dynamics, groundwater storage, and runoff routing offers greater process complexity and structural diversity than the simpler GR4J and HyMod formulations. The inclusion of a model with a contrasting process structure enhances the robustness and hydrological realism of the simulated streamflow products provided in CAMELS-KR.

The simulated streamflow products serve three primary purposes. First, model performance can be used as a diagnostic indicator of dataset quality and catchment complexity. Poor performance may indicate uncertainties in forcing data, streamflow observations, catchment delineation, or strong anthropogenic influences. Second, the simulations provide benchmark datasets for future hydrological and AI-based modeling studies using CAMELS-KR. Third, for catchments with satisfactory model performance, the simulated streamflow can support the reconstruction of missing observations and long-term hydrological analyses.

5.2 Model Setup

Both the LSTM and HBV models were trained using observations from 1 January 1981 to 31 December 2015 and evaluated over an independent period spanning 1 January 2016 to 31 December 2025. For consistency, the mean squared error (MSE) was adopted as the optimization objective for both models. The LSTM was trained regionally across all CAMELS-KR catchments using basin-averaged meteorological forcing data and catchment attributes, whereas the HBV model was calibrated independently for each catchment using observed streamflow records. For the LSTM, the hyperparameters were set as follows: one hidden layer, a hidden size of 128, a learning rate of 0.001, a dropout rate of 0.4, a batch size of 256, a sequence length of 365 d, and the Adam optimization algorithm.

HBV requires potential evapotranspiration (PET) as an input variable. Two PET products available in CAMELS-KR, namely Penman–Monteith-based PET (pet) and GLEAM-based PET (pet_gleam), were evaluated during model development. Preliminary experiments showed that simulations using Penman–Monteith-based PET achieved slightly higher predictive

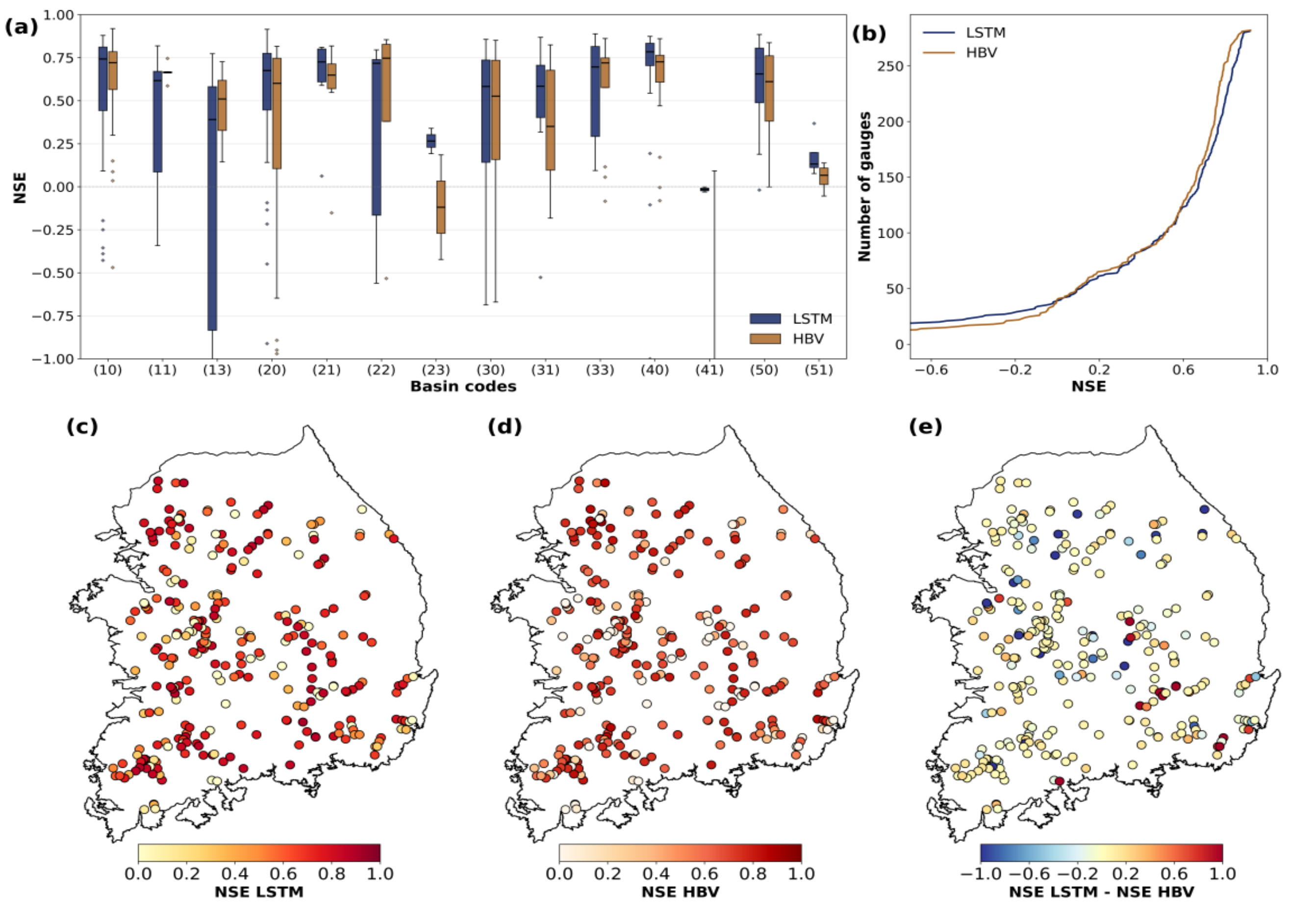
https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

