https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

CAMELS-PE: Hydrometeorological time series and catchment attributes for 136 catchments in Peru

Harold Llauca[1,2], Cristian Montesinos-Caceres[1,2], Max Gutierrez-Reynaga[1], Waldo Lavado-Casimiro[1,2]

[1]National Service of Meteorology and Hydrology of Peru, Lima, 15072, Peru.
[2]Doctoral Program in Water Resources, National Agrarian University La Molina, Lima 15024, Peru.

Correspondence to: Harold Llauca (hllauca@senamhi.gob.pe)

Abstract. Large-sample hydrological datasets are essential for advancing hydrological understanding and modelling across diverse environments, yet they remain scarce in South America, particularly in tropical Andean regions with strong climatic and physiographic gradients. Here, we present CAMELS-PE v1.0.1, a large-sample hydrological dataset for Peru that provides daily hydrometeorological time series and catchment attributes for 136 catchments. The dataset includes observed and simulated streamflow, meteorological forcing variables, geospatial layers, and attributes describing topography, climate, hydrological behaviour, land cover, geology, soils, and human intervention. All variables were generated under a consistent workflow involving temporal harmonisation, catchment-scale aggregation, and standardised formatting, with dedicated screening applied to observed streamflow records. The resulting dataset was evaluated through consistency checks across metadata and catchment attributes, together with plausibility analyses of regional hydroclimatic patterns. By capturing Peru's pronounced environmental contrasts, CAMELS-PE expands the representation of tropical Andean and Amazonian headwater catchments within the CAMELS framework and provides an open benchmark dataset for hydrological modelling, regionalisation, climate–streamflow analysis, prediction in ungauged basins, and machine-learning applications. CAMELS-PE is publicly available through Zenodo at https://doi.org/10.5281/zenodo.21195425 (Llauca et al., 2026) and is supported by the RCamelsPE R package.

1 Introduction

Large-sample hydrological datasets have become fundamental resources for advancing hydrological science, enabling comparative analyses across diverse climatic, physiographic, and environmental conditions (Addor et al., 2020). Over recent decades, initiatives such as the Model Parameter Estimation Experiment (MOPEX), LArge-SaMple DAta for Hydrology and Environmental Sciences (LamaH), Catchment Attributes for Brazil (CABra), the Caravan global community dataset for large-sample hydrology, and the Reference Observatory of Basins for INternational hydrological climate change detection (ROBIN) have provided harmonised hydrometeorological datasets to support hydrological modelling, regionalisation, climate impact assessment, and prediction in ungauged basins (Almagro et al., 2021; Duan et al., 2006; Klingler et al., 2021; Kratzert et al., 2023; Turner et al., 2025). These efforts have helped establish standardised and reproducible data structures for benchmarking hydrological models and analysing hydrological behaviour across large catchment samples.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Among these efforts, the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS) framework has become a widely adopted approach because of its standardised structure and broad applicability for comparative hydrological analyses. The availability of CAMELS-type datasets has supported regionalisation studies, machine-learning approaches, and assessments of hydrological model transferability across contrasting hydroclimatic regions, particularly for runoff prediction in ungauged or poorly gauged catchments (Fathi and Awadallah, 2025). Following the release of the original CAMELS dataset for the United States (Addor et al., 2017), several regional CAMELS datasets were subsequently developed worldwide, including datasets for Great Britain (Coxon et al., 2020) , Australia (Fowler et al., 2021), Switzerland (Höge et al., 2023), and France (Delaigue et al., 2025), thereby expanding the representation of regional hydroclimatic diversity in large-sample hydrological studies.

In South America, CAMELS datasets have mainly focused on Chile (Alvarez-Garreton et al., 2018) and Brazil (Chagas et al., 2020), with a recent extension of the framework to Colombia (Jimenez et al., 2025). These datasets represent major hydroclimatic domains, including extratropical Andean catchments, eastern Brazilian and Amazonian lowland basins, and tropical Andean and Caribbean-influenced catchments in northern South America. Nevertheless, the Pacific-draining catchments, central tropical Andes, and western Amazonian headwaters remain comparatively less represented in large-sample hydrology, despite their high hydrological relevance, complex topography, and pronounced hydroclimatic variability (Espinoza et al., 2020). Expanding CAMELS-type datasets across these regions is therefore essential for supporting more comprehensive comparative hydrological studies across South America.

Peru provides a particularly relevant setting for addressing this gap because it encompasses one of the most hydroclimatically heterogeneous territories in South America. The country includes major Amazonian headwaters, steep elevational gradients, and climate-sensitive hydrological systems influenced by ENSO variability (Sanabria et al., 2018), tropical glacier dynamics (Caro et al., 2024) , and strong atmospheric circulation contrasts (Gutierrez-Villarreal et al., 2025). This diversity is largely controlled by the Andes, which define the country's physiographic and hydrological structure by separating the Pacific, Atlantic, and Titicaca hydrographic regions while strongly modulating regional hydroclimatology and atmospheric circulation (Arias et al., 2021). Moisture transport from the Atlantic drives intense precipitation over the eastern Andes and Amazonian headwaters, whereas the Pacific coast remains predominantly arid due to subsidence and coastal upwelling processes (Sulca et al., 2024). These gradients result in strong contrasts in precipitation, runoff generation, hydrological seasonality, and water availability across Peruvian catchments (Llauca et al., 2021b).

Despite this hydrological diversity, hydrometeorological information in Peru remains fragmented across institutions, data formats, temporal coverages, and processing methodologies, limiting interoperability and reproducibility. These limitations are particularly relevant in data-scarce mountain regions, where hydrological and socio-environmental information remains sparse and heterogeneous (Muñoz et al., 2024). In addition, previous studies in the Peruvian Tropical Andes have shown that complex topography, difficult access, sparse gauges, discontinuous records, and missing data limit the representation of precipitation and streamflow variability, increasing uncertainties in hydrological modelling and forecasting applications

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

(Llauca et al., 2021a, 2023a). Accordingly, Peru still lacks a comprehensive and standardised large-sample hydrological dataset that integrates hydrometeorological time series and catchment attributes within a unified framework.

To address these limitations, this paper presents CAMELS-PE, a large-sample hydrological dataset for Peru. CAMELS-PE provides daily hydrometeorological time series and attributes for 136 catchments, including streamflow observations and simulations, meteorological forcings, and descriptors of hydroclimatic, physiographic, environmental, and human-intervention conditions. The dataset follows internationally adopted CAMELS conventions to facilitate interoperability, reproducibility, and intercomparison across large-sample hydrological studies. The objective of this paper is to introduce CAMELS-PE, document its development and organisation, and discuss its quality, uncertainties, limitations, and reuse potential in large-sample hydrological applications.

2 Study catchments and hydrometric network

The study area encompasses 136 catchments distributed across the three main hydrographic regions of Peru: 67 in the Pacific region, 56 in the Atlantic region, and 13 in the Titicaca region (Figure 1a). In this context, the Atlantic region corresponds primarily to the western headwaters of the Amazon basin, draining eastwards towards the Atlantic Ocean through major Amazonian tributaries. The study catchments span diverse hydroclimatic and physiographic conditions, from hyper-arid coastal systems on the Pacific slope to humid Amazonian headwaters in the eastern Andes and high-elevation endorheic basins surrounding Lake Titicaca. This spatial configuration includes nested catchments across all hydrographic regions, particularly within large Atlantic tributary systems, as well as several northern transboundary basins shared with Ecuador.

The hydrometric network was compiled primarily from streamflow gauging stations operated by the National Service of Meteorology and Hydrology of Peru (SENAMHI), complemented by additional gauges from external institutions with publicly accessible records. From this network, catchments were included in CAMELS-PE when daily streamflow observations were available during 1981–2025 and when their contributing drainage areas could be delineated from reliable outlet locations, without imposing a fixed minimum record-length or completeness threshold. Instead, record-length and completeness diagnostics are provided to support user-defined catchment selection. Although additional gauges exist in Peru, some lack public data, sufficient metadata, or mapped outlet information suitable for inclusion in this first dataset version.

Catchment boundaries were delineated using the DEM-based workflow described in Sect. 3.3.1. For transboundary systems, time series and attributes were calculated over the full upstream drainage area, using national products where available and global products for attributes such as topography, geology, soils, and land cover. Likewise, for nested catchments, all variables represent the complete upstream area draining to each gauge rather than incremental areas between nested outlets, implying spatial dependence among nested basins. This structure is documented through the nested-catchment attributes provided with the dataset.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

For the selected gauges, daily streamflow observations span 1981–2025, although temporal coverage varies considerably among records. This variability reflects differences in gauge operation, record continuity, and data availability across the hydrometric network. Long-term observations exceeding 30 years occur across several hydrographic regions, whereas shorter records (<10 years) are more common in northern and central Andean catchments (Figure 1a). Over time, the number of active gauges increased from the early 1980s onwards, with a stronger expansion during the mid-2010s and nearly 120 active gauges in the early 2020s (Figure 1b). To characterise this heterogeneity, coverage and completeness were assessed using combinations of minimum record length and data availability (Figure 1c). More than 100 gauges satisfy at least 70 % data availability and 5 years of observations, whereas fewer than 25 remain when requiring 30 years and at least 95 % completeness, highlighting the trade-off between temporal continuity and spatial coverage.

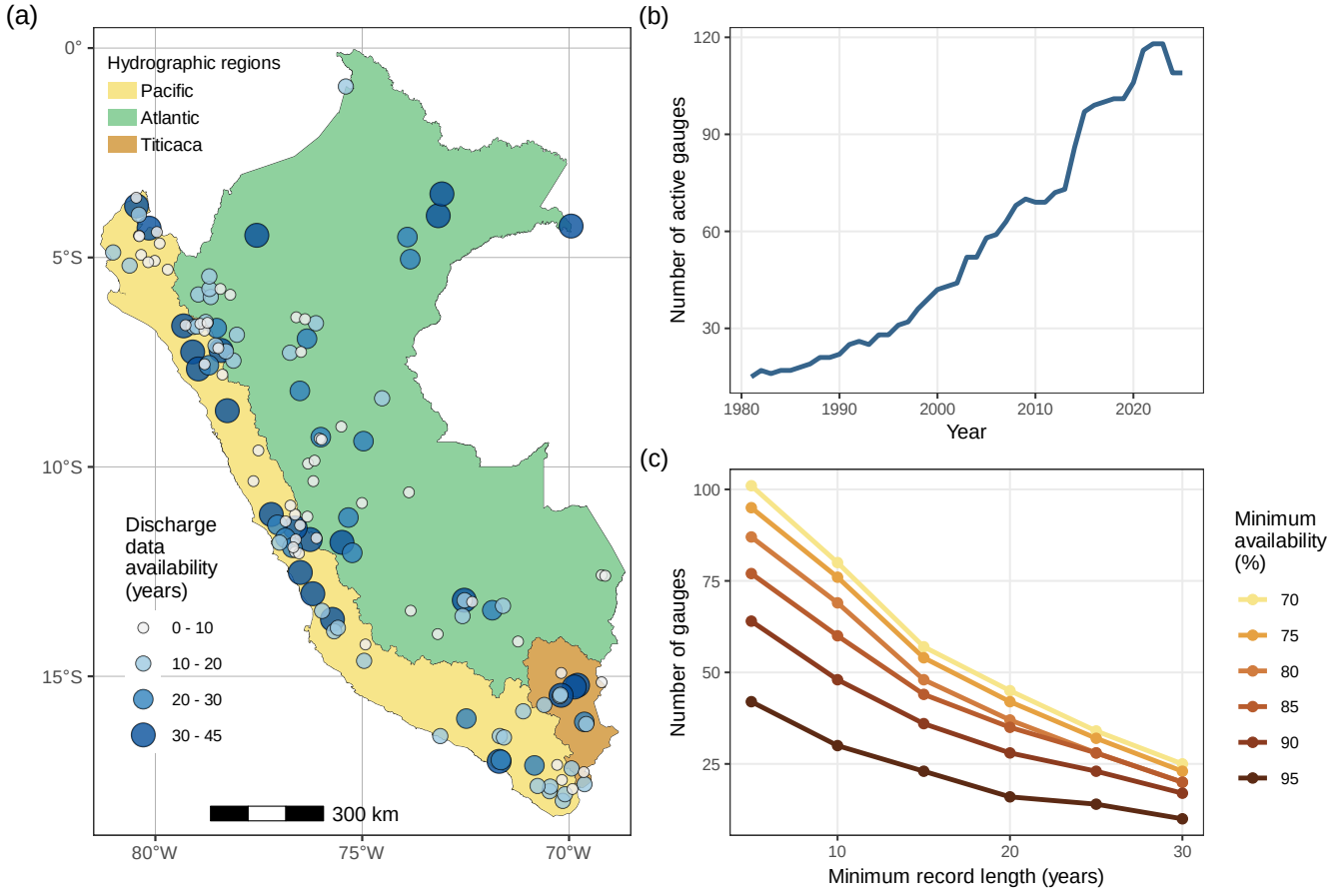


Figure 1: CAMELS-PE hydrometric network and streamflow data availability. (a) Gauging station locations and record length. (b) Annual number of active gauges during 1981–2025. (c) Number of gauges satisfying different record-length and data-availability thresholds.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

3 The CAMELS-PE dataset

3.1 Dataset overview

This first version of CAMELS-PE (v1.0.1) provides hydrometeorological time series, catchment attributes, metadata, dictionaries, and geospatial layers for 136 catchments under a consistent and standardised framework. The dataset includes 10 daily hydrometeorological variables formatted on a common 1981–2025 calendar and 79 catchment attributes describing the hydroclimatic, physiographic, environmental, and human-intervention characteristics of the study catchments (Table 1). Observed and simulated streamflow series are both included, with the latter also used to derive hydrological signatures consistently across all catchments over a common analysis period. Values outside the native temporal coverage of each source product are retained as missing values rather than being temporally extended or infilled. For distribution, CAMELS-PE is organised into four folders containing gauge metadata and dictionaries, catchment attributes, daily time series, and geospatial layers. Tabular data are distributed as comma-separated value files, whereas geospatial layers containing catchment boundaries and outlet locations are provided in GeoPackage format. All components are provided at the catchment scale and linked through a unique gauge identifier, ensuring consistency among time series, attributes, metadata, and spatial files.

