https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Open Access Earth System Science Data Discussions

A multidimensional daily precipitation dataset (1998-2024) over the Third Pole with uncertainty estimates and precipitation phase information

Yi Lyu[1, 2], Bin Yong[1, 2*]
[1] *State Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, China.*
[2]*College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China.*

Correspondence to: Bin Yong (yongbin@hhu.com)

Abstract. Accurate precipitation data for the Tibetan Plateau (TP) is critical for understanding regional water resources and global climate dynamics. However, existing datasets struggle with observational bottlenecks due to complex terrain and high precipitation variability. Here, we introduce the CRISP (Conformal Regression and Integrated Stacking for Precipitation) dataset, providing a continuous, 27-year (1998-2024) daily precipitation record at a 0.1° spatial resolution for the TP. Unlike conventional statistical datasets, CRISP integrates physical atmospheric conditions and terrain features into a machine-learning framework. The CRISP dataset effectively identifies false drizzle signals commonly seen in reanalysis data without missing most real rainfall events. Furthermore, CRISP can provide reliable 90% uncertainty intervals that adaptively adjust to the intensity of rainfall compared to other datasets, and it explicitly classifies precipitation into rain, snow, and mixed phases to directly support cryospheric research. The independent validations indicate that CRISP shows potential in reducing inconsistencies associated with the "Third Pole precipitation paradox" (Miao et al., 2024). By providing consistent precipitation estimates together with uncertainty information and phase classification, CRISP offers a valuable basis for hydroclimatic research over the Third Pole and its downstream regions. The CRISP dataset is openly available at (https://doi.org/10.11888/Atmos.tpdc.303469; Yong & Lyu, 2026).

1 Introduction

The Tibetan Plateau (TP), widely known as the "Third Pole" and the "Asian Water Tower", sustains the livelihoods of nearly two billion people through its wide river systems (Shukla & Sen, 2021; Yao et al., 2022). Because of its extreme elevation and strong land-atmosphere coupling, the TP is highly sensitive to global climate change (You et al., 2021). This sensitivity is evident in rapid glacier retreat, expanding alpine lakes, and pronounced shifts in the regional water balance (Zhang et al., 2015; Su et al., 2022). Understanding these hydroclimatic transformations requires accurate and long-term precipitation estimates, as precipitation is the driver of the energy and water cycles on the TP (Zhang et al., 2017). However, extreme topography, harsh climatic conditions, and the sparse distribution of in situ observations make reliable precipitation quantification a persistent challenge (Gao & Liu, 2013).

To compensate for limited gauge coverage, numerous global precipitation products based on satellite retrievals and atmospheric reanalysis have been developed (Lyu et al., 2024). Such datasets can provide broad spatial coverage and high

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

temporal resolution, showing high potential for regional hydrological studies (Tong et al., 2014). While the reliability of these products is limited by sensor sensitivity and retrieval algorithms, a more critical constraint lies in their heavy dependence on ground-based calibration (Kidd & Levizzani, 2011). These reference observations suffer from systematic biases, as the sparse measurement network fails to capture the precipitation variability over the TP (Tang et al., 2018).

To address these limitations, researchers have developed regional datasets by integrating available dense gauge observations with high-resolution atmospheric simulations. For example, the Tibetan Plateau High-resolution Precipitation (TPHiPr) dataset improves precipitation estimates through dense observation correction (Jiang et al., 2023). Despite these advances, two methodological challenges remain. First, statistical interpolation methods tend to be ineffective in ungauged regions. In these areas, the highly nonlinear relationships between precipitation and complex topography cannot be simulated by simple spatial extrapolation (Das, 2019). Second, many regional products rely on global reanalysis datasets as background fields. This dependence causes them to propagate systematic biases inherited from the parent models, including the overestimation of light precipitation frequency and the underestimation of extreme rainfall intensity (Ward et al., 2011).

In addition to biases in precipitation intensity, another major limitation of existing datasets is the lack of reliable uncertainty quantification. Our previous work introduced a double machine learning framework tailored for the TP and its sub-region (Lyu and Yong, 2024; 2025). While this approach outperformed benchmarks under specific condition (e.g., MSP and TPHiPr) and has seen widespread application, it primarily focused on deterministic accuracy. Without reliable uncertainty intervals, users cannot accurately evaluate how these errors propagate into downstream hydrological models (McMillan et al., 2018). Furthermore, this earlier framework ignored critical physical factors, including wind-induced undercatch and precipitation phase discrimination. The absence of phase information is also particularly restrictive in cryospheric studies, such as snowmelt runoff simulations and glacier mass balance assessments.

To overcome these gaps, we introduce the Conformal Regression and Integrated Stacking for Precipitation (CRISP) framework, offering a multidimensional, daily 0.1° precipitation dataset explicitly designed for the TP. As outlined in the Figure 1, CRISP is introduced as a comprehensive NetCDF4 data cube that advances beyond conventional statistical merging by sharing three core advantages. First, it employs physics-aware corrections by integrating multi-source data with static topography and dynamic atmospheric variables, compensating for wind-induced undercatch in complex terrain. Second, it utilizes distribution-free conformalized quantile regression to generate uncertainty bounds. Third, it incorporates a thermodynamic phase discrimination module to identify the physical state of precipitation. The CRISP data can can meet the requirements of different users. The deterministic layers provide the best estimate (P_Best) and climatology-corrected precipitation (P_Clim) for generalized weather analysis and hydrological modeling. The probabilistic layers offer precipitation probability (Prob) and calibrated 5th and 95th prediction intervals (P_5th, P_95th) for extreme event detection and risk assessment. Finally, the physical state layer (Phase) classifies events into dry, snow, mixed, or rain conditions to support cryospheric studies.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