performance than those using GLEAM-based PET across the study catchments. Consequently, this product was selected as the PET input for the final HBV simulations.

In addition to the reconstructed streamflow time series, CAMELS-KR provides model performance metrics for both models. Furthermore, the calibrated HBV parameter sets and simulated streamflow generated by both models are included to facilitate reproducibility, benchmarking, and future hydrological modeling studies.

5.3 Model Evaluation

The performance of the two models across the 282 CAMELS-KR catchments is summarized in Fig. 5. Overall, the regional LSTM model achieved a median NSE of 0.670 during the evaluation period, with 65.2 % of catchments exhibiting NSE values greater than 0.5 and 43.3 % exceeding 0.7. Negative NSE values were obtained in only 35 catchments, corresponding to 13.8 % of the dataset. The HBV model achieved a median NSE of 0.641, with 66.3 % and 40.4 % of catchments exceeding NSE thresholds of 0.5 and 0.7, respectively.



https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Figure 5 Performance of the regional LSTM network and the locally calibrated HBV model across the 282 CAMELS-KR catchments during the evaluation period spanning January 2016 to December 2025. Panel (a) displays boxplots of the Nash–Sutcliffe efficiency (NSE) for the LSTM network (blue) and the HBV model (yellow) grouped by major river basin. Panel (b) shows the cumulative distribution functions of NSE for the two models across all catchments. Panels (c) and (d) display the spatial distributions of NSE for the LSTM network and the HBV model, respectively. Panel (e) displays the spatial distribution of the NSE difference, where positive values indicate superior performance of the LSTM network and negative values indicate superior performance of the HBV model.

Both models achieved satisfactory predictive performance across a wide range of hydroclimatic and physiographic conditions. However, relatively low performance was observed in several catchments located within the Seomjin River South Sea, Tamjin River, and Heoya-Suyeong basins (Fig. 5a). The reduced performance in these predominantly coastal catchments may reflect complex hydrological and hydraulic processes that are not fully captured by rainfall–runoff models. Streamflow observations in these basins can be influenced not only by upstream rainfall-driven runoff generation but also by downstream boundary conditions, including tidal fluctuations, backwater effects, and, where applicable, estuarine hydraulic processes. In addition, many coastal catchments are characterized by low-relief topography and substantial anthropogenic modification, such as agricultural reservoirs and artificial drainage systems, which can further obscure the natural rainfall–runoff relationship. The regional LSTM model generally outperformed the HBV model in these basins, likely because it was trained using information from all CAMELS-KR catchments and can implicitly learn complex nonlinear relationships. In contrast, HBV employs a simplified conceptual representation of hydrological processes with time-invariant parameters, which may limit its ability to represent dynamic changes. Nevertheless, the spatial patterns of model performance were broadly consistent, with catchments exhibiting good performance in one model generally also showing satisfactory performance in the other. Overall, the performance evaluation suggests that both models capture the dominant rainfall–runoff dynamics across most CAMELS-KR catchments and provide useful supplementary streamflow products for future hydrological analyses.