Table 1. Summary of data provided by CAMELS-PE.

	Variable	Description and data source
Hydrometeorological daily series	Observed streamflow (mm d^{-1})	Records from SENAMHI; variable gauge coverage (1981–2025)
	Simulated streamflow (mm d^{-1})	Simulations from the PISCO-ARNOVIC v1.1 product (1981–2025) for the same gauges
	Precipitation (mm d^{-1})	Catchment average (mm d^{-1}) using PISCOp v2.1 (1981–2025)
	Precipitation variance (mm^2 d^{-2})	Spatial internal variance in precipitation from PISCOp v2.1
	Potential evapotranspiration (mm d^{-1})	Catchment average using PISCOeo_pm v1.0 (1981–2016)
	Minimum, maximum, and mean air temperature (°C)	Catchment average using PISCOt v1.2 (1981–2020)
	Solar radiation (MJ m^{-2} d^{-1})	Catchment average using ERA5-Land (1981–2025)
	Vapour pressure (hPa)	Catchment average using ERA5-Land (1981–2025)
Catchment attributes	Location and topography	22 attributes (Table 2)
	Climatic indices	10 attributes (Table 3)
	Hydrological signatures	13 attributes (Table 4)
	Land-cover characteristics	8 attributes (Table 5)
	Geological characteristics	7 attributes (Table 6)
	Soil characteristics	10 attributes (Table 7)
	Human intervention indicators	9 attributes (Table 8)

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

3.2 Hydrometeorological time series

3.2.1 Streamflow data

Daily observed streamflow records (Table 1) were obtained from the hydrometric network described in Sect. 2. The original observations correspond to daily mean streamflow values in $m^3\ s^{-1}$ and cover different periods depending on gauge operation and data availability, with records extending from 1981 to 2025. Before quality control, discharge records were converted to streamflow expressed as catchment-equivalent runoff depth in $mm\ d^{-1}$ using catchment drainage area.

The time series were subjected to a quality-control procedure adapted from the QARTOD framework (Bushnell et al., 2018), using three fail-only tests applied independently to each gauge: a gross-range test for negative values; a spike test for isolated daily values whose absolute difference from the mean of the previous and following valid observations exceeded four standard deviations of the corresponding gauge series; and a flatline test for sequences of at least seven consecutive valid observations with day-to-day differences smaller than 0.001 $mm\ d^{-1}$. Missing observations were flagged separately and retained as gaps, while the first 30 valid observations of each gauge were excluded from the spike and flatline tests as an initial warm-up period.

Records flagged as failed were removed from the observed streamflow series expressed as catchment-equivalent runoff depth rather than corrected or infilled. This conservative fail-only approach was designed to remove clearly erroneous values while preserving plausible hydrological extremes. Because the flatline test may flag very stable low-flow periods in small, regulated, or groundwater-influenced catchments, flagged records were visually inspected before removal to reduce the risk of excluding plausible hydrological behaviour. Consequently, the quality-controlled streamflow series should be interpreted as screened observational records rather than fully gap-filled or homogenised records. Additional details on the flagging scheme, diagnostics, and effect of the screening procedure are provided in Appendix A.

The dataset also includes simulated daily streamflow series (Table 1), which complement the observed records by providing spatially continuous and temporally complete runoff estimates across all catchments. Simulated streamflow was obtained from the PISCO-ARNOVIC v1.1 hydrological dataset, developed as a national hydrological modelling framework for Peru (Llauca et al., 2023b). The modelling system combines the ARNO/VIC rainfall–runoff model (Todini, 1996) with the RAPID river-routing model (David et al., 2011) and was implemented over 11,913 subbasins across Peru and transboundary basins.

The PISCO-ARNOVIC product was generated using its own forcing configuration based on gridded hydroclimatic inputs from the PISCO modelling framework; therefore, its temporal coverage should be distinguished from the native coverage of the individual meteorological variables distributed in CAMELS-PE and described in Sect. 3.2.2. Simulations in ungauged basins were generated through a similarity-based regionalisation approach previously evaluated using cross-validation procedures (Llauca et al., 2023b), and summary model performance metrics are provided in Appendix B. The original

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

PISCO-ARNOVIC v1.1 outputs consist of routed streamflow series for individual river reaches, from which the series corresponding to each CAMELS-PE catchment outlet were extracted.

For consistency within the dataset, simulated streamflow series were converted from $m^3 s^{-1}$ into catchment-equivalent runoff depths (mm d^{-1}), following the same unit convention applied to the observed streamflow records. Daily observed and simulated streamflow series are provided following standardised formatting and naming conventions to facilitate comparative hydrological analyses and modelling applications across the study domain.

3.2.2 Meteorological data

Daily meteorological forcing data included precipitation, air temperature, and potential evapotranspiration (Table 1), derived from gridded products of the Peruvian Interpolated Data of SENAMHI's Climatological and Hydrological Observations (PISCO). These products were developed by SENAMHI to provide consistent hydroclimatic information across Peru. Following the national meteorological convention, daily values were defined using the local Peruvian meteorological day, from 07:00 local time to 07:00 of the following day.

Precipitation was obtained from PISCOp v2.1 (Aybar et al., 2020), which integrates quality-controlled rain-gauge observations, CHIRP satellite precipitation estimates, and climatologies derived from the TRMM 2A25 radar product within a geostatistical interpolation framework based on residual inverse-distance weighting and ordinary kriging. PISCOp v2.1 provides daily precipitation estimates at 0.1° (~10 km) resolution from 1981 onwards; in CAMELS-PE, precipitation series were extracted for the 1981–2025 period.

Air temperature data were obtained from PISCOt v1.2 (Huerta et al., 2023), which provides daily maximum and minimum air temperature at 0.01° (~1 km) resolution for the 1981–2020 period. The product was developed through a four-stage framework involving quality control, gap-filling, homogenisation of station observations, and interpolation using topographic and remote-sensing covariates, including elevation, latitude, longitude, and MODIS land surface temperature products.

Potential evapotranspiration (PET) was derived from PISCOeo_pm v1.0 (Huerta et al., 2022), which provides daily gridded reference evapotranspiration estimates for Peru at 0.01° (~1 km) resolution for the 1981–2016 period. The product is based on the FAO Penman–Monteith equation and was generated from gridded meteorological variables derived from ground observations and climatologically aided interpolation, including air temperature, sunshine duration, dew-point temperature, and wind speed.

In addition to the PISCO-based variables, CAMELS-PE includes daily solar radiation and vapour pressure (Table 1) derived from ERA5-Land for the 1981–2025 period (Muñoz-Sabater et al., 2021). ERA5-Land was selected instead of the standard ERA5 dataset because of its higher spatial resolution and enhanced representation of land-surface processes. Hourly ERA5-Land outputs were converted to daily values using the 07:00–07:00 local-time aggregation window. Solar radiation was

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

accumulated over each meteorological day and converted from J m^{-2} d^{-1} to MJ m^{-2} d^{-1}, whereas vapour pressure was estimated from hourly dew-point temperature using the Magnus equation (Bolton, 1980). These variables were included to support hydrological applications requiring additional atmospheric predictors, particularly data-driven hydrological modelling approaches.

All meteorological variables were spatially aggregated using area-weighted averages based on the fraction of each grid cell within catchment boundaries, allowing partially intersecting cells to contribute proportionally to each catchment mean. Catchment-average daily series were then generated, with mean air temperature calculated from daily maximum and minimum temperature. All series were formatted on a common 1981–2025 daily calendar, with missing values retained outside the native temporal coverage of each product. The dataset also includes daily spatial precipitation variance within each catchment, calculated from intersecting grid-cell precipitation values as an auxiliary indicator of intra-catchment precipitation heterogeneity (Table 1).

3.2.3 Regional hydroclimatic patterns

Hydrometeorological time series for the 1981–2016 period reveal pronounced hydroclimatic contrasts among the Pacific, Atlantic, and Titicaca hydrographic regions of Peru (Figure 2). This period was selected to ensure consistency with the temporal coverage of the PET dataset, and runoff patterns are described using simulated streamflow to provide comparable information across all catchments.

Catchments in the Atlantic region exhibit the highest mean annual precipitation and runoff values, with median precipitation exceeding 1200 mm yr^{-1} and several catchments surpassing 4000 mm yr^{-1}, while runoff commonly exceeds 500 mm yr^{-1} and locally approaches 3000 mm yr^{-1} (Figure 2a,c). These conditions reflect the humid climate of the Amazonian headwaters and the strong influence of orographic precipitation along the eastern Andes. In contrast, Pacific catchments are characterised by substantially lower precipitation and runoff, with median precipitation generally below 700 mm yr^{-1} and runoff typically below 300 mm yr^{-1}, consistent with the arid to semi-arid conditions of the Pacific slope. The Titicaca region presents intermediate precipitation values around 700–800 mm yr^{-1} but comparatively low runoff, generally below 250 mm yr^{-1}, associated with high-altitude evaporative demand and endorheic hydrological conditions. Potential evapotranspiration exhibits lower variability than precipitation and runoff, with median values ranging approximately between 1200 and 1350 mm yr^{-1} across regions and generally higher values in the Pacific region (Figure 2b).

The relationship between the humidity index (P/PET) and runoff ratio (Q/P) further highlights these regional contrasts, with Amazonian catchments mostly distributed under humid conditions (P/PET > 1) and exhibiting runoff ratios frequently above 0.5, whereas Pacific catchments are mainly water-limited, with lower humidity indices and runoff ratios (Figure 2d). The dispersion observed within each hydrographic region suggests that runoff generation is also influenced by local physiographic and hydrological controls, illustrating the broad hydroclimatic diversity represented within CAMELS-PE.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

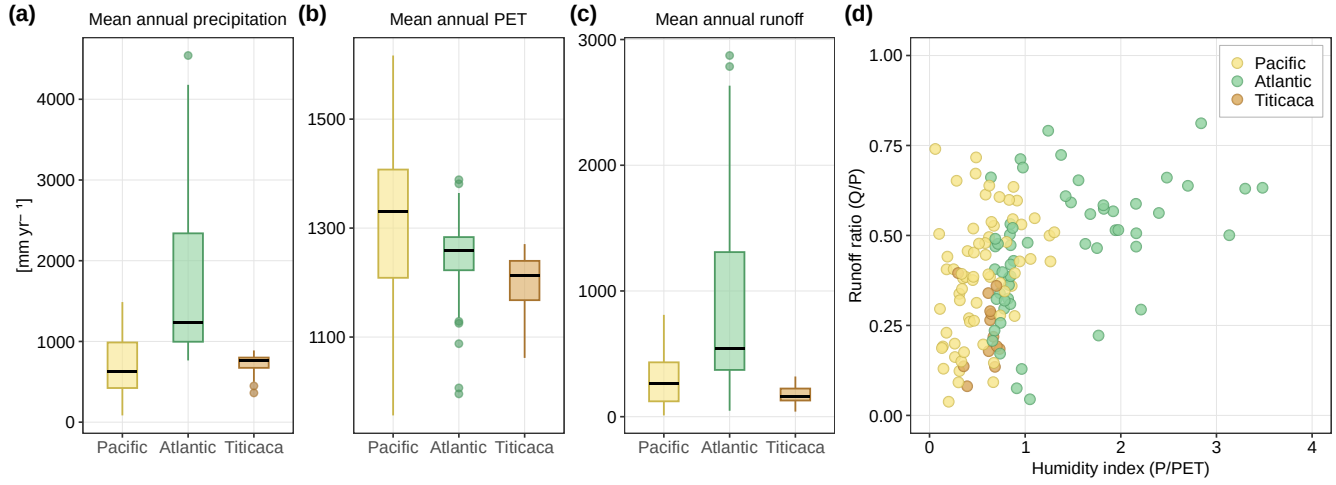


Figure 2: Regional hydroclimatic variability across CAMELS-PE catchments during 1981–2016. Panels show mean annual precipitation (a), potential evapotranspiration (PET) (b), simulated runoff (c), and the relationship between humidity index (P/PET) and simulated runoff ratio (Q/P) (d).