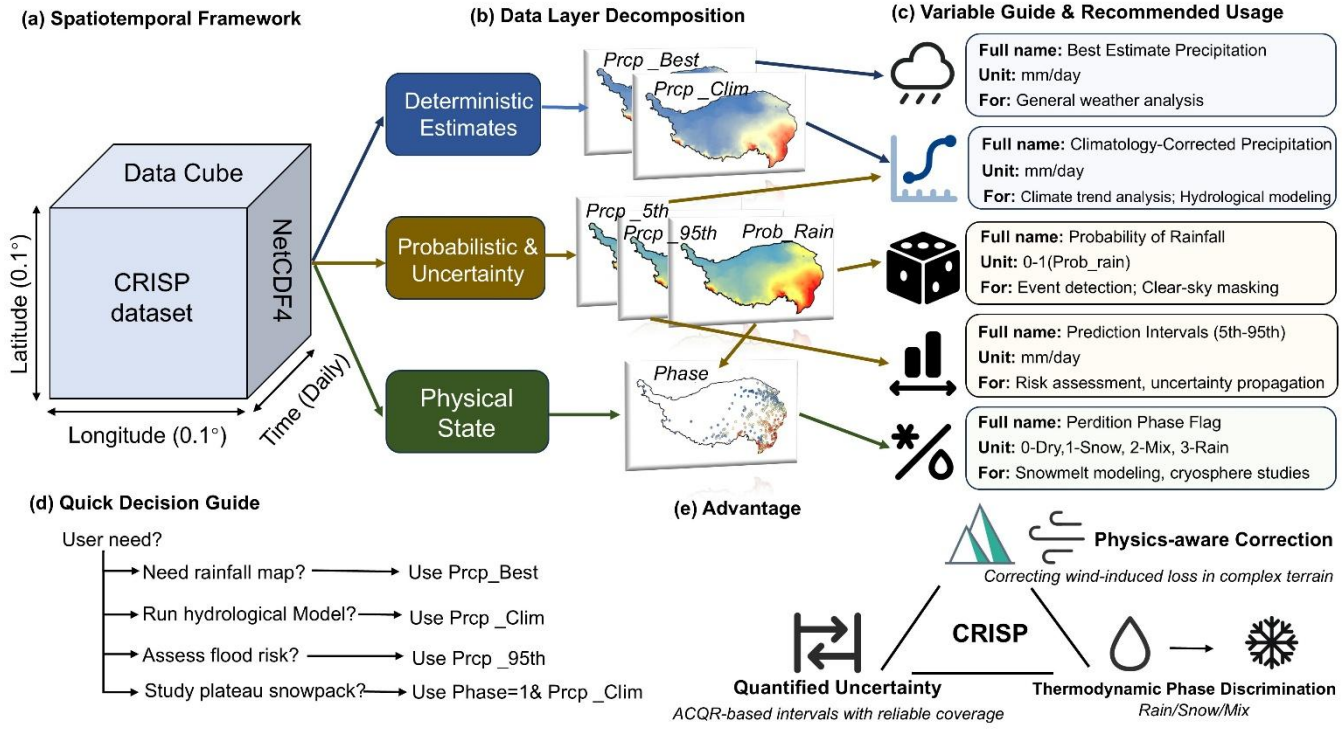


Figure 1. Structural blueprint and multidimensional architecture of the CRISP dataset. (a) The spatiotemporal framework, structured as a high-resolution 0.1° daily data cube in NetCDF4 format. **(b)** Decomposition of the dataset into three functional layers: deterministic estimates, probabilistic uncertainty bounds, and physical state classifications. **(c)** A comprehensive variable guide detailing the output variables (Prcp_Best, Prcp_Clim, Prob_Rain, Prcp_5th, Prcp_95th, and Phase) alongside their recommended hydroclimatic applications. **(d)** A quick decision-making workflow tailored for diverse user needs, from general meteorological analysis to specific cryospheric modeling and risk assessments. **(e)** The three core methodological innovations driving the CRISP framework.

2 Data and Study region

2.1 Dynamic input data

To capture the spatiotemporal pattern of precipitation across the TP, we integrated state-of-the-art satellite retrievals and atmospheric reanalysis. For the satellite-based inputs, we selected two high-resolution, gauge-adjusted products at a 0.1° daily resolution: the Integrated Multi-satellitE Retrievals for GPM (IMERG) Final Run V07 (Huffman et al., 2015) and the Global Satellite Mapping of Precipitation (GSMaP) gauge-adjusted V08 (Kubota et al., 2020). IMERG serves as a Level-3 system that synthesizes constellation microwave estimates, infrared satellite estimates, and monthly rain gauge data from the Global Precipitation Climatology Centre (GPCC). The accuracy of these retrievals is enhanced through calibration using Dual-

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

frequency Precipitation Radar (DPR) and GPM Microwave Imager (GMI) data (Tang et al., 2020). Furthermore, for precipitation phase partitioning, we incorporated the "probabilityLiquidPrecipitation" variable provided by the IMERG dataset. This inclusion introduces a probabilistic constraint on the hydrometeor state to improve the identification of rain-to-snow transitions. Similarly, GSMaP is a quasi-global product adjusted by National Oceanic and Atmospheric Administration (NOAA) gauge observations. While it shares the same resolution of IMERG, it relies solely on DPR data as its calibration standard for microwave retrievals, providing complementary satellite information (Kubota et al., 2017).

To provide dynamic atmospheric forcing and background precipitation fields, we utilized the standard ERA5 (Hersbach et al., 2020) global reanalysis (0.25°, hourly), which ensures physical and thermodynamic consistency through four-dimensional variational data assimilation (4D-Var). Although ERA5-Land (Muñoz-Sabater et al., 2021) offers a higher spatial resolution, the standard ERA5 was chosen for this framework. This decision is because ERA5-Land occasionally exhibits null values over complex TP river networks. More importantly, standard ERA5 avoids the potential biases inherent in ERA5-Land on near-surface atmospheric parameterizations in extreme high-altitude regions (Gomis-Cebolla et al., 2023). Besides total precipitation, we selected five key factors to capture the physical mechanisms of precipitation formation: 2 m air temperature (T) to represent thermal capacity and water vapor limits; wind field vectors (U_{10}, V_{10}) to characterize dynamic moisture transport; 2 m dew point temperature (D) to indicate atmospheric humidity and saturation; surface pressure (SP) as a proxy for atmospheric convergence and vertical motion; and surface net solar radiation (SSR) to account for the surface energy budget driving local convection (Table 1). Specifically, the U_{10}, V_{10} wind fields were vectorially combined to represent the scalar wind speed (WS), characterizing the dynamic intensity of moisture transport:

$$WS = \sqrt{{U_{10}}^2 + {V_{10}}^2} \tag{1}$$

2.2 Static input data

The static environmental framework of the CRISP dataset is derived from the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM). The SRTM provides near-global coverage at a high spatial resolution and serves as the elevation grid (*Elev*) to account for baseline orographic effects across the TP. To match the spatial resolution of the dynamic inputs, the raw DEM was resampled to a 0.1° grid.

To consider the complex physical interactions, we derived three morphometric factors from the elevation grid. The primary terrain factors, slope (*Slp*) and aspect (*Asp*), were calculated based on the spatial gradients of the elevation in the longitudinal (x) and latitudinal (y) directions. The slope represents the maximum rate of change in elevation and the intensity of local orographic forcing. It was computed as the magnitude of the gradient vector transformed into degrees:

$$Slp = \arctan\left(\sqrt{\left(\frac{\partial Elev}{\partial x}\right)^2 + \left(\frac{\partial Elev}{\partial y}\right)^2}\right) \cdot \frac{180}{\pi} \tag{2}$$

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Aspect, representing the orientation of the surface, is a critical parameter for distinguishing humid windward slopes from leeward rain shadows. It was derived according to the direction of the maximum spatial gradient:

$$Asp = atan2d\left(\frac{\partial Elev}{\partial x}, -\frac{\partial Elev}{\partial y}\right) \quad (3)$$

To capture the regional topographic context and meso-scale terrain influences on vertical moisture transport, we implemented a window-based statistical aggregation. A 3×3 moving window was applied to the elevation, slope, and aspect grids to calculate their respective neighborhood means. This moving average approach smoothed local noise while preserving the large-scale geomorphological features that govern adiabatic cooling and regional air mass lifting. Furthermore, Terrain Roughness (TR) was quantified as the local elevation heterogeneity. Following standard geomorphometric practices, it was calculated as the local standard deviation of elevation within the 3×3 window:

$$TR = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\mathrm{Elev} - \overline{\mathrm{Elev}}\right)^2} \quad (4)$$

where N=9 represents the number of pixels within the neighborhood. This descriptor effectively captures surface turbulence and microscale heterogeneity that influence local precipitation distribution.

2.3 Ground and comparative data

To train the models, we utilized daily precipitation records from the National Meteorological Information Centre (NMIC) as the primary target. This dataset integrates observations from approximately 12,000 global stations leveraging the Integrated Surface Dataset (ISD) and the Global Telecommunication System (GTS), providing a robust representation of regional climate characteristics (Yang et al., 2020). We strictly extracted a localized subset of 484 stations geographically distributed within our study domain.

To ensure data integrity, a stringent quality control (QC) protocol was implemented prior to model training. Consequently, as illustrated in Figure 2b, the number of active and highly reliable daily station records available for training dynamically fluctuated over the study period, expanding from 40,100 in 2001 to over 100,000 in recent years. It is important to acknowledge that global gauge-adjusted satellite products evaluated in this study (e.g., IMERG calibrated via GPCC, and GSMaP adjusted via CPC) inherently share a portion of these GTS-based stations for their baseline calibrations. However, traditional large-scale satellite calibration algorithms frequently neglect local wind-induced solid precipitation undercatch and complex sub-grid orographic lifting.