6 Dataset Characteristics and Discussion

The CAMELS-KR dataset encompasses 282 catchments distributed across the major river basins of South Korea, spanning a wide range of physiographic and climatic conditions. Catchment areas range from small headwater basins to large river systems, while elevations extend from lowland coastal regions to mountainous catchments located along the Taebaek and Sobaek mountain ranges. The dataset captures substantial variability in topographic characteristics, land-cover composition, and climatic conditions, providing a representative sample of hydrological environments across the Korean Peninsula. Figure 6 summarizes the distributions of selected catchment attributes, including drainage area, mean elevation, aridity index, precipitation seasonality, percentage of agricultural land, and dominant land-cover type. Forests constitute the dominant land-cover class in most catchments, reflecting the predominantly mountainous landscape of South Korea, whereas the proportion of agricultural land cover exhibits considerable variability among catchments. The precipitation seasonality values highlight the strong concentration of annual precipitation during the summer monsoon period, a defining feature of the Korean hydroclimate. Together, these characteristics demonstrate the diverse environmental conditions represented within CAMELS-

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Open Access Earth System Science Data Discussions

KR and underscore its suitability for large-sample hydrological analyses, comparative hydrology, regionalization studies, and machine-learning applications.

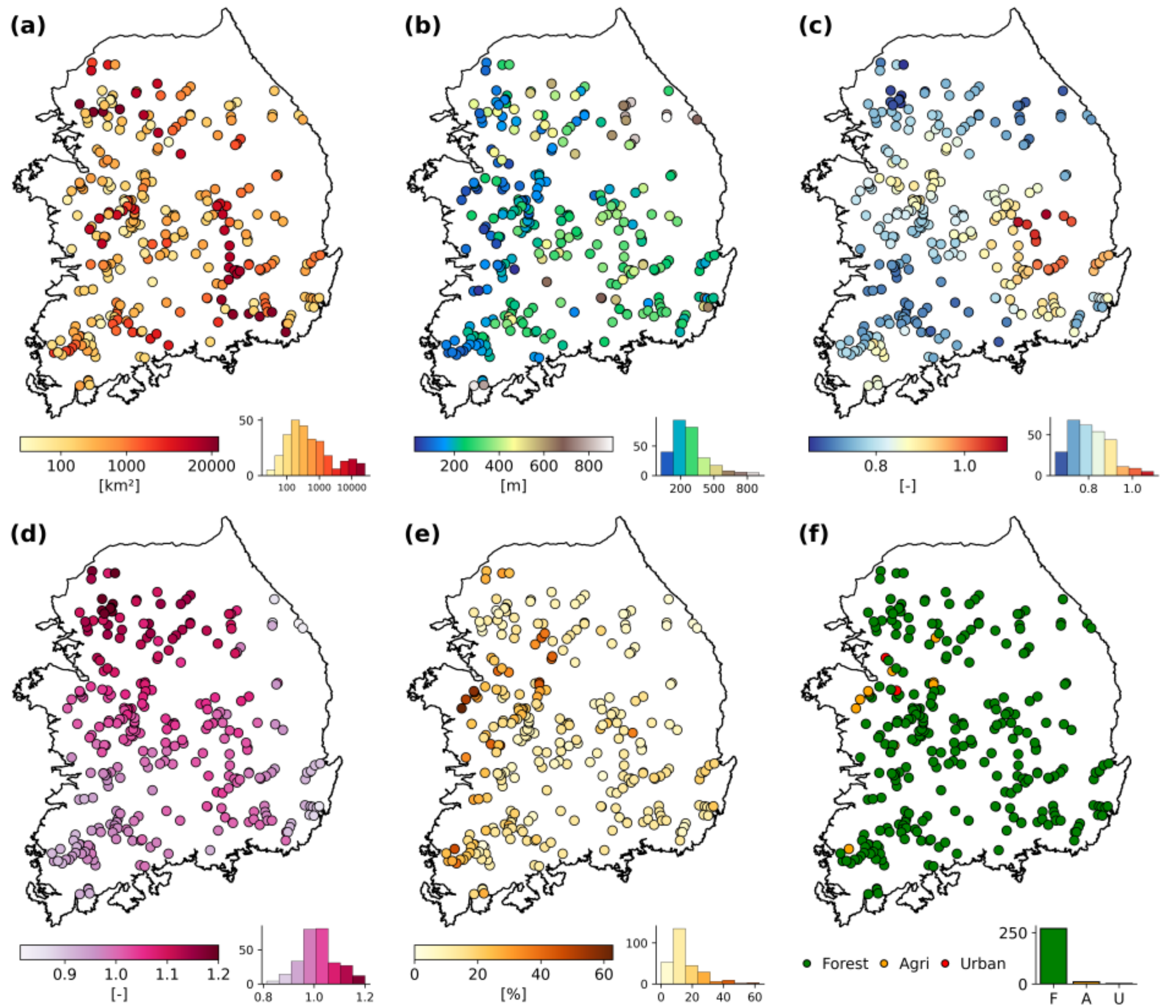


Figure 6 Spatial distributions of representative catchment attributes included in CAMELS-KR. Histograms show the frequency distributions across the 282 catchments. Panels display (a) drainage area, (b) mean elevation, (c) aridity index, (d) precipitation seasonality, (e) percentage of agricultural land, and (f) dominant land-cover type.