3.3 Catchment attributes

Catchment attributes were derived to characterise the hydroclimatic, physiographic, environmental, and human-intervention conditions of the CAMELS-PE catchments. All attributes were calculated at the catchment scale and linked to the time series, metadata, and geospatial layers through the unique gauge identifier. Depending on data availability and thematic scope, attributes were derived from national datasets, global spatial products, or model-consistent hydrological simulations, as described in the following subsections.

3.3.1 Location and topographic characteristics

Location and topographic attributes characterise the spatial organisation, terrain structure, and hydrometric configuration of the CAMELS-PE catchments, providing key information for analysing drainage structure, topographic controls, and potential hydrological response (Table 2). These attributes include gauge metadata, descriptors of catchment topology and river-network connectivity, and morphometric properties derived from the delineated catchments. Gauge metadata, including identifiers, geographic coordinates, elevation, and streamflow record information, were obtained from the official hydrological gauge inventory provided by SENAMHI, whereas official catchment names were obtained from the National Water Authority of Peru (ANA).

Catchment boundaries were delineated from gauge outlet locations using the Forest And Buildings removed Copernicus DEM (FABDEM v1.2; (Hawker et al., 2022) at 1 arc-second spatial resolution, approximately 30 m at the equator. FABDEM was selected to improve terrain representation for hydrological applications by reducing vegetation and building biases in global elevation datasets. Delineation followed a DEM-based workflow involving flow-direction, flow-

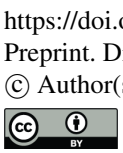
https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

accumulation, and watershed delineation procedures. The resulting boundaries were visually inspected and compared with official ANA hydrographic units where available, particularly for nested, transboundary, coastal, and low-relief catchments, to identify potential inconsistencies in outlet placement or drainage paths. The final catchment polygons were then used to
240 derive river-network topology, identify nested upstream–downstream relationships, and calculate morphometric attributes, including area, perimeter, mean slope, and minimum, mean, median, and maximum elevation.

Table 2. Location and topographic characteristics provided in CAMELS-PE.

Attribute name	Description	Units	Data source
gauge_id	Catchment identifier provided by SENAMHI	–	SENAMHI
gauge_name	Gauge name provided by SENAMHI	–	
gauge_region	Hydrographic region	–	
gauge_lat	Gauge latitude (WGS84)	° N	
gauge_lon	Gauge longitude (WGS84)	° E	
gauge_elev	Gauge elevation	m a.s.l.	
gauge_record_start	Start date of streamflow records	–	
gauge_record_end	End date of streamflow records	–	
gauge_perc_obs	Percentage of valid streamflow records (1981–2025)	%	
name_cat	Official catchment name	–	ANA
is_nested	Indicates whether the gauge belongs to a nested catchment system	TRUE / FALSE	Derived from spatial topology
nested_group_id	Gauge identifier of the terminal downstream gauge within the nested catchment	–	
nested_group_size	Number of gauges belonging to the same nested catchment group	–	
downstream_gauge_id	Gauge identifier of the immediately downstream nested catchment	–	
upstream_gauge_id	Gauge identifier of the immediately upstream nested catchment	–	
area	Catchment area	km²	FABDEM v1.2 (1 arc second)
perimeter	Catchment perimeter	km	
elev_min	Catchment minimum elevation	m a.s.l.	
elev_mean	Catchment mean elevation	m a.s.l.	
elev_median	Catchment median elevation	m a.s.l.	
elev_max	Catchment maximum elevation	m a.s.l.	
slope_mean	Catchment mean slope	m km^{-1}	

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.



The resulting attributes reveal marked physiographic variability across the CAMELS-PE catchments. Gauge and mean catchment elevations reflect the dominant topographic gradients imposed by the Andes, with the highest elevations concentrated in the central and southern Andes and progressively lower elevations towards the Amazonian lowlands and coastal regions (Figure 3a,c). Mean catchment elevations commonly exceed 4000 m a.s.l. in several southern Andean catchments, whereas many Amazonian and coastal catchments remain below 1000 m a.s.l. Catchment area also varies substantially, ranging from less than 100 km² in small headwater catchments to more than 10^5 km² in large Amazonian drainage systems (Figure 3b). Relatively small and steep catchments predominate along the western Andean flank, whereas larger drainage systems are mainly distributed across the eastern Andes and Atlantic region. This contrast reflects the asymmetric hydrographic organisation of the Peruvian Andes, where short Pacific-draining catchments contrast with extensive tributary systems draining towards the Amazon basin.

Mean catchment slope is generally highest along the western Andean flank and in mountain transition zones, where values frequently exceed 20 m km^{-1}, suggesting conditions favourable to faster runoff concentration (Figure 3d). In contrast, lower slopes dominate the Amazonian lowlands and the Altiplano surrounding Lake Titicaca, where flatter terrain may favour slower runoff concentration and greater hydrological storage potential.

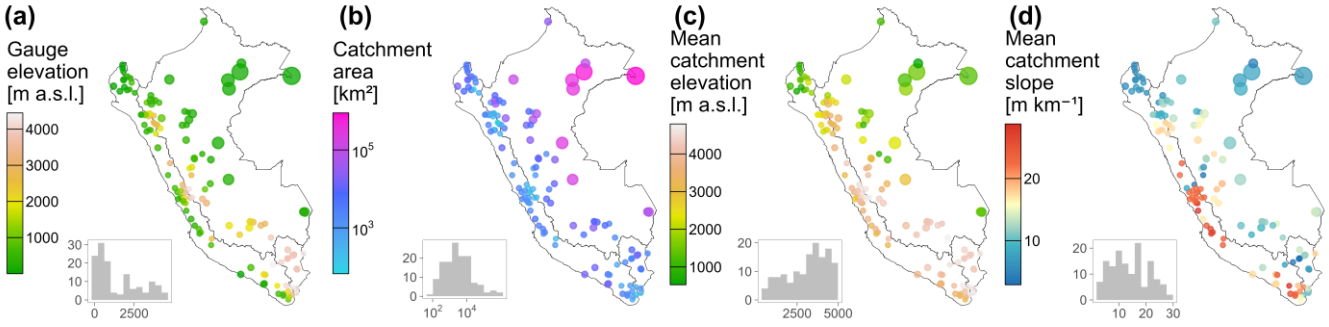


Figure 3: Topographic attributes of CAMELS-PE catchments. Points represent catchment outlets; point size represents drainage area. Colours indicate gauge elevation (a), drainage area (b), mean catchment elevation (c), and mean slope (d). Histograms show catchment-scale distributions.

3.3.2 Climatic indices

Climatic indices provide key information for understanding the long-term atmospheric controls on runoff generation, hydrological seasonality, and water availability across river basins. The selected indices, commonly reported in previous CAMELS datasets, characterise mean climatic conditions, aridity, precipitation seasonality, and the frequency, duration, and timing of wet and dry periods across the study catchments (Table 3). All indices were computed from catchment-scale daily time series for the 1981–2016 period, corresponding to the temporal availability of the PISCOeo_pm dataset. Aridity was estimated as the ratio between mean potential evapotranspiration and mean precipitation, whereas precipitation seasonality was calculated from sinusoidal representations of the annual precipitation and temperature cycles following Woods (2009). Additional threshold-based indices were derived from daily precipitation series to describe the occurrence and persistence of

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

wet and dry conditions. High-precipitation events were defined using thresholds based on multiples of mean daily precipitation, while dry days were identified as days with precipitation below 1 mm d^{-1}.

Table 3. Climatic indices provided in CAMELS-PE.

Attribute name	Description	Units	Data source
p_mean	Mean daily precipitation	mm d^{-1}	PISCOp v2.1
pet_mean	Mean daily potential evapotranspiration (PET)	mm d^{-1}	PISCOeo_pm v1.0
aridity	Aridity, computed as the ratio of mean PET to mean precipitation	–	PISCOeo_pm v1.0 and PISCOp v2.1
p_seasonality	Seasonality and timing of precipitation (estimated using sine curves to represent the annual temperature and precipitation cycles; values are typically between −1 (precipitation out of phase with temperature) and 1 (precipitation in phase with temperature, i.e., simultaneous peaks); values close to 0 indicate uniform precipitation throughout the year). See Eq. (14) in Woods (2009).	–	
high_prec_freq	Frequency of high precipitation days (≥ 5 times the mean daily precipitation)	days yr^{-1}	
high_prec_dur	Average duration of high precipitation events (number of consecutive days ≥ 5 times the mean daily precipitation)	days	PISCOp v2.1
high_prec_timing	Season during which most high precipitation days (≥ 5 times the mean daily precipitation) occur	season	
low_prec_freq	Frequency of dry days (< 1 mm d^{-1})	days yr^{-1}	
low_prec_dur	Average duration of dry periods (number of consecutive days < 1 mm d^{-1})	days	
low_prec_timing	Season during which most dry days (< 1 mm d^{-1}) occur	season	

Strong hydroclimatic gradients are evident across the CAMELS-PE catchments, consistent with Sect. 3.2.3. Mean daily precipitation is markedly higher in Atlantic region and Amazonian headwaters, where median values generally exceed 3 mm d^{-1} and several basins surpass 10 mm d^{-1}, whereas much lower precipitation dominates the Pacific and southern Titicaca regions (Figure 4a). These patterns reflect the influence of Atlantic moisture transport and orographic precipitation along the eastern Andes. Accordingly, the aridity index indicates predominantly humid conditions across the Atlantic region, while Pacific coastal catchments are characterised by higher aridity values associated with high evaporative demand and limited precipitation inputs (Figure 4b). Aridity values exceed 1 in many Pacific catchments and locally surpass 5 in some coastal basins, highlighting the arid to hyper-arid climate of the Peruvian Pacific slope. Precipitation seasonality is generally positive across most catchments, indicating that precipitation and temperature occur broadly in phase during the annual cycle, with wet conditions concentrated during the austral summer months (Figure 4c). In contrast, the wider distribution of seasonality values in Atlantic catchments reflects the more persistent and less seasonally constrained precipitation regime of humid tropical environments.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