Prior to serving as the training target, these raw gauge observations were systematically adjusted to compensate for wind-induced undercatch. Relying on concurrent near-surface air temperature (T, °C) and 10-m wind speed (WS), which were directly extracted from the NMIC station streams and systematically substituted with co-located ERA5 reanalysis data in instances of missing station observations. We also derived the gauge-height wind speed (WS_g, constrained to a minimum of

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

0.1m/s) via an empirical scaling factor of 0.7. The aerodynamic catch ratio (CR) was dynamically calculated based on precipitation phase thresholds:

$$CR_{rain} = \exp(-0.03 \cdot WS_g), T > 2°C \quad (5)$$

$$CR_{snow} = \exp(-0.20 \cdot WS_g), T < -2°C \quad (6)$$

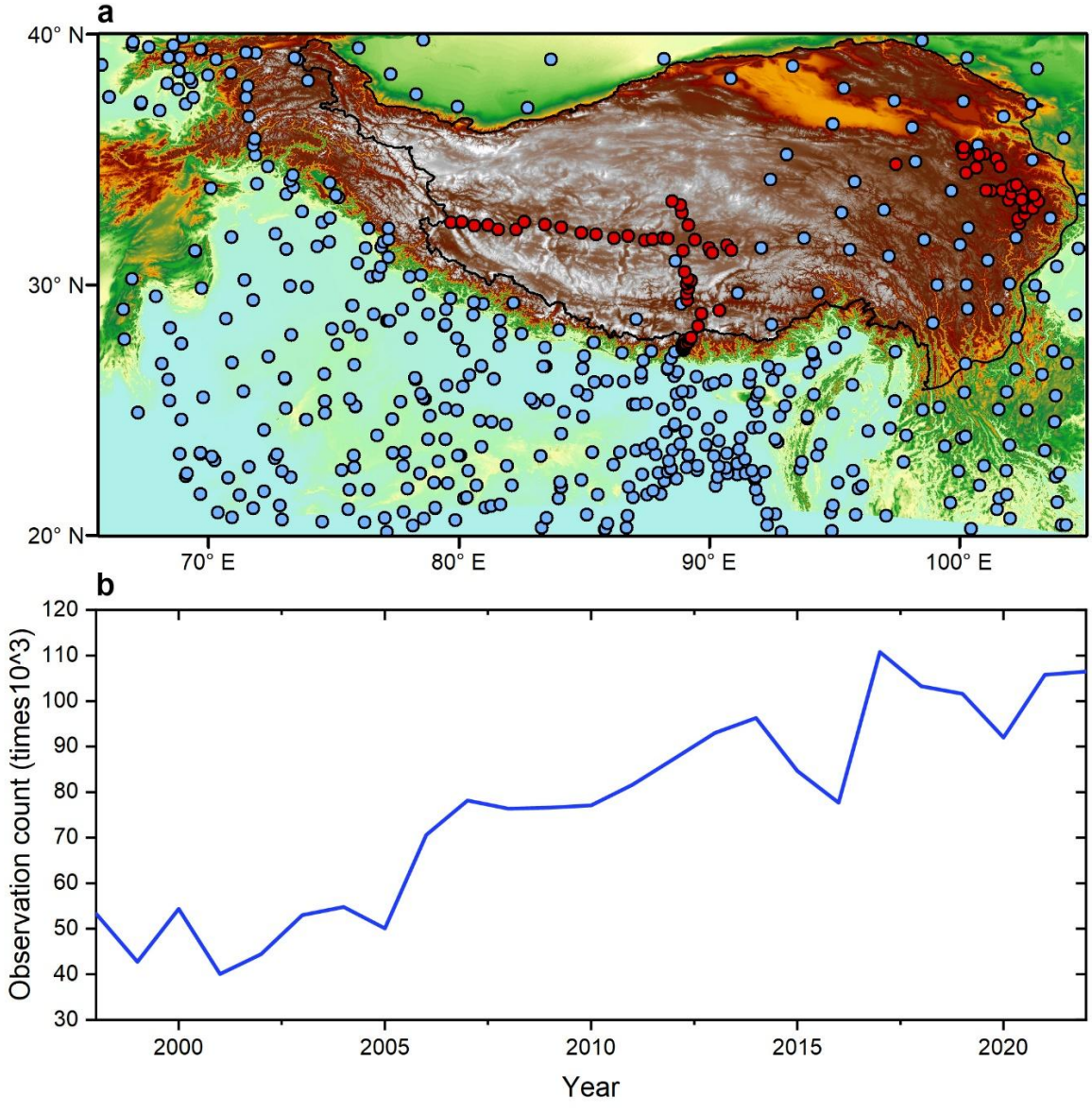
For mixed-phase precipitation (-2℃⩽T⩽2℃), CR was determined through linear interpolation:

$$CR_{mixed} = CR_{snow} + \frac{T-(-2)}{4} \cdot (CR_{rain} - CR_{snow}) \quad (7)$$

To prevent numerical instability under extreme wind conditions, a strict lower bound of 0.2 was enforced for all catch ratios. The actual precipitation mass (P_c) was then reconstructed as $P_c = P_{obs} / CR$. Subsequently, to align these point-based observations with the 0.1° gridded framework, a nearest-neighbor approach was employed to map each station to its corresponding grid cell. In instances where multiple stations fell within the same 0.1° grid cell, their corrected precipitation values were arithmetically averaged to generate a single, unique grid-level training target. By utilizing these physics-corrected stations as our training target, the CRISP framework explicitly learns to model and correct the unresolved physical residuals typical of solid precipitation in alpine environments.

For independent validation, we incorporated supplementary datasets completely excluded from the model training phase due to their limited temporal coverage. These include hydrological observations from the Yellow River Conservancy Commission (Lyu & Yong, 2025) and a specialized 53-station high-altitude network (Zhan et al., 2022). The latter consists of spatial transects equipped with tipping-bucket rain gauges (0.2 mm precision), providing critical warm-season hourly data from 2017 to 2020 (locations detailed in Figure 2).

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026



155

Figure 2. Spatial distribution and temporal availability of precipitation observations across the Tibetan Plateau. (a) Topographic map of the study domain featuring the spatial distribution of rain gauges. Blue circles represent the integrated long-term station network (derived from NMIC, ISD, and GTS) utilized for model training and cross-validation. Red circles denote the independent observation gauges, including the Yellow River Conservancy Commission stations and the specialized 160 53-station high-altitude network, reserved exclusively for independent evaluation. **(b)** Temporal evolution of the available observation records for the integrated long-term network. The elevation background is derived from the SRTM DEM.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

To evaluate the performance of CRISP against state-of-the-art regional efforts, we selected the TPHiPr dataset as our primary comparative baseline. TPHiPr is a high-resolution (1/30°, daily) precipitation product explicitly designed for the Third Pole, covering the period from 1979 to 2020 (Jiang et al., 2023). It was produced by merging the atmospheric simulation-based ERA5_CNN with observations from more than 9,000 rain gauges using climatologically aided interpolation and random forest methods. Previous evaluations demonstrate that TPHiPr achieves remarkable accuracy in densely gauged areas, significantly outperforming global products such as IMERG, MSWEP, and ERA5 (Cheng et al., 2025). Because TPHiPr represents the highest standard of deterministic statistical interpolation currently available for the plateau, it serves as an ideal reference to evaluate the added value of the physics-aware corrections and conformalized uncertainty bounds introduced by the CRISP framework.

Table 1. Detailed information of used data.