Figure 7 further illustrates the hydroclimatic diversity represented within CAMELS-KR using key climatic indices, including mean annual precipitation, mean annual potential evapotranspiration (PET), and the seasonality index. Catchments located in

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

southern and mountainous regions generally receive greater annual precipitation than those situated in northwestern and inland lowland regions. This pattern primarily reflects the combined influence of the East Asian summer monsoon and orographic precipitation associated with the major mountain ranges. In contrast, PET exhibits a pronounced latitudinal and elevational gradient, with higher values in warmer southern regions and lower values in high-elevation catchments. The precipitation seasonality index exhibits lower values over mountainous catchments despite their high annual precipitation. This pattern reflects the fact that these catchments receive substantial precipitation throughout the year from multiple hydroclimatic mechanisms, including the East Asian summer monsoon and orographic precipitation, as well as winter snowfall in high-elevation regions. Consequently, annual precipitation is distributed more evenly across seasons than in lowland catchments. Figure 7b presents the distribution of CAMELS-KR catchments in Budyko space. Most catchments fall within the theoretical water and energy limits, demonstrating the overall consistency of the forcing data, streamflow observations, and catchment delineations. The broad distribution across Budyko space reflects considerable variability in long-term water partitioning and catchment functioning, highlighting the influence of climate, topography, and land cover on hydrological responses.

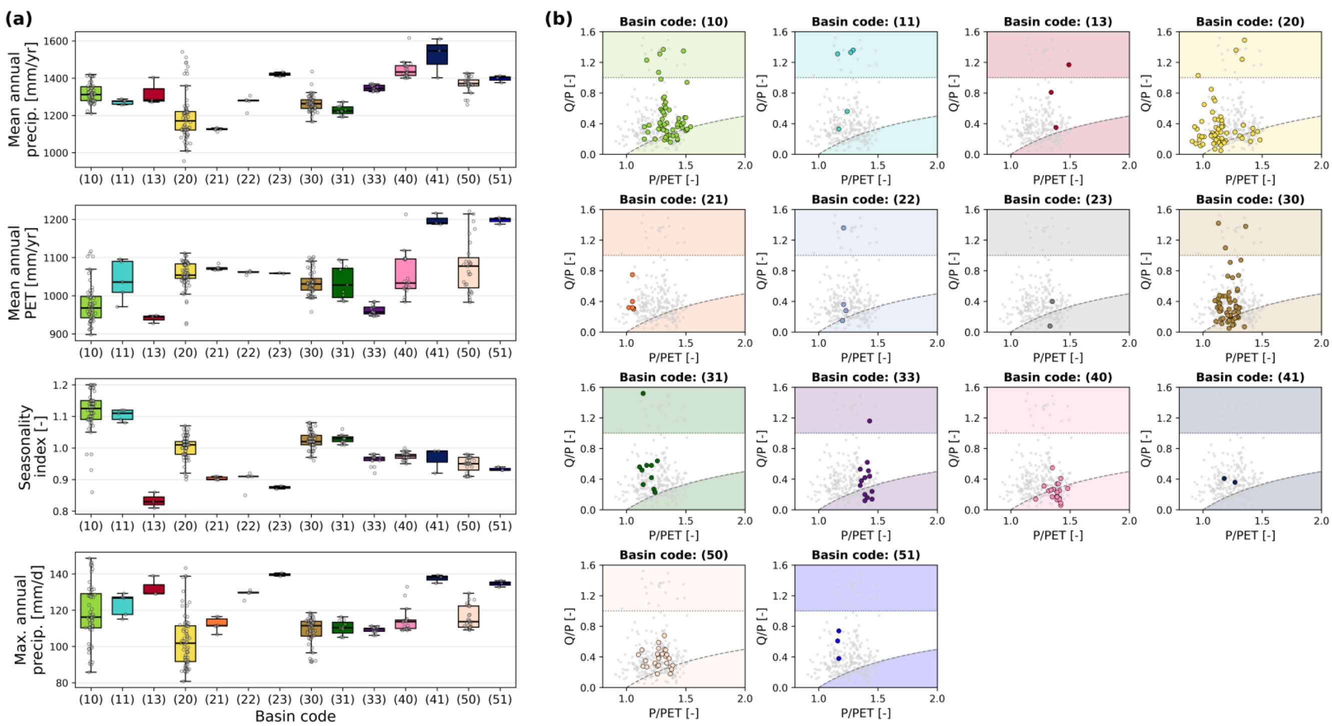


Figure 7 Climatic characteristics of the CAMELS-KR catchments across the major river basins. Panels display (a) boxplots of selected climatic attributes for each major river basin (see Fig. 1 for basin locations), and (b) distribution of catchments in Budyko space. In (b), large colored symbols denote catchments within the given major river basin, whereas small gray symbols represent catchments from all other basins.