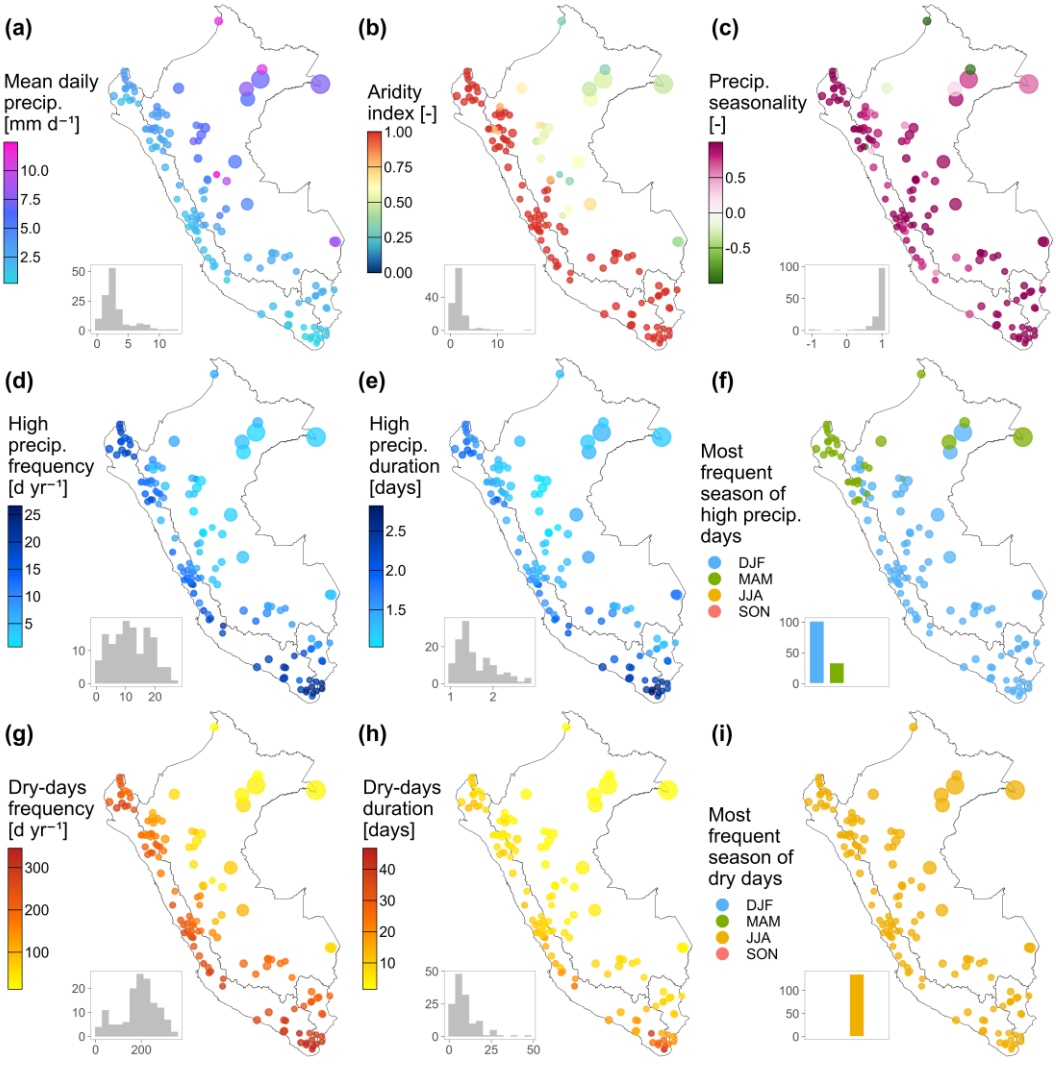


Figure 4: Climatic indices of CAMELS-PE catchments. Colours indicate precipitation, aridity, precipitation seasonality, high-precipitation characteristics, and dry-period characteristics. Histograms show catchment-scale distributions.

High-precipitation events are generally more frequent and persistent in northern Pacific and southern Andean catchments (Figure 4d,e), consistent with strong convective activity, moisture transport, orographic enhancement, and the recurrent influence of El Niño-related extreme rainfall along the northern Peruvian coast (Rollenbeck et al., 2022). Across most catchments, high-precipitation events occur predominantly during the austral summer season (DJF) (Figure 4f), consistent with the seasonal intensification of tropical moisture influx over the tropical Andes. In contrast, dry-day frequency and duration are considerably larger in the Pacific and Titicaca regions, where dry conditions commonly exceed 200 days yr^{-1} in several catchments, reflecting the prolonged dry seasons typical of the western Andean slope and Altiplano environments

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

(Figure 4g,h). The maximum occurrence of dry days is predominantly associated with the austral winter season (JJA) throughout the country (Figure 4i).

3.3.3 Hydrological signatures

Hydrological signatures provide integrated descriptors of runoff generation, flow regulation, hydrological variability, and the occurrence of flow extremes across river basins. To ensure consistency with the climatic indices presented in Sect. 3.3.2, all signatures were calculated for the 1981–2016 period. Although observed streamflow is provided as the primary streamflow record in CAMELS-PE, hydrological signatures were derived from simulated streamflow to ensure temporal completeness and spatial comparability across all catchments. These signatures should therefore be interpreted as model-consistent hydrological descriptors rather than observation-derived metrics.

Table 4. Hydrological signatures provided in CAMELS-PE.

Attribute name	Description	Units	Data source
q_mean	Mean daily streamflow	mm d^{-1}	
runoff_ratio	Runoff ratio, defined as the ratio between mean daily streamflow and mean daily precipitation	–	
stream_elas	Elasticity of streamflow with respect to precipitation (i.e., responsiveness of streamflow to precipitation changes at the annual scale, using mean daily streamflow as reference). See Eq. (7) in Sankarasubramanian et al. (2001), with $\bar{P}/\bar{Q}$ used instead of $\bar{Q}/\bar{P}$	–	
slope_fdc	Slope of the flow duration curve calculated between the log-transformed 33rd and 66th percentiles of streamflow	–	
baseflow_index	Baseflow index, estimated as the ratio of mean baseflow to mean total streamflow, where baseflow is derived using the digital filter of Ladson et al. (2013)	–	
hfd_mean	Mean half-flow date, defined as the day of the year when cumulative streamflow (starting on 1 September) reaches 50 % of the annual total	day of the year	PISCO-ARNOVIC v1.1
Q5	5th percentile of streamflow (low-flow)	mm d^{-1}	
Q95	95th percentile of streamflow (high-flow)	mm d^{-1}	
high_q_freq	Frequency of high-flow days (> 9 times the median daily streamflow)	days yr^{-1}	
high_q_dur	Average duration of high-flow events (number of consecutive days exceeding 9 times the median daily streamflow)	days	
low_q_freq	Frequency of low-flow days (< 0.2 times the mean daily streamflow)	days yr^{-1}	
low_q_dur	Average duration of low-flow events (number of consecutive days below 0.2 times the mean daily streamflow)	days	
zero_q_freq	Percentage of days with zero streamflow.	%	

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.


The selected signatures describe complementary aspects of catchment hydrological behaviour, including runoff production and efficiency, flow seasonality and persistence, responsiveness to climatic variability, and the magnitude and duration of hydrological extremes (Table 4). Streamflow elasticity was estimated following Sankarasubramanian et al. (2001), whereas flow regulation characteristics were represented using the slope of the flow duration curve and the baseflow index derived from the digital filtering approach proposed by Ladson et al. (2013). Additional indicators describing high-flow and low-flow conditions were computed using threshold-based analyses of the daily streamflow series, together with percentile-based metrics characterising extreme discharge conditions. Here, Q5 and Q95 refer to the empirical 5th and 95th percentiles of daily streamflow values, respectively, rather than flow-duration exceedance probabilities.

Hydrological signatures reveal substantial variability in runoff generation and flow-regime behaviour across the CAMELS-PE catchments. Mean daily streamflow and runoff ratio are generally higher in humid Amazonian and eastern Andean catchments, where runoff ratios frequently exceed 0.5 and mean daily discharge locally surpasses 6 mm d^{-1}, whereas substantially lower runoff production dominates the Pacific slope and Titicaca region (Figure 5a,b). The timing of runoff generation also exhibits a marked spatial organisation, with earlier runoff concentration observed in Pacific catchments and progressively later half-flow dates towards the Atlantic region, commonly exceeding day 190 of the hydrological year in humid basins (Figure 5c). These patterns are consistent with the combined influence of precipitation seasonality, catchment storage, and sustained runoff generation in wetter environments. Flow duration curve slope and streamflow elasticity are generally larger in highly seasonal and water-limited catchments, particularly along the Pacific region, where elasticity values locally exceed 4, suggesting stronger runoff sensitivity to precipitation variability and more variable streamflow regimes (Figure 5d,f). In contrast, higher baseflow index values, commonly above 0.7, are predominantly observed in humid Amazonian catchments, suggesting greater flow regulation and more persistent delayed-flow contributions under wetter climatic conditions (Figure 5e).

Hydrological extremes also exhibit strong spatial organisation across the study domain. High-flow frequency, duration, and Q95 values are generally larger in humid Andean and Amazonian catchments, where Q95 values locally exceed 30 mm d^{-1} and high-flow events may persist for more than 20 consecutive days (Figure 5g–i). These conditions are consistent with sustained precipitation inputs and prolonged runoff generation in humid environments. Conversely, low-flow frequency and duration are considerably higher in Pacific and southern catchments, where low-flow conditions commonly exceed 150 days yr^{-1} and event durations may surpass 60 consecutive days in highly seasonal basins (Figure 5j,k). Spatial patterns of Q5 further indicate reduced low-flow availability in arid and intermittent catchments, with several Pacific basins exhibiting values close to zero during dry periods (Figure 5l). White points in the duration maps indicate catchments where the selected high-flow or low-flow thresholds were not exceeded during the analysis period, resulting in zero event frequency and undefined event-duration metrics. Overall, these signatures highlight the strong hydroclimatic controls on runoff variability across Peru, ranging from persistent and hydrologically regulated humid systems to highly seasonal and intermittently flowing catchments influenced by climatic water limitations.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

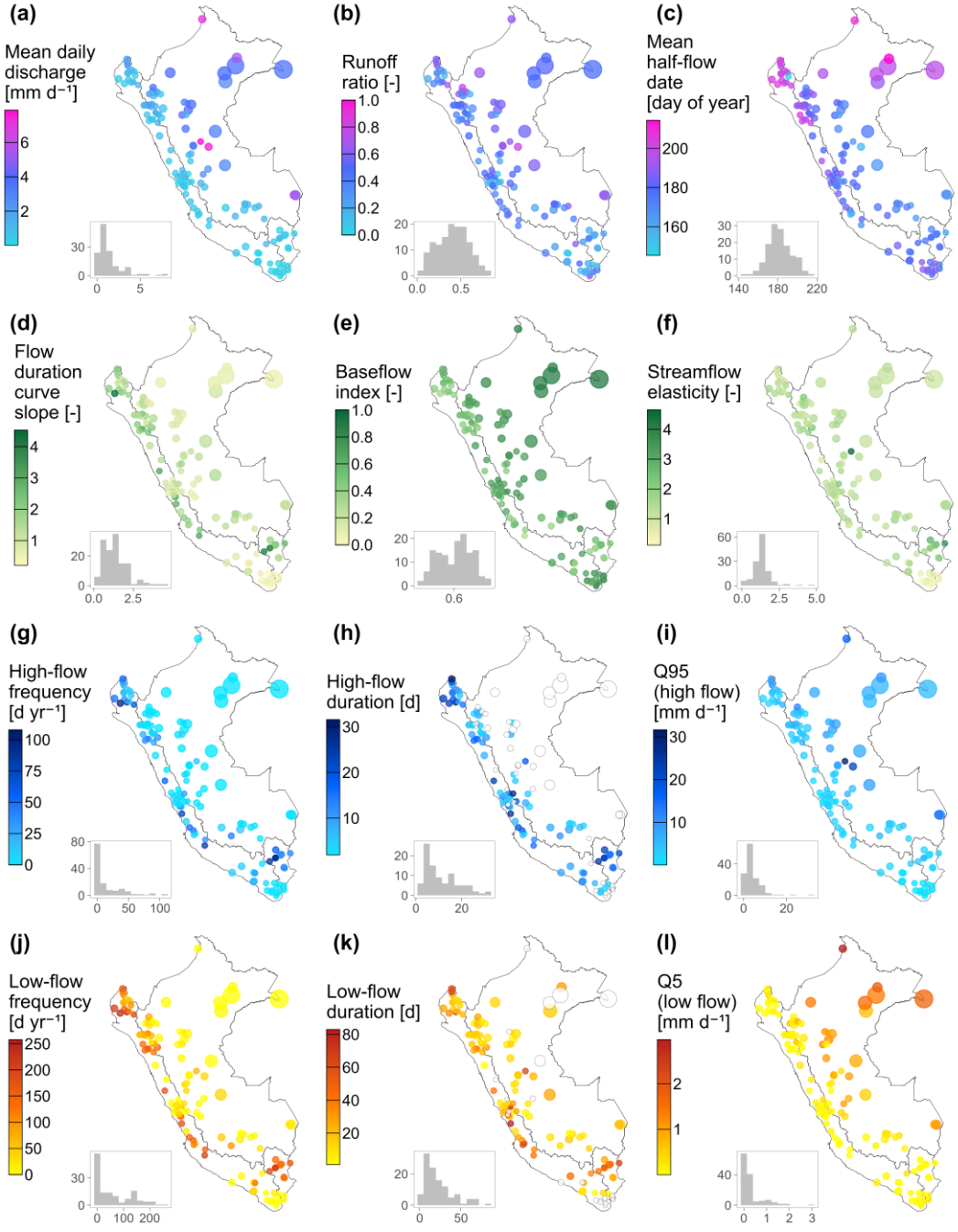


Figure 5: Hydrological signatures of CAMELS-PE catchments derived from simulated streamflow. Colours indicate runoff, flow seasonality, flow regulation, flow variability, and high- and low-flow characteristics. Histograms show catchment-scale distributions.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

3.3.4 Land-cover characteristics

Land-cover attributes were derived to characterise the dominant vegetation and land-use conditions influencing hydrological processes across the study catchments. Variations in vegetation cover and land use influence evapotranspiration, infiltration, soil moisture dynamics, runoff generation, and flow regulation within river basins. Land-cover information was obtained from MapBiomas Peru Collection 3.0 (Proyecto MapBiomas Perú, 2025) and, for transboundary drainage areas extending into Ecuador, from MapBiomas Ecuador Collection 3.0 (Proyecto MapBiomas Ecuador, 2025). These products provide annual land-use and land-cover classifications at 30 m spatial resolution for 1985–2024, based on Landsat imagery, Random Forest classification algorithms, and temporal consistency filters implemented in Google Earth Engine. In this study, the 2024 land-cover classification was used.