Acronym	Full Name	Variable	Used Data	Original Resolution	Category
IMERG	Integrated Multi-satellitE Retrievals for GPM Final Run V07	Precipitation	Precipitation	Daily /0.1°	Dynamic input
GSMaP	Global Satellite Mapping of Precipitation (GSMaP) gauge-adjusted V08	Precipitation	Precipitation	Daily /0.1°	
ERA5	European Centre for Medium-range Weather Forecast Re Analysis 5	Total precipitation 2m temperature 2m dewpoint temperature Surface pressure 10m u-component of wind 10m v-component of wind Surface net solar radiation	Total precipitation 2m temperature 2m dewpoint temperature Surface pressure Scalar wind speed Surface net solar radiation	Daily /0.25°	
DEM	Shuttle Radar Topography Mission Digital Elevation Model	Elevation	Elevation Slope Aspect Terrain roughness	90m	Static input
NMIC	daily precipitation records from the National Meteorological Information Centre	Precipitation	Precipitation	Daily /point	Label
TPHiPr	a long-term (1979–2020) high-resolution (1/30°, daily) precipitation dataset	Precipitation	Precipitation	Daily /1/30°	Reference

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

2.4 Study region

The analysis domain spans 20° to 40°N and 65° to 105°E, which intentionally extends beyond the geographical boundary of the plateau to minimize potential edge effects during data processing and model simulation. For subsequent statistical validation and comparison, all datasets were strictly clipped to the actual geographic extent of the Tibetan Plateau, which encompasses approximately 2.5 million km² with a mean elevation exceeding 4,000 m (Wu et al., 2015; Figure 2). Widely recognized as the "Asian Water Tower", this region sustains the headwaters of major continental rivers, including the Indus, Ganges, Yellow, and Yangtze Rivers (Cui & Graf, 2009). Because of its massive elevated heating source and profound topographic relief, the plateau critically modulates the Asian monsoon system and mid-latitude westerlies, showing high sensitivity to global climate change (Yong et al., 2024). The interaction between these large-scale atmospheric circulations and the extreme orography produces unique precipitation patterns. The region transitions sharply from monsoon-dominated, humid margins in the southeast to hyper-arid interiors and westerly-dominated, glacierized mountains in the west. Accurate precipitation quantification across this domain is therefore essential for understanding regional water resources and hydroclimatic variability affecting nearly two billion people across Asia (Liu & Chen, 2000).

3 Data processing and methodology

The development of the CRISP dataset is driven by a hybrid structure that seamlessly integrates machine learning with physics-aware constraints. As illustrated in Figure 3, the framework consists of three sequential modules: spatiotemporal alignment and feature engineering, the CRISP machine learning engine, and physics-guided post-processing.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

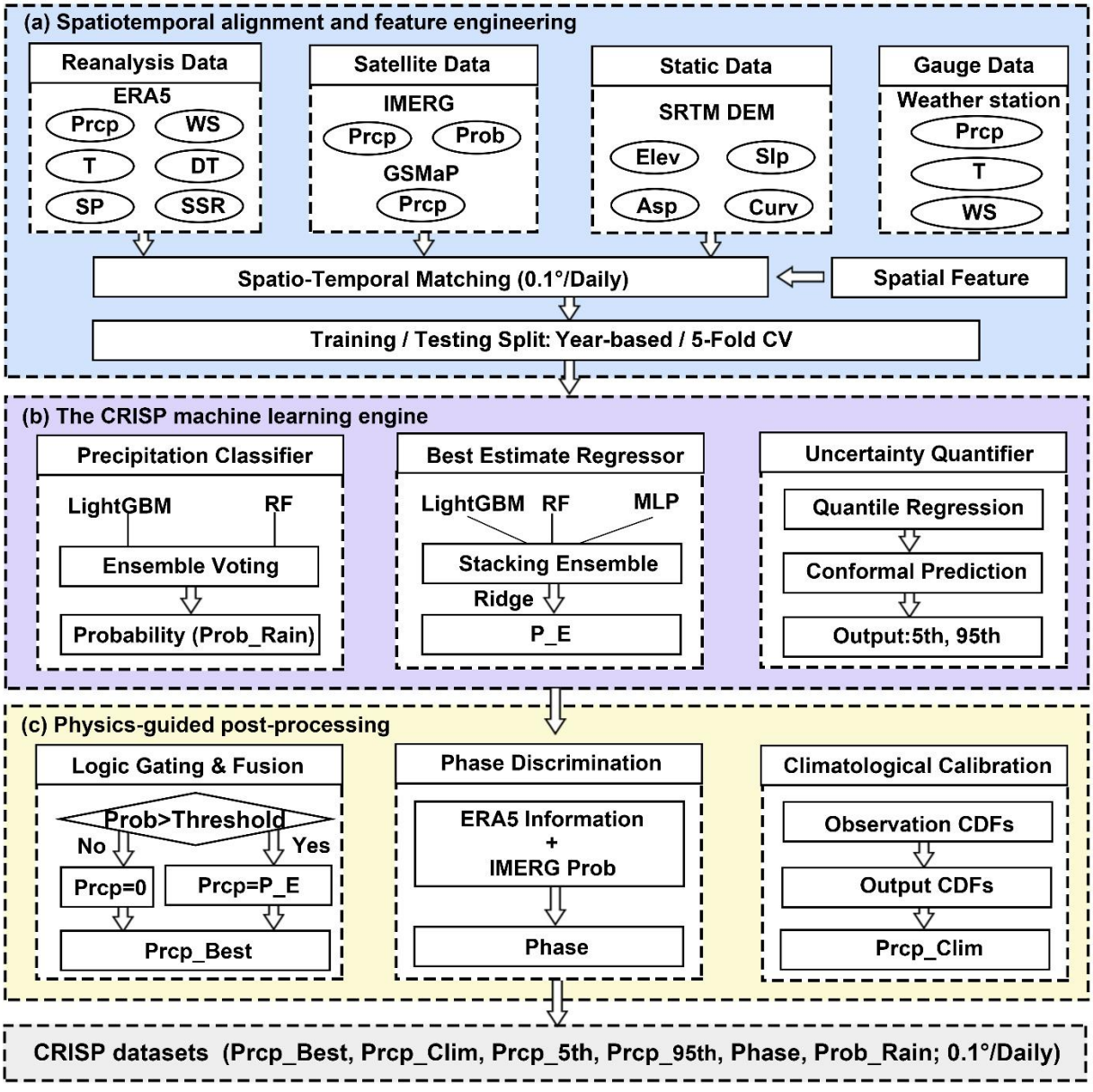


Figure 3. Methodological flowchart of the CRISP framework. (a) Spatiotemporal alignment and feature engineering, integrating static geomorphometric factors with dynamic atmospheric forcing to support physical parameterizations. **(b)** The CRISP machine learning engine, featuring an ensemble classifier for rain probability (Prob_Rain), a stacking regressor for optimal deterministic estimates (P_E), and an Asymmetric Conformalized Quantile Regression (ACQR) module for uncertainty bounds (5th and 95th percentiles). **(c)** Physics-guided post-processing, encompassing expected-value logic gating to suppress false alarms, thermodynamic phase discrimination for cryospheric applications, and Cumulative Distribution Function (CDF)-based climatological calibration to long-term research.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

3.1 Spatiotemporal alignment and feature engineering

To construct a consistent multidimensional data cube, all dynamic atmospheric and static geomorphometric variables were systematically aligned. Variables with coarser native resolutions, such as the 0.25° ERA5 reanalysis, were resampled to a uniform 0.1° grid utilizing bilinear interpolation to ensure spatial consistency (Kirkland, 2010). The station observations were subjected to quality control procedures, including the removal of abnormal records and missing-value screening, prior to model development and validation. Concurrently, all temporal variables were aggregated to a daily resolution using Coordinated Universal Time (UTC) to maintain temporal consistency among multi-source datasets.

To explicitly incorporate mesoscale atmospheric backgrounds and capture local topographic gradients, we derived neighborhood statistics for the three primary precipitation inputs (IMERG, GSMaP, and ERA5). A 5×5 moving window was applied to compute the spatial mean, representing the broad regional precipitation forcing, while a 3×3 window was utilized to calculate the spatial standard deviation, quantifying microscale variability. These derived spatial metrics were integrated with static factors (elevation, slope, aspect, and terrain roughness) and dynamic atmospheric state variables (wind speed, air temperature, dew point, surface pressure, and solar radiation). This integration provides the necessary physical context for the models to learn and compensate for highly non-linear hydroclimatic processes, particularly orographic lifting and wind-induced gauge undercatch.

To ensure the spatiotemporal robustness of the CRISP framework, we implemented a multi-tiered validation strategy. During model development, a standard 5-fold cross-validation was employed to guarantee spatial stability across diverse geographic samples (Fushiki, 2011). Furthermore, recognizing the strong temporal autocorrelation inherent in hydroclimatic data, the final validation was executed using an annual-based hold-out methodology. Specifically, the long-term observational records were temporally partitioned: data spanning the period from 1998 to 2015 were utilized for model training and structural optimization, while the remaining calendar years from 2016 to 2020 were strictly isolated as an independent test set. By isolating entire calendar years for independent testing, this approach ensures an unbiased assessment of the dataset's reliability across varying climatic years.