In addition to natural catchment characteristics, CAMELS-KR provides detailed information on anthropogenic water infrastructure, including major dams and agricultural reservoirs. Figure 8 illustrates the spatial distribution of water infrastructure density across the catchments. Many Korean river basins are influenced by varying degrees of flow regulation associated with multipurpose dams, water-supply reservoirs, and agricultural storage systems. The density and storage capacity of these facilities vary substantially among catchments, indicating considerable heterogeneity in human influences on streamflow regimes. The widespread presence of engineered water infrastructure reflects the extensive development and management of water resources in South Korea. While some catchments remain relatively natural, others are strongly influenced by reservoir operations and water-management practices, resulting in a broad gradient of anthropogenic regulation. The inclusion of water infrastructure attributes enables investigations into the effects of human interventions on hydrological processes, model transferability, and the performance of both process-based and data-driven hydrological models. Together, the diverse physiographic, climatic, and anthropogenic characteristics represented in CAMELS-KR provide a valuable benchmark dataset for comparative hydrology, regionalization studies, and next-generation hydrological modeling in monsoon-dominated environments.

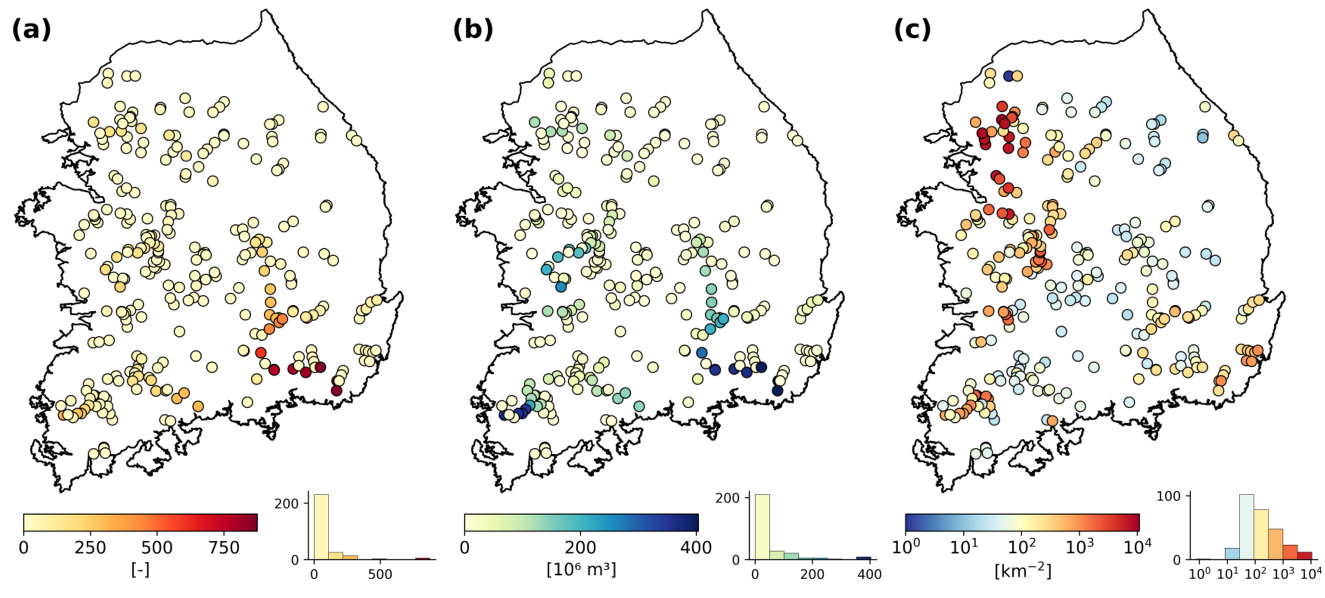


Figure 8 Spatial distribution of human influence-related attributes included in CAMELS-KR. Panels display (a) the number of dams and reservoirs, (b) the total storage capacity of reservoirs, and (c) catchment population density in 2024.