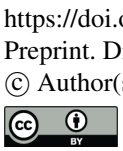
The original MapBiomas legend, which includes up to 26 classes, was reclassified into seven broader categories: croplands, forest cover, grasslands and shrublands, non-forest natural formations, non-vegetated areas, water bodies, and non-identified land-cover (Table 5). Catchment-scale attributes were calculated as the percentage of total catchment area occupied by each category using area-weighted spatial aggregation, and the dominant land-cover category was also identified for each catchment. Only the most representative categories are presented in Figure 6, whereas additional classes with limited spatial coverage are included as attributes in the distributed dataset and listed in Table 5.

Table 5. Land-cover characteristics provided in CAMELS-PE.

Attribute name	Description	Units	Data source
agricul_perc	Percentage of the catchment covered by croplands	%	MapBiomas Peru and MapBiomas Ecuador
forest_perc	Percentage of the catchment covered by forest cover	%	
non_veget_perc	Percentage of the catchment covered by non-vegetated area	%	
non_woody_perc	Percentage of the catchment covered by grasslands and shrublands	%	
water_perc	Percentage of the catchment covered by water bodies	%	
non_forest_perc	Percentage of the catchment covered by non-forest natural formation	%	
non_identi_perc	Percentage of the catchment covered by non-identified land-cover	%	
land_dominant_class	Dominant land-cover class	–	

The categories shown in Figure 6 reveal strong spatial organisation associated with the climatic and physiographic gradients of Peru. Forest cover is predominantly concentrated in humid Amazonian and eastern Andean catchments, where forest percentages frequently exceed 75 % and locally approach complete basin coverage in lowland Amazonian systems (Figure 6b). These densely vegetated catchments may be associated with higher evapotranspiration, greater infiltration opportunities, and more sustained delayed-flow contributions. In contrast, cropland coverage is generally higher along the Pacific slope and inter-Andean valleys, with several catchments exceeding 40 % agricultural coverage, reflecting the concentration of irrigated

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.



agriculture and human activities in these regions (Figure 6a). Catchments with higher agricultural coverage may therefore be more likely to exhibit human influences on runoff seasonality, water abstraction, and flow regulation.

Natural grassland and shrubland formations dominate many high-elevation Andean and southern catchments, particularly within the Titicaca region and central Andes, where puna ecosystems are widespread (Figure 6c). These high-altitude ecosystems may influence soil water storage, infiltration, and seasonal runoff regulation, particularly in headwater environments. Non-vegetated areas are mainly associated with arid coastal catchments and high-mountain environments characterised by sparse vegetation cover, with several Pacific catchments exceeding 40 % non-vegetated surfaces (Figure 6d). Such conditions are commonly associated with reduced vegetation-mediated infiltration and more rapid runoff response during intense precipitation events.

Water bodies occupy only a small proportion of most catchments, although slightly larger percentages occur in southern Andean basins influenced by lakes, wetlands, reservoirs, and glaciers connected to high-elevation or Titicaca systems (Figure 6e). Dominant land-cover classes further emphasise these regional patterns, with forests and grassland/shrubland formations prevailing across most catchments, whereas croplands dominate a smaller number of basins mainly distributed along the Pacific region (Figure 6f).

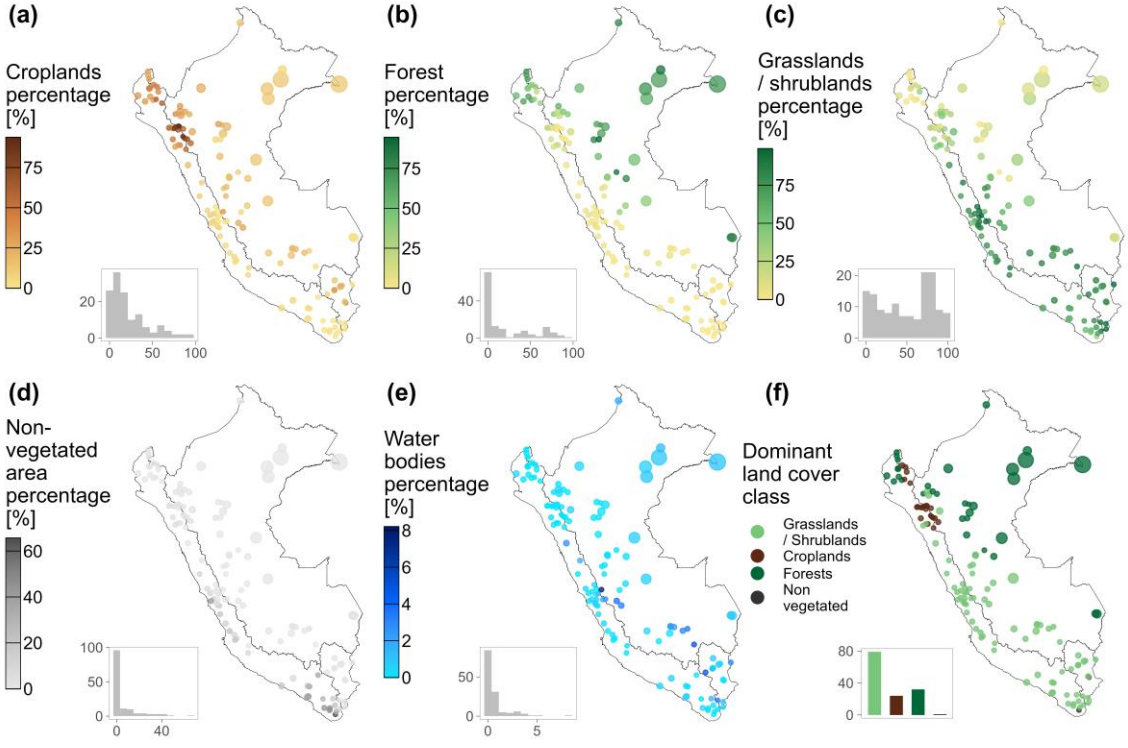


Figure 6: Land-cover attributes of CAMELS-PE catchments. Colours indicate cropland, forest, grassland/shrubland, non-vegetated area, water bodies, and dominant land-cover class. Histograms show catchment-scale distributions.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

3.3.5 Geological characteristics

Geological attributes were derived to describe catchment-scale lithological composition and subsurface hydrogeological properties (Table 6). Lithological information was obtained from the Global Lithological Map (GLiM v1.0; Hartmann and Moosdorf 2012), a global vector database compiled from 92 geological maps and organised into hierarchical lithological classes. GLiM polygons were spatially intersected with the delineated catchment boundaries to calculate the areal proportion of each lithological class. From these intersections, five lithological attributes were derived for each catchment: the dominant geological class and its percentage, the second most frequent geological class and its percentage, and the percentage of intermediate volcanic rocks. The latter was included as a separate attribute because intermediate volcanic formations, mainly andesitic rocks and related units, are widespread across the Peruvian Andes and may provide a useful first-order descriptor of lithological variability in Andean headwater catchments.

Subsurface hydrogeological properties were obtained from the GLobal HYdrogeology MaPS dataset (GLHYMPS v2.0; Huscroft et al. 2018), which provides global estimates of near-surface permeability and porosity based on lithological and hydrogeological information. GLHYMPS 2.0 incorporates the Global Unconsolidated Sediments Map to improve the representation of unconsolidated deposits and provides permeability estimates calibrated using regional groundwater models. In CAMELS-PE, porosity and permeability were spatially aggregated to the catchment scale using area-weighted averaging, with permeability retained in logarithmic form following the original GLHYMPS specification. These variables complement the GLiM-based lithological classes by providing continuous descriptors of subsurface conditions that may influence infiltration, subsurface storage potential, hydrological connectivity, and runoff generation.

Table 6. Geological characteristics provided in CAMELS-PE.

Attribute name	Description	Units	Data source
geol_class_1st	Most common geological class in the catchment	–	GLiM v1.0
geol_class_1st_perc	Percentage of the catchment covered by the most common geological class	%	
geol_class_2nd	Second most common geological class in the catchment	–	
geol_class_2nd_perc	Percentage of the catchment covered by the second most common geological class	%	
inter_volca_rocks_perc	Percentage of the catchment covered by intermediate volcanic rocks	%	
geol_porosity	Subsurface porosity of the catchment	–	GLHYMPS v2.0
geol_permeability	Average catchment area-weighted subsurface permeability (log10 scale)	log10(m^2)	

The GLiM-derived attributes reveal a clear spatial organisation that broadly follows the main physiographic domains of Peru. Mixed sedimentary rocks are frequent across several Atlantic and northern catchments, particularly in larger drainage areas extending towards the Amazonian lowlands, whereas intermediate volcanic rocks are more common along the central

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

and southern Andes. Unconsolidated sediments are mostly associated with lowland, coastal, and alluvial environments, while carbonate sedimentary rocks appear as dominant or secondary classes in some Andean catchments, especially in the central part of the country. The percentage of the dominant class generally ranges from about 10 % to 50 % of catchment area, with most catchments concentrated between approximately 20 % and 35 % (Figure 7c), indicating that many catchments integrate multiple lithological units rather than being represented by a single geological class.

The second dominant geological class commonly covers about 10 %–40 % of catchment area and rarely approaches 50 % (Figure 7d), further emphasising the lithological heterogeneity of Peruvian catchments, particularly along the Andean corridor. Intermediate volcanic rocks show a marked spatial concentration along the central and southern Andes, with several catchments exceeding 25 % and some locally exceeding 50 % of catchment area (Figure 7e). From a hydrological perspective, this lithological diversity may influence infiltration, subsurface storage, hydrological connectivity, and runoff pathways. Unconsolidated sediments may be associated with greater shallow storage where alluvial deposits are extensive, whereas volcanic, plutonic, and metamorphic terrains may indicate more fracture-controlled subsurface pathways. Although GLHYMPS-derived porosity and permeability are not shown in Figure 7, they are retained in the dataset as complementary continuous descriptors of subsurface hydrogeological conditions.

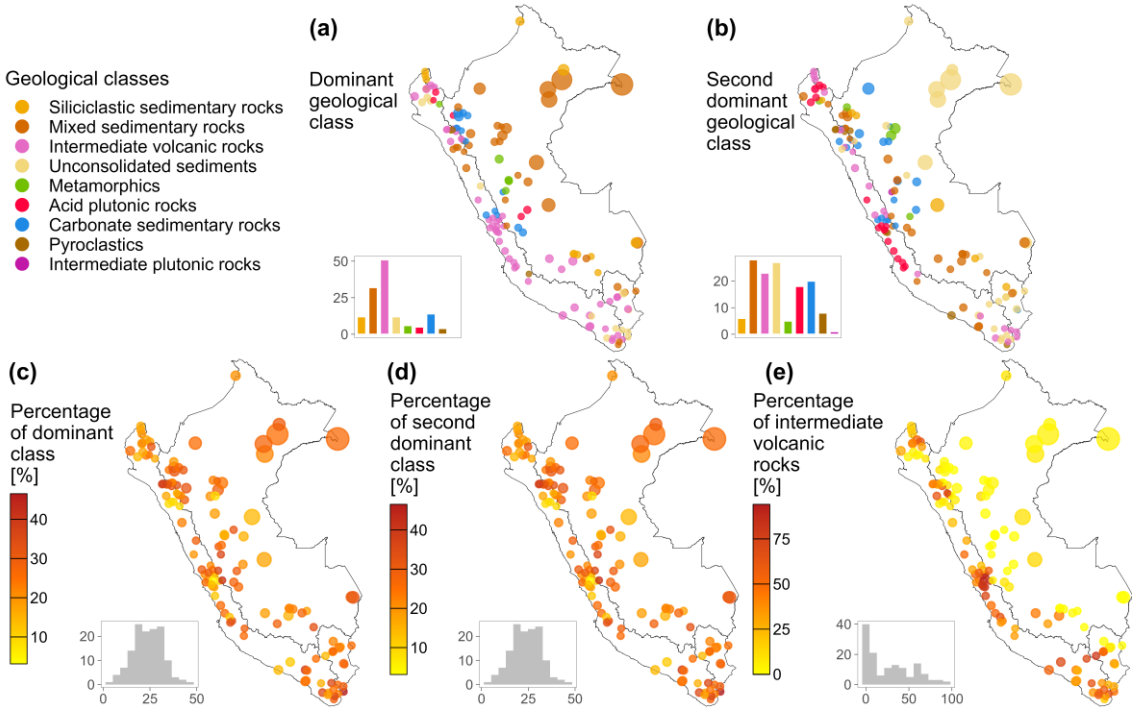


Figure 7: Geological attributes of CAMELS-PE catchments. Colours indicate dominant and secondary geological classes and their relative coverage. Histograms show catchment-scale distributions.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

3.3.6 Soil characteristics

Soil attributes were derived to describe the dominant soil orders and their spatial distribution within each CAMELS-PE catchment (Table 7). Soil information was obtained from the Digital Soil Open Land Map (DSOLMap; López-Ballesteros et al. 2023), a global digital soil product at 250 m spatial resolution developed to support hydrological modelling applications. DSOLMap integrates Open Land Map data and provides soil information and physical–hydraulic property estimates across standard depths using three-dimensional machine-learning algorithms calibrated with global soil-profile observations. In CAMELS-PE, DSOLMap soil-class information was used to identify the dominant USDA soil order and to calculate the percentage of catchment area covered by each selected order. The considered classes include Inceptisols, Entisols, Alfisols, Ultisols, Aridisols, Gelisols, Oxisols, Mollisols, and Vertisols. Soil classes were spatially intersected with the delineated catchment boundaries, and the proportional area occupied by each soil order was calculated relative to the total drainage area. These attributes provide first-order descriptors of soil development, weathering status, drainage conditions, and potential water-retention behaviour relevant to infiltration, soil-water storage, evapotranspiration, and runoff generation. Only the most representative soil orders are presented in Figure 8, whereas additional soil-order classes with limited spatial coverage are included as attributes in the distributed dataset and listed in Table 7.