Following the completion of the validation protocols, the CRISP framework was re-trained utilizing the full integrated long-term observational dataset to generate the final production version. This comprehensive dataset was employed for the 27-year multidimensional data cube and served as the foundation for the subsequent independent station evaluations and physical consistency tests.

3.2 The CRISP machine learning engine

The core estimation engine utilizes an integrated machine learning approach to separately model precipitation occurrence, magnitude, and uncertainty. To ensure optimal performance and prevent regional overfitting, the structural parameters of all algorithms were tuned using Bayesian optimization (Frazier, 2018) prior to final dataset generation.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

3.2.1 Precipitation classifier

Because the TP exhibits a high frequency of clear sky days, estimating precipitation as a pure regression problem typically leads to a severe overestimation (Lyu & Yong, 2024). To resolve this, we conducted a soft voting ensemble classifier combining Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017) and Random Forest (RF) (Breiman, 2001) algorithms. This classifier evaluates the dynamic and static features to estimate the absolute probability of precipitation occurrence (Prob_Rain) for every grid cell. To account for the imbalanced nature of precipitation events, we implemented an unequal error-penalty mechanism during training. By assigning a higher penalty to missed actual rainfall, the classifier maintains high detection sensitivity while effectively reducing the "drizzle effect" commonly seen from reanalysis background fields (see Table 2).

3.2.2 Best estimate regressor

For the deterministic quantification of precipitation magnitude, a stacking ensemble architecture was used. Stacking effectively minimizes the structural biases of individual algorithms by training a meta-learner to optimally combine their predictions. The base layer integrates LightGBM, RF, and a Multi-layer Perceptron (MLP) (Kruse et al., 2022) to capture diverse non-linear relationships. After getting the first guess results from these optimized models, a Ridge regression model serves as the final meta-estimator (McDonald, 2009), applying L2 regularization to fuse the first guess results into a robust and highly accurate estimate (P_E).

3.2.3 Uncertainty quantifier

A primary advancement of the CRISP dataset is the quantification of estimation uncertainty without relying on strict parametric assumptions. Recognizing that the absolute error magnitude for heavy rainfall is larger than that for light drizzle, we integrated an Asymmetric Conformalized Quantile Regression (ACQR) algorithm (Romano et al., 2019). In detail, two independent LightGBM models were trained using a quantile objective function to directly estimate the raw 5th and 95th percentiles of the precipitation distribution. Subsequently, to guarantee the mathematical validity of the uncertainty bounds and prevent potential data leakage, a comprehensive and strictly independent calibration pool was constructed. Rather than arbitrarily sacrificing training data, we utilized the out-of-fold (OOF) predictions generated during the aforementioned 5-fold cross-validation phase. This cross-conformal approach ensures that the asymmetric conformity scores are computed exclusively on instances completely unseen by the respective quantile models. Because applying a single, global error penalty would inappropriately force the lower bounds of light rain events to zero, we calculated distinct calibration scores for the lower and upper bounds. This asymmetric calibration ensures that the generated estimation intervals naturally widen for extreme events and narrow for trace precipitation, providing users with mathematically reliable 90% coverage intervals.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

3.3 Physics-guided post-processing

To ensure that the statistical outputs from the machine learning engine follow the hydroclimatic reality strictly, a three-dimensional physics-guided post-processing module was implemented prior to finalizing the data cube.

3.3.1 Physical value constraints

Because statistical models occasionally produce trace numerical artifacts instead of true zeros, physical value constraints were enforced to reduce false precipitation signals. Grid cells with a predicted rain probability (Prob_Rain) below an optimized threshold (here, 40%) were masked to zero. For wet grid cells, the deterministic estimate was secured by setting a meteorological trace minimum of 0.1 mm/day. Concurrently, boundary consistency algorithms were applied to logically constrain the uncertainty intervals. A strict topological sorting constraint was enforced across all pixels to guarantee that the lower bound never exceeds the best estimate, and the best estimate never exceeds the upper bound ($Prcp\_5th \leq Prcp\_Best \leq Prcp\_95th$), preventing physically impossible overlapping in complex terrain.

3.3.2 Thermodynamic phase discrimination

To support cryospheric research, a dedicated module was incorporated to explicitly classify the physical state of precipitation into four categories: Dry (0), Snow (1), Mixed (2), and Rain (3). In alpine regions, lower atmospheric pressure and reduced humidity intensify the sublimational cooling of falling hydrometeors, allowing solid precipitation to survive at relatively warmer near-surface temperatures. To account for this, the critical 2-m air temperature threshold for snowfall was dynamically parameterized using the SRTM DEM. The threshold shifts linearly from -2.0°C at lower elevations to -1.0°C for regions above 4,000 m, while pure rainfall is defined strictly above 2.0°C. To address ERA5 uncertainties in resolving micro-topographic inversions, this logic is further constrained by the IMERG liquid probability layer: probabilities >90% are forcibly reclassified as rain, and <10% as snow. Crucially, to prevent circular reasoning during evaluation (e.g., Figure 10), our ground-truth phase validation ("Obs") relies on completely independent data. The benchmarks are extracted exclusively from the official Present Weather codes recorded by the NMIC network, which are determined via standardized manual observations, rather than temperature-based derivations.

3.3.3 Climatological calibration

While the baseline models optimize daily event-level precision, long-term hydroclimatic applications require strict volumetric consistency. To eliminate slight magnitude biases over multi-decadal scales, a climatological calibration module was developed using non-parametric Empirical Cumulative Distribution Function (CDF) mapping (Popat & Hartmann, 2024). A baseline CDF was constructed by extracting exact quantiles from both the preliminary estimations and the corresponding gauge observations. A robust transfer function was then established to map the estimated quantiles directly to the observed

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

distribution. The resulting climatology-corrected layer (Prcp_Clim) perfectly preserves the daily spatial heterogeneity generated by our models while restoring the long-term regional climatological volume.

Table 2. Results of hyperparameter tuning.

Model	Task	Hyperparameter	Range	Optimal value
LightGBM	Classifier	n_estimators	[200, 2000]	500
		class_weight	-	0: 1.0, 1: 1.5
	Regressor	n_estimators	[200, 2000]	1575
		learning_rate	[0.005, 0.05]	0.016
		num_leaves	[31, 150]	70
		max_depth	[7, 15]	7
		subsample	[0.6, 0.9]	0.652
		colsample_bytree	[0.6, 0.9]	0.65
		reg_alpha	[0.001, 1]	0.005
		reg_lambda	[0.001, 1]	0.003
RF	Classifier	n_estimators	[150, 400]	200
		max_depth	[12, 25]	15
		class_weight	-	0: 1.0, 1: 1.5
	Regressor	n_estimators	[150, 400]	332
		max_depth	[12, 25]	23
		min_samples_split	[5, 15]	8
		min_samples_leaf	[4, 12]	10
MLP	Regressor	hidden_layer_sizes	-	100, 50, 25
		max_iter	-	1000
		alpha	[0.0001, 0.01]	0.002
		learning_rate_init	[0.0005, 0.005]	0.001

3.4 Implementation details and hyperparameter optimization

To ensure computational reproducibility, the CRISP framework was implemented in the Python programming environment, using the scikit-learn and LightGBM libraries. Given the massive spatiotemporal data size across the study period, hyperparameter optimization was executed on a representative subset of the training data. We employed the Optuna framework to conduct Bayesian optimization, systematically navigating the high-dimensional hyperparameter space to minimize the Root Mean Square Error (RMSE) for the regression modules and optimize the performance metrics.