7 Possible Future Extensions

The CAMELS-KR dataset was designed as a flexible and extensible framework for large-sample hydrological studies in South Korea. The standardized data structure and reproducible processing workflow facilitate the incorporation of additional data sources and future updates without requiring substantial modifications to the existing dataset architecture. As national hydrological monitoring networks continue to expand and new environmental datasets become available, CAMELS-KR can be readily updated to provide enhanced spatial coverage and extended temporal records.

One important direction for future development is the integration of groundwater information. Groundwater plays a critical role in sustaining streamflow during dry periods and influences hydrological responses across many Korean catchments. Future versions of CAMELS-KR may incorporate groundwater-level observations from the National Groundwater Information Center (GIMS; https://www.gims.go.kr/) as well as groundwater-storage estimates derived from satellite products such as the Gravity Recovery and Climate Experiment (GRACE). The inclusion of groundwater information would enable more comprehensive investigations of surface–subsurface interactions and improve assessments of drought and water-resource vulnerability.

Another promising extension is the incorporation of additional water-management and water-quality information. While the current version includes major dams and agricultural reservoirs, future releases may expand to include operational reservoir records, water withdrawals, return flows, and water-quality observations obtained from the Water Environment Information System (WEIS; https://water.nier.go.kr/) operated by the National Institute of Environmental Research (NIER). The inclusion of water-quality variables, such as nutrient concentrations, suspended solids, and other key indicators of river condition, would broaden the applicability of CAMELS-KR beyond hydrological analyses to integrated water quantity–quality studies. Together, these additions would support investigations of anthropogenic influences on hydrological and biogeochemical processes and facilitate interdisciplinary research linking hydrology, water resources, and aquatic environmental management.

The reconstructed streamflow products included in CAMELS-KR also provide opportunities for future development. As advances in machine learning and hybrid modeling continue, additional streamflow products generated using emerging deep-learning architectures or physics-informed models could be incorporated into future releases. These products would provide valuable benchmarks for model intercomparison and uncertainty assessment. Taken together, we encourage the hydrological community to contribute additional datasets, derived products, and methodological improvements to future versions of CAMELS-KR. Continuous expansion and community-driven development will enhance the scientific value of the dataset and strengthen its role as a benchmark resource for comparative hydrology and data-driven hydrological modeling in East Asian monsoon environments.

8 Data availability

The CAMELS-KR dataset is freely available at Zenodo under https://doi.org/10.5281/zenodo.21930882 (Lee et al., 2026). The dataset includes (i) catchment boundaries for the 282 delineated catchments, (ii) daily meteorological forcing data, including

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Open Access Earth System Science Data Discussions

precipitation, temperature, humidity, wind speed, solar radiation, vapor pressure, sunshine duration, potential evapotranspiration, and actual evapotranspiration, (iii) daily streamflow and water-level observations, (iv) catchment attributes describing topography, climate, hydrology, land cover, soils, and water infrastructure, (v) reconstructed streamflow products generated using LSTM and HBV models, and (vi) documentation files describing the dataset structure, variables, units, and processing procedures. All data files are provided in comma-separated value (CSV) format, while catchment boundaries are distributed as shapefiles. A comprehensive README file is included to facilitate data access and interpretation.

9 Conclusions

CAMELS-KR represents the first comprehensive large-sample hydrological dataset developed for South Korea. The dataset includes 282 quality-controlled catchments associated with 282 hydrological gauging stations and provides daily hydroclimatic time series, catchment boundaries, and a diverse set of catchment attributes describing topography, climate, hydrology, land cover, soils, and water infrastructure. In addition, CAMELS-KR includes reconstructed streamflow products generated using both a regionally trained LSTM model and a locally calibrated HBV model, providing benchmark datasets for both data-driven and process-based hydrological modeling.

The selected catchments span a broad range of physiographic and hydroclimatic conditions across the Korean Peninsula, encompassing substantial variability in climate, topography, land cover, and human influences. In particular, South Korea represents a distinctive hydrological environment characterized by a monsoon-dominated climate, steep mountainous terrain, short and responsive river systems, rapid urbanization, and extensive water-resource regulation through dams and agricultural reservoirs. These characteristics distinguish Korean catchments from many regions represented in existing CAMELS datasets and provide valuable opportunities for investigating hydrological processes in highly dynamic and human-influenced environments.

By providing open access to harmonized hydrometeorological data, catchment attributes, and reconstructed streamflow products, CAMELS-KR is expected to support a wide range of hydrological applications, including comparative hydrology, regionalization studies, prediction in ungauged basins, climate-impact assessments, and the development and benchmarking of next-generation AI-based hydrological models. The inclusion of detailed water-infrastructure attributes further enables investigations of anthropogenic influences on catchment behavior and hydrological predictability.