Table 7. Soil characteristics provided in CAMELS-PE.

Attribute name	Description	Units	Data source
inceptisols_perc	Percentage of catchment covered by inceptisols	%	Digital Soil Open Land Map (DSOLMap)
entisols_perc	Percentage of catchment covered by entisols	%	
alfisols_perc	Percentage of catchment covered by alfisols	%	
ultisols_perc	Percentage of catchment covered by ultisols	%	
aridisols_perc	Percentage of catchment covered by aridisols	%	
gelisols_perc	Percentage of catchment covered by gelisols	%	
oxisols_perc	Percentage of catchment covered by oxisols	%	
mollisols_perc	Percentage of catchment covered by mollisols	%	
vertisols_perc	Percentage of catchment covered by vertisols	%	
soil_dominant_class	Dominant soil order	–	

The soil orders shown in Figure 8 reveal a clear regional organisation associated with Peru's climatic and geomorphological gradients. Inceptisols are the most widespread soil order in the CAMELS-PE catchments and dominate many Andean and inter-Andean basins (Figure 8a,d). Their coverage ranges from values below 25 % in some coastal and Amazonian catchments to more than 75 % in several central and southern Andean basins, locally approaching complete catchment coverage. This pattern is consistent with relatively young to moderately developed soils in mountainous terrain, where steep

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

slopes and active erosion may limit advanced soil development. Entisols are more spatially restricted but show high coverage in several arid coastal and southern catchments, locally exceeding 75 % and approaching 100 % in some basins (Figure 8b). Their occurrence reflects poorly developed soils commonly associated with dry climates, sparse vegetation, alluvial deposits, or recently reworked surfaces.

Ultisols are concentrated mainly in humid Amazonian catchments and eastern Andean slopes, where their coverage commonly exceeds 50 % and locally surpasses 75 % of catchment area (Figure 8c,d). This distribution is consistent with highly weathered tropical soils developed under humid conditions and sustained leaching. The dominance map indicates that Inceptisols represent the largest number of catchments, followed by Ultisols, whereas Entisols dominate a smaller group mainly located in arid or semi-arid environments. From a hydrological perspective, these soil-order patterns provide first-order information on potential differences in infiltration, soil-water storage, evapotranspiration, and runoff generation. Overall, the soil attributes capture a transition from poorly developed arid and high-relief soils to highly weathered humid tropical soils, complementing the climatic, topographic, geological, and land-cover gradients represented in CAMELS-PE.

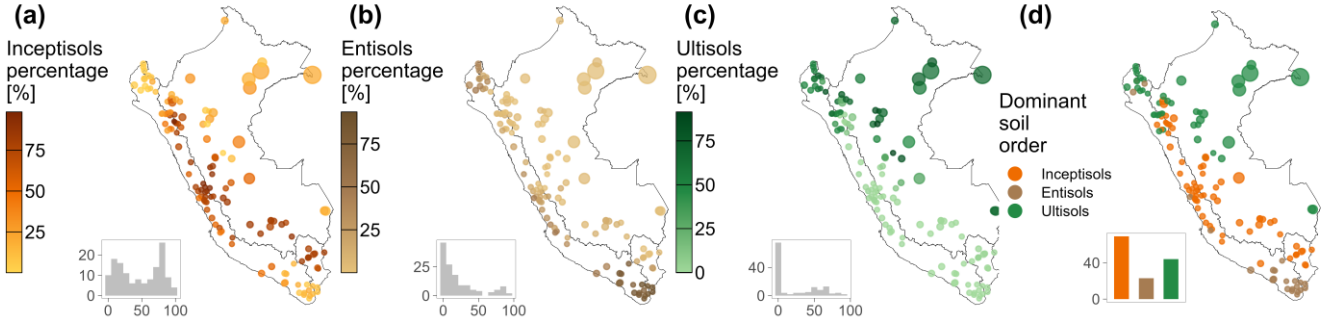


Figure 8: Soil attributes of CAMELS-PE catchments. Colours indicate soil-order percentages and dominant soil order. Histograms show catchment-scale distributions.

3.3.7 Human-intervention indicators

Human-intervention attributes were derived to represent recorded anthropogenic pressures on water resources within the CAMELS-PE catchments (Table 8). The information was obtained from georeferenced vector datasets provided by ANA Peru, including Water Availability Accreditation (ADH) records and reservoir inventories. In Peru, ADH is an administrative procedure conducted prior to the allocation of water-use rights to evaluate whether sufficient water resources are available for a proposed activity; therefore, these attributes should be interpreted as indicators of potential water-allocation pressure rather than direct measurements of actual water demand, abstraction, consumption, or hydrological alteration.

ADH records were spatially intersected with the delineated catchment boundaries, and duplicate records generated by the intersection procedure were removed using unique record identifiers. Records were separated into surface-water and groundwater accreditations according to the reported source type, with springs and aquifers classified as groundwater sources and the remaining records classified as surface-water sources. For each catchment, the number of ADH records, total

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

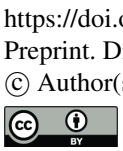
accredited volume, and dominant use category were calculated separately for surface water and groundwater; accredited volumes were converted to $hm^3\ yr^{-1}$, and missing or invalid volume values were excluded from the aggregation. Reservoir attributes were derived from the ANA reservoir vector layer using the same catchment-intersection approach, from which the number of reservoirs, total storage capacity, and dominant reservoir use category were calculated. Use categories were harmonised into common classes, and catchments without ADH records or reservoirs were assigned zero counts and volumes, with the corresponding use category reported as "None".

Table 8. Human-intervention indicators provided in CAMELS-PE.

Attribute name	Description	Units	Data source
surface_adh_n	Number of surface-water availability accreditation records within the catchment	–	ANA
surface_adh_vol	Total administratively accredited volume associated with surface-water availability accreditation records within the catchment	$hm^3\ yr^{-1}$	
surface_adh_use	Dominant use category of surface-water availability accreditation records within the catchment	–	
groundwater_adh_n	Number of groundwater availability accreditation records within the catchment	–	
groundwater_adh_vol	Total administratively accredited volume associated with groundwater availability accreditation records within the catchment	$hm^3\ yr^{-1}$	
groundwater_adh_use	Dominant use category of groundwater availability accreditation records within the catchment	–	
reservoir_n	Number of reservoirs located within the catchment	–	
reservoir_vol	Total reservoir storage capacity within the catchment	hm^3	
reservoir_use	Dominant reservoir use category within the catchment	–	

The resulting attributes reveal strong contrasts in water-allocation pressure and hydraulic infrastructure across the CAMELS-PE catchments. Surface-water ADH records range from zero in many catchments to more than 300–400 records in a few basins, while administratively accredited surface-water volumes are mostly below 100 $hm^3\ yr^{-1}$ but exceed 300 $hm^3\ yr^{-1}$ in some catchments (Figure 9a,b). Groundwater ADH records show an even wider range, from zero to more than 900 records in catchments with the largest counts, although administratively accredited groundwater volumes are generally lower, commonly below 10 $hm^3\ yr^{-1}$ and locally exceeding 20 $hm^3\ yr^{-1}$ (Figure 9d,e). These records are concentrated mainly along the Pacific slope and northern coastal catchments, where arid to semi-arid conditions coincide with intensive agricultural, urban, and mining activities. By contrast, many Atlantic and Titicaca catchments exhibit fewer ADH records and lower administratively accredited volumes, indicating lower administrative water-allocation pressure across much of these regions.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.



Dominant use categories indicate that agriculture is the main surface-water ADH use in most catchments, followed by domestic use, whereas aquaculture, energy, industrial, and other uses occur more locally (Figure 9c). Groundwater ADH use shows a stronger contribution from domestic use, although agriculture remains important in several catchments (Figure 9f).

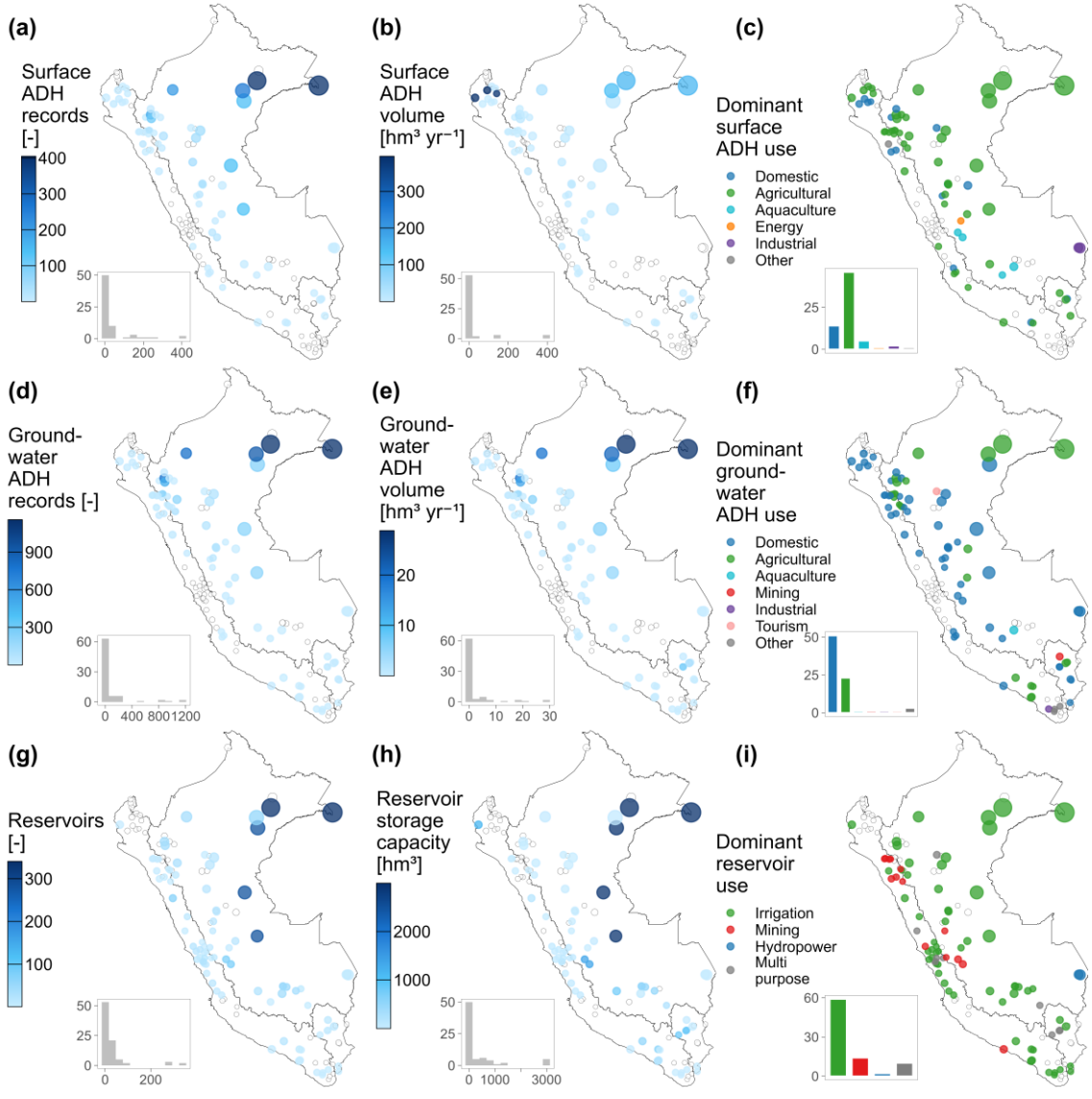


Figure 9: Human-intervention attributes of CAMELS-PE catchments. Colours indicate ADH records and administratively accredited ADH volumes, reservoir counts, storage capacity, and dominant use categories; point size represents drainage area. Histograms show catchment-scale distributions.