For the regressor, the Bayesian optimization converged on an ensemble architecture that effectively balances structural complexity with robust regularization. Specifically, the tree-based algorithms (LightGBM and RF) favored a large ensemble size of base estimators, constrained by conservative learning rates and stochastic subsampling mechanisms. Concurrently, the MLP component converged on a three-tier hierarchical architecture. This progressive dimensionality reduction allows our

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

dataset to effectively get thermodynamic information from the raw forcing data without suffering from parameter expansion. The finalized optimal hyperparameters for all base learners are systematically summarized in Table 2.

Similarly, the precipitation classifier inherited these robust baseline configurations but was tailored to address a meteorological challenge (i.e., the high frequency of dry days compared to actual precipitation events). To prevent the model from passively defaulting to dry days, we introduced a targeted asymmetric penalty mechanism. Rather than treating dry and wet conditions equally, the algorithm (class_weight) was explicitly programmed to penalize missed rainfall events 1.5 times more heavily than false alarms. This strategy ensures that the model maintains a high sensitivity to actual precipitation while effectively reducing drizzle false alarms, perfectly aligning with the complex hydroclimatic characteristics of the region.

3.5 Data format and structure

To facilitate seamless integration with Earth system models and climate data operators (e.g., CDO, xarray), the CRISP dataset is encapsulated in the Network Common Data Form version 4 (NetCDF4) format, strictly adhering to the Climate and Forecast (CF) metadata conventions (version 1.8). To optimize I/O performance and storage efficiency, the multidimensional data cubes employ spatial chunking alongside internal zlib compression.

The dataset is distributed as annual files (e.g., CRISP_1998.nc). The spatial framework consists of a 400 ×200 grid cell array mapped to the standard World Geodetic System 1984 (WGS84; EPSG:4326) Coordinate Reference System (CRS). To prevent spatial ambiguity, longitude and latitude arrays define the grid cell centers (e.g., 65.05° to 104.95° E) and are structurally supported by explicit coordinate bounds. The temporal dimension (time) serves as a continuous axis across all separate files. It is recorded as 64-bit integers representing "days since 1998-01-01 00:00:00" for every file, utilizing the proleptic_gregorian calendar to automatically resolve leap days without breaking the temporal concatenation.

Within this consistent coordinate framework, the data cube explicitly houses six primary physical variables to support diverse hydroclimatic applications. All continuous variables are stored as single-precision floating-point numbers (single / float32) to optimize computational memory without reducing precision. These include:

1. Prcp_Best: The optimal deterministic daily precipitation estimates.

2. Prcp_Clim: The climatologically calibrated precipitation for long-term volumetric consistency.

3. Prcp_5th and Prcp_95th: The explicitly quantified asymmetric uncertainty bounds (5th and 95th percentiles).

4. Prob_Rain: The absolute predicted probability of precipitation occurrence.

To maximize storage efficiency, the discrete thermodynamic phase classification (**Phase**) is stored as an 8-bit integer (int8). Following CF conventions for categorical data, it is explicitly defined by flag_values (0, 1, 2, 3) and their corresponding flag_meanings ("dry snow mixed rain"). Finally, a static regional land mask is applied: oceanic regions and grid cells extending beyond the valid terrestrial study domain are uniformly masked, with both the standard _FillValue and missing_value attributes explicitly set to NaN.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

330 **4 Results and discussions**

4.1 Overall spatiotemporal patterns and statistical performance

To comprehensively evaluate the developed dataset, we first examined its ability to capture the spatiotemporal patterns and its overall deterministic accuracy against reference products.

As shown in the spatial distribution of mean annual precipitation (Figure 4a-f), CRISP accurately captures the southeast-to-
335 northwest moisture gradient across the TP. Compared to the satellite retrievals (IMERG and GSMaP), CRISP shows enhanced topographic performance, especially characterizing precipitation features along the Himalayas and the Nyainqêntanglha Mountains. Furthermore, while the ERA5 reanalysis effectively captures the spatial dynamics of the precipitation, it exhibits a systematic wet bias across the southeastern plateau. CRISP successfully improves the tendency of overestimation by utilizing satellite constraints with our algorithms, resulting in a high-resolution spatial pattern that closely aligns with the dense gauge
340 network (Figure 4f).

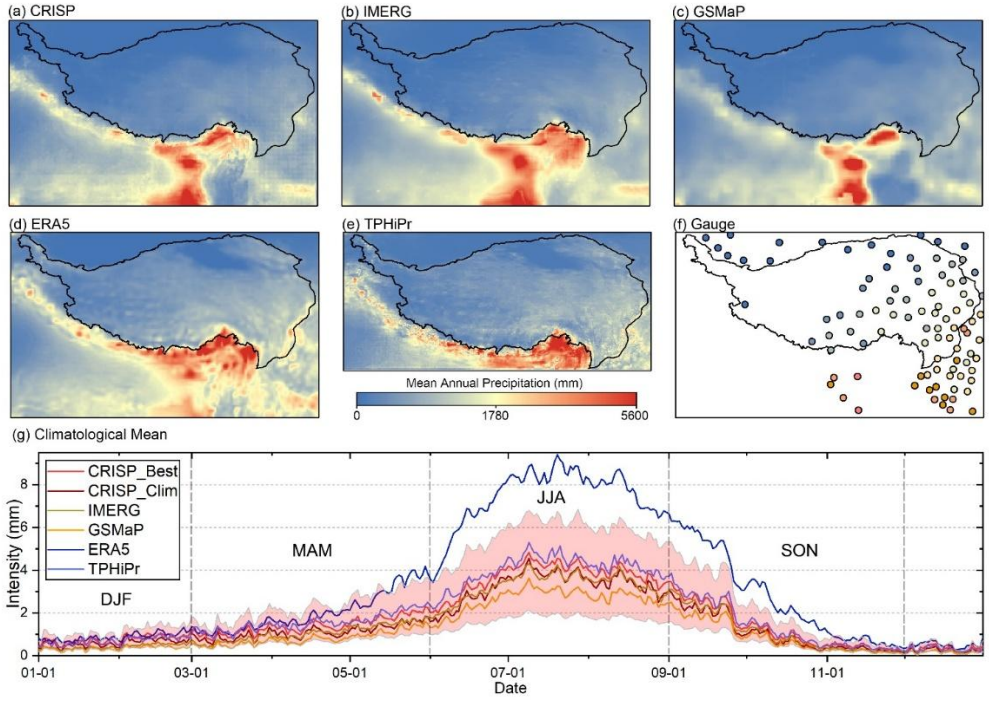


Figure 4. Spatiotemporal distribution of mean annual precipitation and daily climatological time series across the Tibetan Plateau (1998-2024). (a-f) Spatial patterns of multi-year mean annual precipitation derived from CRISP, IMERG, GSMaP, ERA5, TPHiPr (1998-2020), and NMIC Gauge, respectively. **(g)** Climatological mean daily precipitation time series
345 covering the entire annual cycle. Solid lines represent the deterministic best estimates from the evaluated products. The red shaded area indicates the estimation interval (5th to 95th percentiles) dynamically.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

The temporal robustness of the dataset is proved through the climatological mean daily precipitation time series (Figure 4g). The precipitation over the TP is heavily affected by the Asian summer monsoon, with the majority of rainfall concentrated in the JJA (i.e., June-July-August) season. ERA5 generally overestimates the daily intensity throughout the entire annual cycle. In contrast, the deterministic estimate of CRISP (i.e., Prcp_Best, hereafter CRISP_Best) and its climatologically calibrated variable (i.e., Prcp_Clim, hereafter CRISP_Clim) show a reasonable intensity, tracking the multi-product closely. Crucially, Figure 4g also visualizes the 90% estimation intervals generated by the ACQR module (shaded red area). These uncertainty ranges successfully include the deterministic estimates across all seasons, widening during the intense monsoon summer (reflecting higher variability) and narrowing during the drier winter months.

In addition to visual consistency, quantitative statistical evaluations also validate the accuracy of the CRISP product. As shown in the overall metric comparisons (Figure 5), CRISP outperforms IMERG, GSMaP, ERA5, and the regional benchmark TPHiPr across primary accuracy indicators in the validation periods. CRISP achieves the highest Correlation Coefficient (CC= 0.77) and the lowest Root Mean Square Error (RMSE < 8.5 mm/day), showing high agreement with ground observations (Figure 5a, 5c; The relevant evaluation metrics can be found in the Appendix A & Table A1.)