More broadly, CAMELS-KR fills an important geographic gap within the global CAMELS initiative by providing standardized hydrological information from monsoon-dominated, highly urbanized, and strongly regulated catchments that have been underrepresented in existing large-sample hydrological datasets. We anticipate that CAMELS-KR will serve as a valuable

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

community resource for comparative hydrology, machine-learning hydrology, and climate-impact research, while contributing to a more comprehensive understanding of hydrological processes across diverse environmental settings worldwide.

Author contributions

S.L. developed the dataset, conducted the analyses, and led the manuscript preparation under the supervision of K.H.A. S.L. performed the geospatial analyses, and G.J. carried out the streamflow reconstruction analyses. All authors contributed to the interpretation of the results, reviewed the manuscript, and approved the final version.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00208210).

References

Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci. 21, 5293–5313.

Alcantara, A.L., Ahn, K.-H., 2021. Future flood riverine risk analysis considering the heterogeneous impacts from tropical cyclone and non-tropical cyclone rainfalls: Application to daily flows in the Nam River Basin, South Korea. Adv. Water Resour. 103983.

Alcantara, A.L., Ahn, K.-H., 2020. Probability Distribution and Characterization of Daily Precipitation Related to Tropical Cyclones over the Korean Peninsula. Water 12, 1214.

Arias-Rodriguez, L.F., Tüzün, U.F., Duan, Z., Huang, J., Tuo, Y., Disse, M., 2023. Global water quality of inland waters with harmonized landsat-8 and sentinel-2 using cloud-computed machine learning. Remote Sens. 15, 1390.

Brunner, M.I., Slater, L., Tallaksen, L.M., Clark, M., 2021. Challenges in modeling and predicting floods and droughts: A review. Wiley Interdiscip. Rev. Water 8, e1520.

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

https://doi.org/10.5194/essd-2026-544
Preprint. Discussion started: 18 August 2026

Chagas, V.B., Chaffe, P.L., Addor, N., Fan, F.M., Fleischmann, A.S., Paiva, R.C., Siqueira, V.A., 2020. CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. Earth Syst. Sci. Data Discuss. 2020, 1–41.

Chang, M., Ho, C.-H., Ho, J., Cha, E.-J., 2023. Centennial analysis in tropical cyclone-induced precipitation in Korea. Weather Clim. Extrem. 39, 100549.

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

Do, H., Kim, J., 2020. Long-term change of warm-season precipitation climatology in South Korea, in: EGU General Assembly Conference Abstracts. p. 20963.

Feng, D., Beck, H., Lawson, K., Shen, C., 2023. The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. Hydrol. Earth Syst. Sci. 27, 2357–2373.

Fowler, K.J., Acharya, S.C., Addor, N., Chou, C., Peel, M.C., 2021. CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia. Earth Syst. Sci. Data Discuss. 2021, 1–30.

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

Han River Flood Control Office, 2025. Korea Annual Hydrological Report 2024.

Han River Flood Control Office, 2024. Korea Annual Hydrological Report 2023.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M.G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A.E., others, 2023. CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland. Earth Syst. Sci. Data Discuss. 2023, 1–46.

Hwang, H.-T., Lee, E., Berg, S.J., Sudicky, E.A., Kim, Y., Park, D., Lee, H., Park, C., 2021. Estimating anthropogenic effects on a highly-controlled basin with an integrated surface-subsurface model. J. Hydrol. 603, 126963.

Khoshkalam, Y., Rousseau, A.N., Rahmani, F., Shen, C., Abbasnezhadi, K., 2023. Applying transfer learning techniques to enhance the accuracy of streamflow prediction produced by long Short-term memory networks with data integration. J. Hydrol. 622, 129682.

Kim, J.-S., Kang, H.-W., Son, C.-Y., Moon, Y.-I., 2016. Spatial variations in typhoon activities and precipitation trends over the Korean Peninsula. J. Hydro-Environ. Res. 13, 144–151.

Kuentz, A., Arheimer, B., Hundecha, Y., Wagener, T., 2017. Understanding hydrologic variability across Europe through catchment classification. Hydrol. Earth Syst. Sci. 21, 2863–2879.

Lee, D.-G., Ahn, K.-H., 2022. Assessment of Suitable Gridded Climate Datasets for Large-Scale Hydrological Modelling over South Korea. Remote Sens. 14, 3535.

Lee, S., Jung, G., Ahn, K.-H., 2026. CAMELS-KR: Catchment attributes, meteorology, and reconstructed streamflow for large-sample hydrology in South Korea (version1.1). https://doi.org/10.5281/zenodo.21930882

Lee, Y., Peñuela, A., Pianosi, F., Rico-Ramirez, M.A., 2023. Catchment-scale skill assessment of seasonal precipitation forecasts across South Korea. Int. J. Climatol. 43, 5092–5111.