Reservoir-related indicators also show marked spatial heterogeneity: the number of reservoirs ranges from zero in many catchments to more than 300 in a few basins, while total storage capacity ranges from negligible values to more than 2000–3000 hm³ in large regulated systems (Figure 9g,h). Irrigation is the dominant reservoir use across most catchments with

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.


reservoirs, whereas mining, hydropower, and multi-purpose reservoirs are locally important in specific Andean and coastal basins (Figure 9i).

These attributes provide first-order indicators of potential flow alteration, storage regulation, and water-allocation pressure. However, because ADH records represent administrative accreditations rather than observed withdrawals, the absence of ADH records should not be interpreted as the absence of water demand or human intervention. In some highly intervened catchments, particularly along the central Pacific coast, the lack of ADH records may reflect limited new water-availability accreditations, constraints on additional water availability, administrative reporting differences, or data-coverage limitations. Therefore, spatial differences in these indicators may reflect differences in administrative reporting, formalisation of water uses, and data availability, in addition to actual water-resource pressure.

4 Data quality, uncertainty and limitations

Although CAMELS-PE provides broad national coverage, it should not be interpreted as a spatially uniform or exhaustive representation of all Peruvian river basins. Gauge availability remains uneven across the country and is constrained by the historical development, accessibility, and operation of the hydrometric network.

Uncertainty also arises from catchment delineation and spatial aggregation. Although catchment boundaries were visually checked against official ANA hydrographic units where available, residual uncertainty may remain in outlet placement and drainage-path representation, especially in low-gradient, transboundary, coastal, and large Atlantic-draining catchments. In addition, several CAMELS-PE catchments are nested and share upstream drainage areas; therefore, they should not be treated as statistically independent samples without considering their spatial dependence. Users applying statistical analyses, machine-learning models, regionalisation methods, or model benchmarking experiments should account for this structure, for example by using the provided nested-catchment attributes or by filtering the dataset to non-nested terminal basins when independent samples are required.

Observed streamflow records vary substantially in length, continuity, and completeness. The quality-control procedure described in Sect. 3.2.1 and Appendix A was implemented as a conservative fail-only screening approach to remove clearly erroneous values while preserving plausible hydrological extremes. This is particularly relevant because streamflow records from conventional and automatic stations may be affected by sensor failures, reporting artefacts, rating-curve issues, or extrapolation uncertainty during high flows. Missing observations were retained as gaps; therefore, the observed streamflow series should be interpreted as quality-controlled records with variable temporal completeness, not as homogeneous or gap-filled products.

Simulated streamflow was included to provide spatially complete and temporally homogeneous runoff estimates for all catchments. The PISCO-ARNOVIC simulations support comparative analyses and the derivation of model-consistent hydrological signatures over a common period, especially where observed records are short or discontinuous. Appendix B

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.


indicates generally reasonable agreement between simulated and observed streamflow at the national scale, although performance varies among catchments and hydrological regimes. Consequently, hydrological signatures derived from simulated streamflow should be interpreted as model-consistent descriptors rather than direct observation-based metrics, particularly for low-flow, high-flow, and baseflow-related indicators.

Meteorological forcings and climatic indices were derived mainly from PISCO gridded products and aggregated to the catchment scale. These products benefit from Peru's national meteorological station network, but uncertainties remain in regions with sparse observations, complex topography, strong elevation gradients, or limited ground validation, such as parts of the Atlantic region, high Andean basins, and remote Amazonian headwaters. ERA5-Land variables were included to support applications requiring additional atmospheric predictors, particularly data-driven modelling, but should also be interpreted with caution in complex terrain.

Catchment attributes derived from spatial products, especially geology, hydrogeology, soils, land cover, and elevation, should be interpreted as catchment-scale descriptors rather than substitutes for detailed local surveys. Aggregation to the catchment scale may reduce some fine-scale positional uncertainty but does not eliminate uncertainty in the underlying source maps. Similarly, climatic indices and hydrological signatures inherit uncertainties from the PISCO meteorological inputs and PISCO-ARNOVIC streamflow simulations, respectively.

Human-intervention attributes should also be interpreted carefully. Because spatially explicit information on actual withdrawals, consumptive use, and demand points is not consistently available at the national scale, CAMELS-PE uses ANA Water Availability Accreditation (ADH) records as proxies of potential water-allocation pressure. These records should not be interpreted as direct measurements of abstraction, consumption, or hydrological alteration. Spatial differences in ADH records may also reflect administrative reporting, formalisation of water uses, historical data availability, or already limited water availability in highly intervened basins. Reservoir-related values are mapped at the CAMELS-PE catchment outlets and should therefore be interpreted as catchment-scale summaries of reported reservoir infrastructure, rather than exact reservoir locations, polygons, or inundation extents.

These limitations should be considered when using CAMELS-PE for comparative hydrology, model benchmarking, regionalisation, and data-driven hydrological applications.

5 Code and data availability

The CAMELS-PE v1.0.1 dataset is available from Zenodo under the Creative Commons Attribution 4.0 International License (CC BY 4.0) at https://doi.org/10.5281/zenodo.21195425 (Llauca et al., 2026). The Zenodo repository contains the complete dataset release, including gauge metadata and dictionaries, catchment attributes, daily hydrometeorological time series, and geospatial layers with catchment boundaries and outlet locations. The RCamelsPE R package provides functions to access, visualise, and download CAMELS-PE. The package website, including documentation and usage examples, is

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

available at https://hllauca.github.io/RCamelsPE/, and the source code is hosted at https://github.com/hllauca/RCamelsPE. The package supports reproducible use of the dataset, although the full processing workflow used to generate all source-derived products is not distributed as part of this release.

6 Conclusions

This study presents CAMELS-PE, a large-sample hydrological dataset developed for Peru under the CAMELS framework. It provides daily hydrometeorological time series, metadata, geospatial layers, and catchment attributes for 136 catchments across the Pacific, Atlantic, and Titicaca hydrographic regions. By integrating observed and simulated streamflow with climatic, topographic, hydrological, geological, soil, land-cover, and human-intervention attributes, CAMELS-PE represents Peru's environmental and hydrological diversity within a standardised and reusable structure.

The dataset characterises strong spatial contrasts among catchments, from humid Amazonian headwaters with high runoff generation to highly seasonal Pacific catchments with greater water limitation and high-elevation Titicaca basins with endorheic conditions. These contrasts support the study of hydrological processes across diverse environments and the evaluation of model transferability, regionalisation, and prediction in ungauged basins.

CAMELS-PE facilitates reproducibility and intercomparison with other CAMELS datasets worldwide. Its reuse should consider uneven gauge availability, variable streamflow completeness, uncertainties in gridded and spatial products, model-based hydrological signatures, nested catchment dependence, and the interpretation of ADH records as proxies of potential water-allocation pressure. Overall, CAMELS-PE provides open and standardised information to support hydrological modelling, machine-learning applications, climate–streamflow analyses, drought and flood assessment, water-resources studies, and regional collaboration across tropical and Andean South America.

Author contributions

HLL and WLC initiated the investigation. HLL, CMC, and WLC designed the study. HLL, CMC and MGR processed the data. HLL created the figures. HLL and CMC computed the catchment attributes. HLL prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This research received no external funding.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

References

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

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, Hydrol. Sci. J., 65, 712–725, https://doi.org/10.1080/02626667.2019.1683182, 2020.

595 Almagro, A., Oliveira, P. T. S., Meira Neto, A. A., Roy, T., and Troch, P.: CABra: a novel large-sample dataset for Brazilian catchments, Hydrol. Earth Syst. Sci., 25, 3105–3135, https://doi.org/10.5194/hess-25-3105-2021, 2021.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrol. Earth Syst. Sci., 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018,

600 2018.

Arias, P. A., Garreaud, R., Poveda, G., Espinoza, J. C., Molina-Carpio, J., Masiokas, M., Viale, M., Scaff, L., and van Oevelen, P. J.: Hydroclimate of the Andes Part II: Hydroclimate Variability and Sub-Continental Patterns, Front. Earth Sci., 8, https://doi.org/10.3389/feart.2020.505467, 2021.

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., and Felipe-Obando, O.: Construction of a high-resolution

605 gridded rainfall dataset for Peru from 1981 to the present day, Hydrol. Sci. J., 65, 770–785, https://doi.org/10.1080/02626667.2019.1649411, 2020.

Bolton, D.: The computation of equivalent potential temperature, Mon. Weather Rev., 108, 1046–1053, https://doi.org/10.1175/1520-0493(1980)108%253C1046:tcoept%253E2.0.co;2, 1980.

Bushnell, M., Heitsenrether, R., Thomas, J., Galvarino, C., Burger, E., Dorton, J., and Leonard, L.: Status and near-term

610 plans for the U.S. IOOS Quality Assurance / Quality Control of Real-Time Oceanographic Data (QARTOD) project, in: OCEANS 2018 MTS/IEEE Charleston, OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, 22 October 2018, 1–6, https://doi.org/10.1109/oceans.2018.8604787, 2018.

Caro, A., Condom, T., Rabatel, A., Champollion, N., García, N., and Saavedra, F.: Hydrological response of Andean catchments to recent glacier mass loss, Cryosphere, 18, 2487–2507, https://doi.org/10.5194/tc-18-2487-2024, 2024.

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

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671

620 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., and Eijkhout, V.: River Network Routing on the NHDPlus Dataset, J. Hydrometeorol., 12, 913–934, https://doi.org/10.1175/2011JHM1345.1, 2011.

Delaigue, O., Guimarães, G. M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., and Andréassian, V.: CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, Earth Syst. Sci. Data, 17, 1461–1479, https://doi.org/10.5194/essd-17-1461-2025, 2025.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, https://doi.org/10.1016/j.jhydrol.2005.07.031, 2006.

Espinoza, J. C., Garreaud, R., Poveda, G., Arias, P. A., Molina-Carpio, J., Masiokas, M., Viale, M., and Scaff, L.: Hydroclimate of the Andes part I: Main climatic features, Front. Earth Sci., 8, https://doi.org/10.3389/feart.2020.00064, 2020.

Fathi, M. M. and Awadallah, A. G.: Regionalizing hydrologic information for runoff predictions beyond continental boundaries using machine learning, Adv. Water Resour., 206, 105162, https://doi.org/10.1016/j.advwatres.2025.105162, 2025.

Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth Syst. Sci. Data, 13, 3847–3867, https://doi.org/10.5194/essd-13-3847-2021, 2021.

Gutierrez-Villarreal, R. A., Junquas, C., Espinoza, J.-C., Baby, P., and Armijos, E.: Influence of local topographic structures on the atmospheric mechanisms related to the Andean-Amazon rainiest zone, Atmos. Res., 320, 108068, https://doi.org/10.1016/j.atmosres.2025.108068, 2025.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface: TECHNICAL BRIEF, Geochem. Geophys. Geosyst., 13, https://doi.org/10.1029/2012gc004370, 2012.

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., and Neal, J.: A 30 m global map of elevation with forests and buildings removed, Environ. Res. Lett., 17, 024016, https://doi.org/10.1088/1748-9326/ac4d4f, 2022.

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

Huerta, A., Bonnesoeur, V., Cuadros-Adriazola, J., Gutierrez, L., Ochoa-Tocachi, B. F., Román-Dañobeytia, F., and Lavado-Casimiro, W.: PISCOeo_pm, a reference evapotranspiration gridded database based on FAO Penman-Monteith in Peru, Scientific Data, 9, 1–18, https://doi.org/10.1038/s41597-022-01373-8, 2022.

© Author(s) 2026. CC BY 4.0 License.

Huerta, A., Aybar, C., Imfeld, N., Correa, K., Felipe-Obando, O., Rau, P., Drenkhan, F., and Lavado-Casimiro, W.: High-resolution grids of daily air temperature for Peru - the new PISCOt v1.2 dataset, Sci. Data, 10, 847, https://doi.org/10.1038/s41597-023-02777-w, 2023.

Huscroft, J., Gleeson, T., Hartmann, J., and Börker, J.: Compiling and mapping global permeability of the unconsolidated and consolidated earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0), Geophys. Res. Lett., 45, 1897–1904, https://doi.org/10.1002/2017gl075860, 2018.