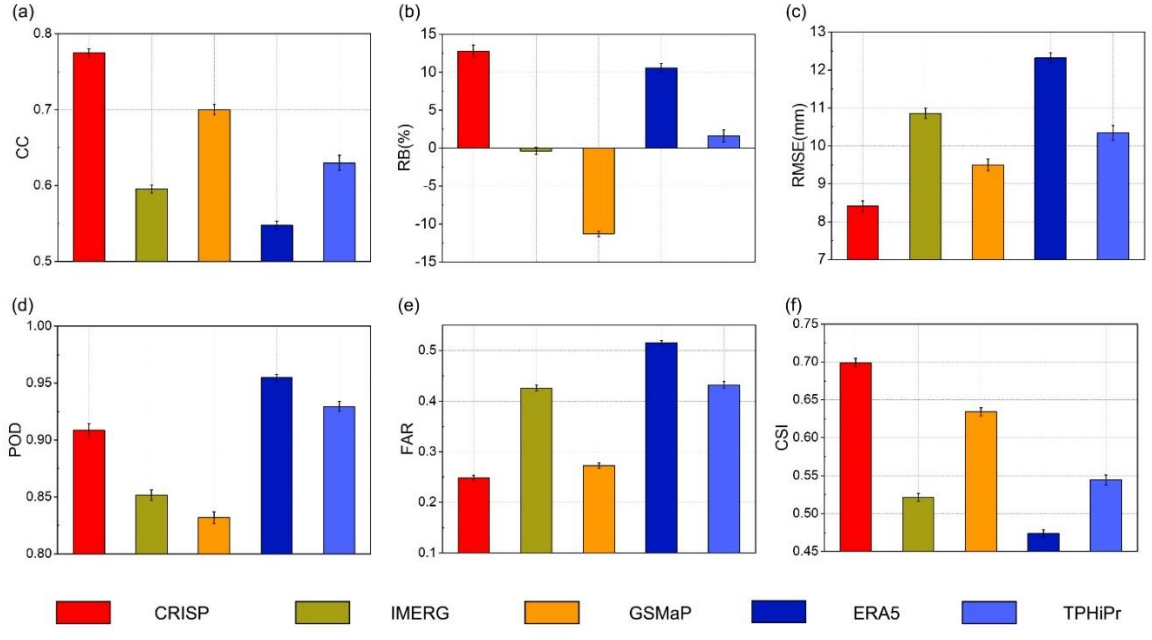


Figure 5. Quantitative statistical evaluation of daily precipitation products during the validation period (2016-2020). The performance metrics evaluate both continuous statistical accuracy and precipitation detection skills, including: **(a)** Correlation Coefficient (CC), **(b)** Relative Bias (RB, %), **(c)** Root Mean Square Error (RMSE, mm/day), **(d)** Probability of Detection (POD), **(e)** False Alarm Ratio (FAR), and **(f)** Critical Success Index (CSI). The error bar is obtained through 10,000 bootstrap resampling processes of gridded daily datasets.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Notably, the evaluation of precipitation detection metrics validates the algorithmic design of the CRISP precipitation classifier. Generally, reanalysis products like ERA5 suffer from the "drizzle effect", as evidenced by its near-perfect Probability of Detection (POD > 0.95) but high False Alarm Ratio (FAR > 0.5) (Figure 5d, 5e). Conversely, owing to the asymmetric class-weighting penalty applied in the classification module, CRISP achieves a high POD (0.91) while maintaining the lowest FAR (< 0.25) among all evaluated products. This optimal balance achieves the highest Critical Success Index (CSI = 0.70; Figure 5f), proving that CRISP effectively identifies actual precipitation events from the noise.

4.2 Climatological performance

While the deterministic output (CRISP_Best) can minimize daily event-level residuals and capturing spatial patterns (as demonstrated in Section 4.1), data-driven regression models inherently prioritize variance reduction. To ensure that the dataset is suitable for long-term climate trend analysis and hydrological modeling, we evaluated the performance of the climatological calibration results (CRISP_Clim, 1998-2024).

As shown in the precipitation frequency distributions (Figure 6a), traditional reanalysis products like ERA5 and regional downscaled datasets like TPHiPr show a severe overestimation of trace and light precipitation (< 2 mm/day). This "drizzle effect" changes their entire rainfall climatological distribution. CRISP_Best successfully reduces this high false alarm at the lowest intensities, aligning much closer to the ground observations. However, CRISP_Best demonstrates a slight overestimation in the moderate-to-heavy precipitation bins (i.e., 2 to 15 mm/day) compared to the calibrated version.

This phenomenon is further quantified in the frequency bias analysis (Figure 6b). The uncalibrated CRISP_Best curve exhibits a positive bias peak in the moderate intensity range. Following the application of the CDF mapping module, CRISP_Clim reduces this structural distortion clearly. The frequency bias of CRISP_Clim is almost reduced, pulling the curve close to the zero-bias baseline across the entire intensity range.

The total results of this calibration are validated by the CDF curves (Figure 6c). Due to the known drizzle effect, the cumulative probability for ERA5 increases far too quickly compared to the actual ground truth. Conversely, the CDF curve of CRISP_Clim perfectly aligns the Observation curve. This precise distributional alignment guarantees that the CRISP framework preserves the daily high accuracy generated by the regressor process while keeping the long-term regional climatological persistence.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