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

Loritz, R., Dolich, A., Acuña Espinoza, E., Ebeling, P., Guse, B., Götte, J., Hassler, S.K., Hauffe, C., Heidbüchel, I., Kiesel, J., others, 2024. CAMELS-DE: hydro-meteorological time series and attributes for 1555 catchments in Germany. Earth Syst. Sci. Data Discuss. 2024, 1–30.

Mangukiya, N.K., Kumar, K.B., Dey, P., Sharma, S., Bejagam, V., Mujumdar, P.P., Sharma, A., 2025. CAMELS-IND: hydrometeorological time series and catchment attributes for 228 catchments in Peninsular India. Earth Syst. Sci. Data 17, 461–491.

Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E., others, 2025. GLEAM4: global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. Sci. Data 12, 416.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., others, 2024. Global prediction of extreme floods in ungauged watersheds. Nature 627, 559–563.

Noh, G.-H., Ahn, K.-H., 2020. New gridded rainfall dataset over the Korean peninsula: Gap infilling, Reconstruction, and Validation. Int. J. Climatol.

Park, C., Son, S.-W., Kim, H., Ham, Y.-G., Kim, J., Cha, D.-H., Chang, E.-C., Lee, G., Kug, J.-S., Lee, W.-S., others, 2021. Record-breaking summer rainfall in South Korea in 2020: Synoptic characteristics and the role of large-scale circulations. Mon. Weather Rev. 149, 3085–3100.

Parmes, H.O., Alcantara, A., Son, S., Ahn, K.-H., 2026. Rapid intensification restructures the upper tail of tropical-cyclone storm surge extremes. Environ. Res. Lett. 21. https://doi.org/10.1088/1748-9326/ae798b

Pellarin, T., Román-Cascón, C., Baron, C., Bindlish, R., Brocca, L., Camberlin, P., Fernández-Prieto, D., Kerr, Y.H., Massari, C., Panthou, G., others, 2020. The Precipitation Inferred from Soil Moisture (PrISM) near real-time rainfall product: Evaluation and comparison. Remote Sens. 12, 481.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289.

Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240.

Sentelhas, P.C., Gillespie, T.J., Santos, E.A., 2010. Evaluation of FAO Penman–Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. Agric. Water Manag. 97, 635–644.

Seppä, I., Gonzales Inca, C., Uusikivi, J., Alho, P., 2025. CAMELS-FI: hydrometeorological time series and landscape properties for 320 catchments in Finland. Earth Syst. Sci. Data Discuss. 2025, 1–33.

Sun, M., Yan, B., Zhou, X., Chang, J., Du, S., 2025. Hybrid-GR4J: A hybrid hydrological model integrating GR4J and deep learning. Environ. Model. Softw. 106636.

Tsai, C.-L., Kim, K., Park, Heechul, Park, Hongmok, Bang, W., Lee, G., 2025. Association between upstream conditions and the intensity of orographic precipitation in the main mountain ranges of South Korea. Earth Space Sci. 12, e2024EA003989.

Yin, Z., Lin, P., Riggs, R., Allen, G.H., Lei, X., Zheng, Z., Cai, S., 2024. A synthesis of Global Streamflow Characteristics, Hydrometeorology, and Catchment Attributes (GSHA) for large sample river-centric studies. Earth Syst. Sci. Data 16, 1559–1587.

Yoo, S., Ahn, K.-H., 2025. Exploring event-based calibration approaches for semi-distributed urban hydrologic models in data scarce urban watersheds under climate change. J. Hydrol. 660, 133498.

Yoon, S., Ahn, K.-H., 2025. Exploring the influence of training sampling strategies on time-series deep learning model in hydrology. J. Hydrol. 653, 132774.

Yoon, S., Ahn, K.-H., 2024. Self-training approach to improve the predictability of data-driven rainfall-runoff model in hydrological data-sparse regions. J. Hydrol. 130862.

Yoon, S., Kim, D., Ahn, K.-H., 2026. A heterogeneous weighting strategy for leveraging Cross-Basin data enhances the Usability of deep learning hydrological models. J. Hydrol. 135097.

Yun, S.-L., McCloskey, T.A., Cohen, M.C., Liu, K., Shin, H.-S., Ryu, J., 2025. Natural and anthropogenic factors controlling organic carbon storage in riverine wetlands along South Korea’s four rivers. Sci. Rep. 15, 539.