Jimenez, D. A., Meneses, J. E., Solha, P. H. B., Avila-Diaz, A., Quesada, B., Melo Brentan, B., and Ferreira Rodrigues, A.: CAMELS-COL: A large-sample hydrometeorological dataset for Colombia, Earth Syst. Sci. Data Discuss., 1–38, https://doi.org/10.5194/essd-2025-200, 2025.

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

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Sci. Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A standard approach to baseflow separation using the Lyne and hollick filter, Australas. J. Water Resour., 17, 25–34, https://doi.org/10.7158/13241583.2013.11465417, 2013.

Llauca, H., Lavado-Casimiro, W., León, K., Jimenez, J., Traverso, K., and Rau, P.: Assessing Near Real-Time Satellite Precipitation Products for Flood Simulations at Sub-Daily Scales in a Sparsely Gauged Watershed in Peruvian Andes, Remote Sensing, 13, 826, https://doi.org/10.3390/rs13040826, 2021a.

Llauca, H., Lavado-Casimiro, W., Montesinos, C., Santini, W., and Rau, P.: PISCO_HyM_GR2M: A Model of Monthly Water Balance in Peru (1981–2020), Water, 13, 1048, https://doi.org/10.3390/w13081048, 2021b.

Llauca, H., Arestegui, M., and Lavado-Casimiro, W.: Constraining Flood Forecasting Uncertainties through Streamflow Data Assimilation in the Tropical Andes of Peru: Case of the Vilcanota River Basin, Water, 15, 3944, https://doi.org/10.3390/w15223944, 2023a.

Llauca, H., Leon, K., and Lavado-Casimiro, W.: Construction of a daily streamflow dataset for Peru using a similarity-based regionalization approach and a hybrid hydrological modeling framework, Journal of Hydrology: Regional Studies, 47, 101381, https://doi.org/10.1016/j.ejrh.2023.101381, 2023b.

Llauca, H., Montesinos-Caceres, C. Gutierrez-Reynaga, M., and Lavado-Casimiro, W.: CAMELS-PE: Catchment Attributes and Meteorology for Large-sample Studies in Peru, Zenodo [data set], https://doi.org/10.5281/zenodo.20058779, 2026.

López-Ballesteros, A., Nielsen, A., Castellanos-Osorio, G., Trolle, D., and Senent-Aparicio, J.: DSOLMap, a novel high-resolution global digital soil property map for the SWAT + model: Development and hydrological evaluation, Catena, 231, 107339, https://doi.org/10.1016/j.catena.2023.107339, 2023.

Proyecto MapBiomas Ecuador: Mapa anual de cobertura y uso de la tierra en Ecuador – Colección 3, available at: https://plataforma.ecuador.mapbiomas.org/, last access: 14 December 2025, 2025.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Proyecto MapBiomas Perú: Mapa anual de cobertura y uso de la tierra en el Perú – Colección 3, available at: https://plataforma.peru.mapbiomas.org/, last access: 14 December 2025, 2025.

Muñoz, R., Vaghefi, S. A., Drenkhan, F., Santos, M. J., Viviroli, D., Muccione, V., and Huggel, C.: Assessing water management strategies in data-scarce mountain regions under uncertain climate and Socio-economic changes, Water Resour. Manage., https://doi.org/10.1007/s11269-024-03853-5, 2024.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Jean-Noël Thépaut: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Rollenbeck, R., Orellana-Alvear, J., Bendix, J., Rodríguez, R., Pucha-Cofrep, F., Guallpa, M., Fries, A., and Célleri, R.: The coastal El Niño event of 2017 in Ecuador and Peru: A weather radar analysis, Remote Sens, 14, 824, https://doi.org/10.3390/rs14040824, 2022.

Sanabria, J., Bourrel, L., Dewitte, B., Frappart, F., Rau, P., Solis, O., and Labat, D.: Rainfall along the coast of Peru during strong El Niño events, Int. J. Climatol., 38, 1737–1747, https://doi.org/10.1002/joc.5292, 2018.

Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water Resour. Res., 37, 1771–1781, https://doi.org/10.1029/2000wr900330, 2001.

Sulca, J., Takahashi, K., Espinoza, J.-C., Tacza, J., Zubieta, R., Mosquera, K., and Apaestegui, J.: A multiple linear regression model for the prediction of summer rainfall in the northwestern Peruvian Amazon using large-scale indices, Clim. Dyn., 62, 4431–4451, https://doi.org/10.1007/s00382-023-07044-7, 2024.

Todini, E.: The ARNO rainfall—runoff model, J. Hydrol., 175, 339–382, https://doi.org/10.1016/S0022-1694(96)80016-3, 1996.

Turner, S., Hannaford, J., Barker, L. J., Suman, G., Killeen, A., Armitage, R., Chan, W., Davies, H., Griffin, A., Kumar, A., Dixon, H., Albuquerque, M. T. D., Almeida Ribeiro, N., Alvarez-Garreton, C., Amoussou, E., Arheimer, B., Asano, Y., Berezowski, T., Bodian, A., Boutaghane, H., Capell, R., Dakhlaoui, H., Daňhelka, J., Do, H. X., Ekkawatpanit, C., El Khalki, E. M., Fleig, A. K., Fonseca, R., Giraldo-Osorio, J. D., Goula, A. B. T., Hanel, M., Horton, S., Kan, C., Kingston, D. G., Laaha, G., Laugesen, R., Lopes, W., Mager, S., Rachdane, M., Markonis, Y., Medeiro, L., Midgley, G., Murphy, C., O'Connor, P., Pedersen, A. I., Pham, H. T., Piniewski, M., Renard, B., Saidi, M. E., Schmocker-Fackel, P., Stahl, K., Thyer, M., Toucher, M., Tramblay, Y., Uusikivi, J., Venegas-Cordero, N., Visessri, S., Watson, A., Westra, S., and Whitfield, P. H.: ROBIN: Reference observatory of basins for international hydrological climate change detection, Sci. Data, 12, 654, https://doi.org/10.1038/s41597-025-04907-y, 2025.

Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, Adv. Water Resour., 32, 1465–1481, https://doi.org/10.1016/j.advwatres.2009.06.011, 2009.

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Open Access Earth System Science Data Discussions

Appendix A. Streamflow quality-control procedure and diagnostics

The streamflow quality-control procedure described in Sect. 3.2.1 was applied independently to each observed streamflow series. This appendix summarises the flagging scheme and diagnostic tests used to screen the original records. Failed records were removed from the quality-controlled streamflow series, whereas missing observations were retained as gaps.

Before quality control, discharge records were converted from $m^3\ s^{-1}$ to catchment-equivalent runoff depth (mm d^{-1}) using catchment drainage area. This normalisation allowed the screening criteria, particularly the flatline threshold, to be applied consistently across catchments with contrasting drainage areas and flow magnitudes. The tests were applied to streamflow expressed as catchment-equivalent runoff depth, hereafter denoted as q_t.

The primary flags were defined as follows: Flag 1 indicates that the observation passed all tests; Flag 4 indicates that the observation failed at least one test and was removed from the distributed series; and Flag 9 indicates that the observation was missing in the original record and retained as a gap.

Three fail-only tests were applied to each gauge series. The gross-range test identified physically invalid negative streamflow values:

$$q_t < 0 \Rightarrow \text{fail.} \tag{A1}$$

The spike test identified isolated anomalous values by comparing each observation with the mean of the previous and following valid observations:

$$\left| q_t - \frac{(q_{t-1} + q_{t+1})}{2} \right| > 4\sigma_q \Rightarrow \text{fail,} \tag{A2}$$

where q_t is the streamflow value expressed as catchment-equivalent runoff depth at day t, q_{t-1} and q_{t+1} are the previous and following valid observations, and σ_q is the standard deviation of the corresponding gauge series expressed in mm d^{-1}. The first 30 valid observations of each gauge were excluded from this test as an initial warm-up period.

The flatline test identified persistent repeated or near-repeated values. A sequence was flagged when at least seven consecutive valid observations showed day-to-day differences smaller than 0.001 mm d^{-1}:

$$|q_t - q_{t-1}| < 0.001 \text{ mm d}^{-1}, \tag{A3}$$

for at least seven consecutive days. The first 30 valid observations of each gauge were also excluded from this test. Because stable low-flow periods may occur naturally in some catchments, particularly in small, regulated, or groundwater-influenced systems, flatline flags were interpreted conservatively and visually inspected before removal to reduce the risk of excluding plausible hydrological behaviour.

The final quality-controlled streamflow series, expressed as catchment-equivalent runoff depth, was generated as:

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

$$q_t^{QC} = \begin{cases} q_t, & \text{if } flag_t = 1 \\ NA, & \text{if } flag_t = 4 \text{ or } 9 \end{cases} \quad \text{(A4)}$$

Thus, failed observations and originally missing values are both represented as missing values in the distributed streamflow series expressed as catchment-equivalent runoff depth, although they correspond to different flag categories during the quality-control procedure.

To quantify the effect of the screening procedure, quality-control flags were summarised across all observed streamflow records after conversion to catchment-equivalent runoff depth (Table A1). Missing observations represented 61.51 % of the complete 1981–2025 calendar, reflecting variable temporal coverage and continuity across the hydrometric records. Among originally available observations, 6.53 % were flagged as failed and removed from the quality-controlled streamflow series. Most failures were associated with the flatline test, which accounted for 99.38 % of all failed records, whereas spike and gross-range failures were rare. These diagnostics indicate that the procedure mainly removed persistent repeated or near-repeated values rather than physically invalid negative values or isolated spikes.

Table A1. Summary of streamflow quality-control results across the CAMELS-PE hydrometric records.

Diagnostic	Value
Total daily entries in common calendar	2,235,296
Originally available observations	860,306
Missing observations	1,374,990
Observations passing all tests	804,129
Observations flagged as failed	56,177
Failed observations relative to available observations	6.53 %
Missing observations relative to complete calendar	61.51 %
Median failed observations per gauge	1.91 %
Interquartile range of failed observations per gauge	0.14–8.19 %
Maximum failed observations in one gauge	41.63 %
Flatline failures	55,826
Spike failures	336
Gross-range failures	15

Appendix B. Performance of the PISCO-ARNOVIC simulated streamflow series

The PISCO-ARNOVIC simulated streamflow series were evaluated against the available quality-controlled observations at the daily time step, using the overlapping period between observed and simulated records for each gauge. Missing values were excluded, and both series were evaluated in catchment-equivalent runoff depth units (mm d^{-1}).

https://doi.org/10.5194/essd-2026-386
Preprint. Discussion started: 10 July 2026
© Author(s) 2026. CC BY 4.0 License.

Performance was assessed using the Kling–Gupta efficiency (KGE), the square-root transformed Nash–Sutcliffe efficiency (sqrtNSE), and percentage bias (PBIAS). These metrics provide complementary diagnostics of integrated model performance, hydrograph agreement with reduced dominance of high-flow errors, and volumetric bias, respectively.

The Kling–Gupta efficiency was calculated as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \quad \text{(B1)}$$

where r is the Pearson correlation coefficient, α is the ratio between the standard deviations of simulated and observed streamflow, and β is the ratio between their means.

The square-root transformed Nash–Sutcliffe efficiency was calculated as:

$$sqrtNSE = 1 - \frac{\sum_t\left(\sqrt{q_{obs,t}} - \sqrt{q_{sim,t}}\right)^2}{\sum_t\left(\sqrt{q_{obs,t}} - \overline{\sqrt{q_{obs}}}\right)^2}, \quad \text{(B2)}$$

The square-root transformation was used instead of the standard NSE to reduce the dominance of high-flow errors.

Percentage bias was calculated as:

$$PBIAS = 100\frac{\sum_t(q_{obs,t} - q_{sim,t})}{\sum_t q_{obs,t}}, \quad \text{(B3)}$$

Following this convention, positive PBIAS indicates underestimation and negative PBIAS indicates overestimation of total runoff volume.

Table B1. Summary performance statistics of the PISCO-ARNOVIC simulated streamflow series.

Metric	Mean	Median	SD	P25	P75
KGE	0.68	0.75	0.20	0.55	0.84
sqrtNSE	0.48	0.60	0.34	0.28	0.76
PBIAS (%)	-0.67	-0.30	2.28	-1.10	0.30

The diagnostics indicate generally reasonable national-scale agreement between simulated and observed streamflow for deriving model-consistent hydrological signatures. Median KGE reached 0.75 and median sqrtNSE reached 0.60, while median PBIAS was close to zero (-0.30 %), indicating limited systematic volumetric bias. Nevertheless, the spread in KGE and sqrtNSE values shows that performance varies among catchments. Therefore, simulated streamflow and derived signatures should be interpreted as model-consistent descriptors for comparative large-sample analyses rather than as direct substitutes for observed streamflow in catchment-specific applications.