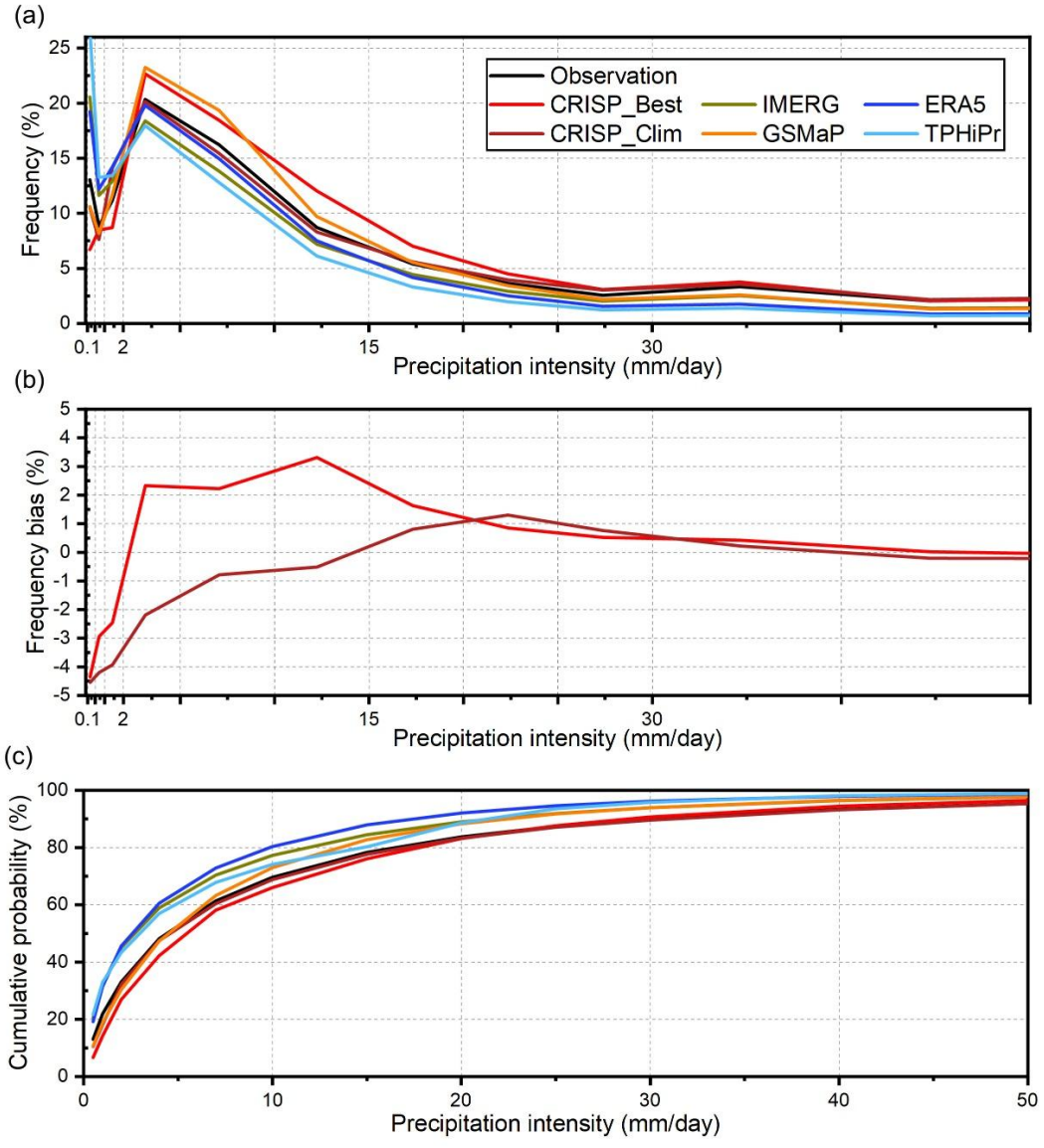


Figure 6. Climatological distribution and calibration performance of daily precipitation intensities. (a) Probability density functions (PDF) displaying the occurrence frequency across different precipitation intensity bins. **(b)** Frequency bias of the uncalibrated preliminary estimate (CRISP_Best) and the climatologically calibrated output (CRISP_Clim) relative to the gauge baseline. **(c)** Empirical Cumulative Distribution Functions (CDF) of the evaluated datasets.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

4.3 Performance across topographic gradients and seasonal cycles

The orographic relief and seasonal variability of the TP pose the most well-known challenges for precipitation estimation. To evaluate the robustness of the physics-aware corrections integrated for the CRISP dataset, we further calculated the validation metrics across different elevation bins and seasonal cycles during the validation period (2016-2020).

The performance gradients along varying elevations (Figure 7) reveal similarity problems in reference datasets. As elevation exceeds 3,000 m, the precipitation detection skills of satellite retrievals (IMERG and GSMaP) decrease rapidly, characterized by a sharp drop in POD (Figure 7d) and CSI (Figure 7f). This phenomenon occurs because spaceborne infrared and passive microwave sensors struggle to distinguish cold, high-altitude precipitation surfaces from underlying snow cover. Conversely, the ERA5 reanalysis captures the occurrence well but suffers from an exponentially increasing wet bias (Figure 7b), overestimating precipitation intensities at elevations above 4,000 m due to unresolved sub-grid orographic lifting.

In contrast, CRISP dataset maintains high stability across the entire topographic profile. By incorporating static geomorphometric factors (elevation, slope, terrain roughness) and dynamic atmospheric forcing (wind speed, temperature), the CRISP learns the localized characteristics of orographic precipitation. Consequently, CRISP dataset maintains the highest CC (CC > 0.8, Figure 7a) and consistently low RMSE (Figure 7c) even in the most extreme elevation bin (>4,000 m), proving its unparalleled capability to correct topography-induced measurement biases in complex alpine environments.

This robust physical adaptability is further supported by the seasonal Taylor diagrams (Figure 8), which integrating correlation, centered root-mean-square difference, and standard deviation into a single multidimensional assessment. During the summer monsoon season (JJA, Figure 8b), characterized by abundant liquid precipitation, all products exhibit their best relative performance, but CRISP still performs closest to the observation reference point.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

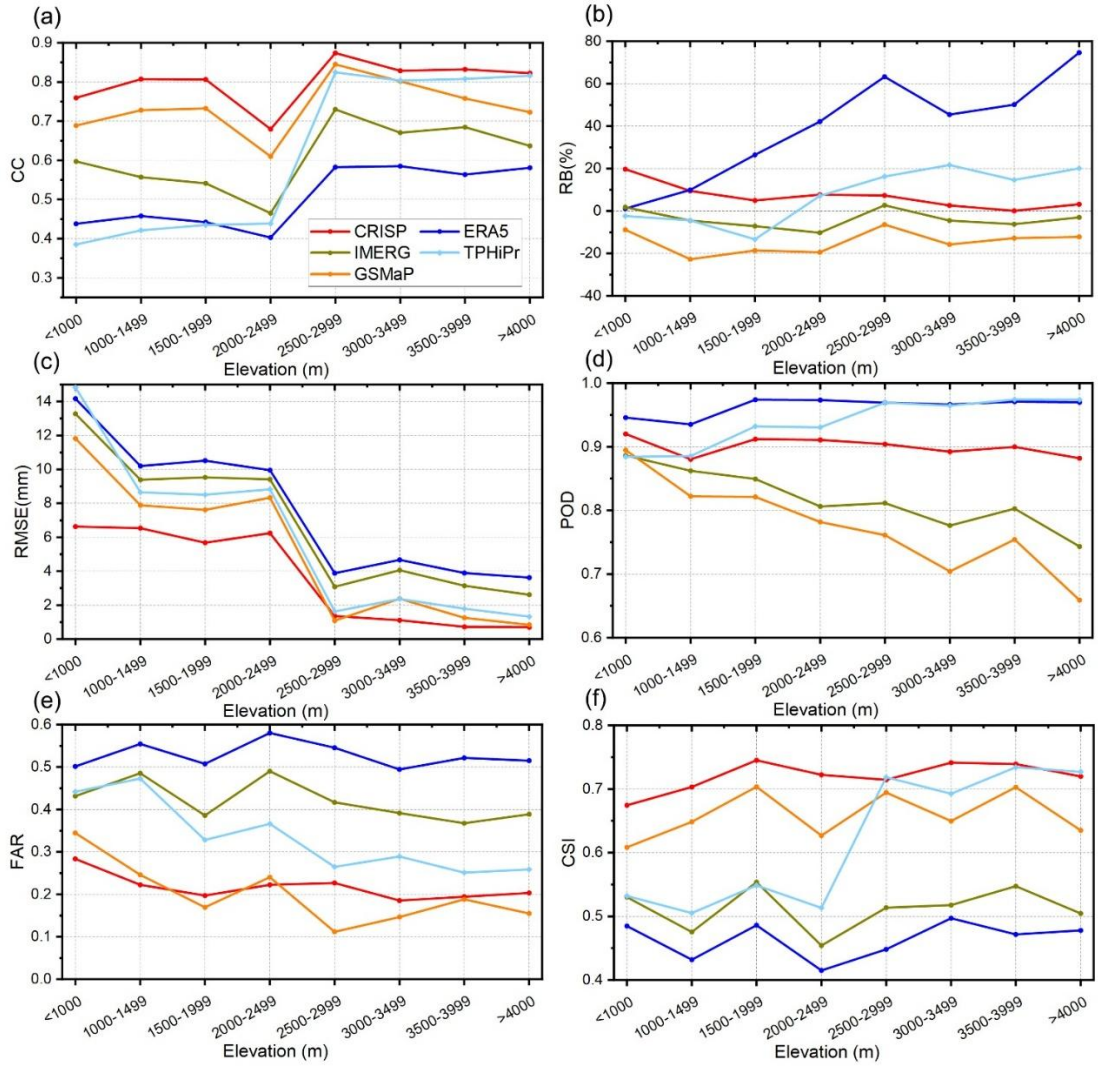


Figure 7. Stepwise performance evaluation of daily precipitation across varying elevation gradients during the validation period (2016-2020). The six panels display the progression of (a) Correlation Coefficient (CC), (b) Relative Bias (RB, %), (c) Root Mean Square Error (RMSE, mm/day), (d) Probability of Detection (POD), (e) False Alarm Ratio (FAR), and (f) Critical Success Index (CSI) across different elevation bins.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

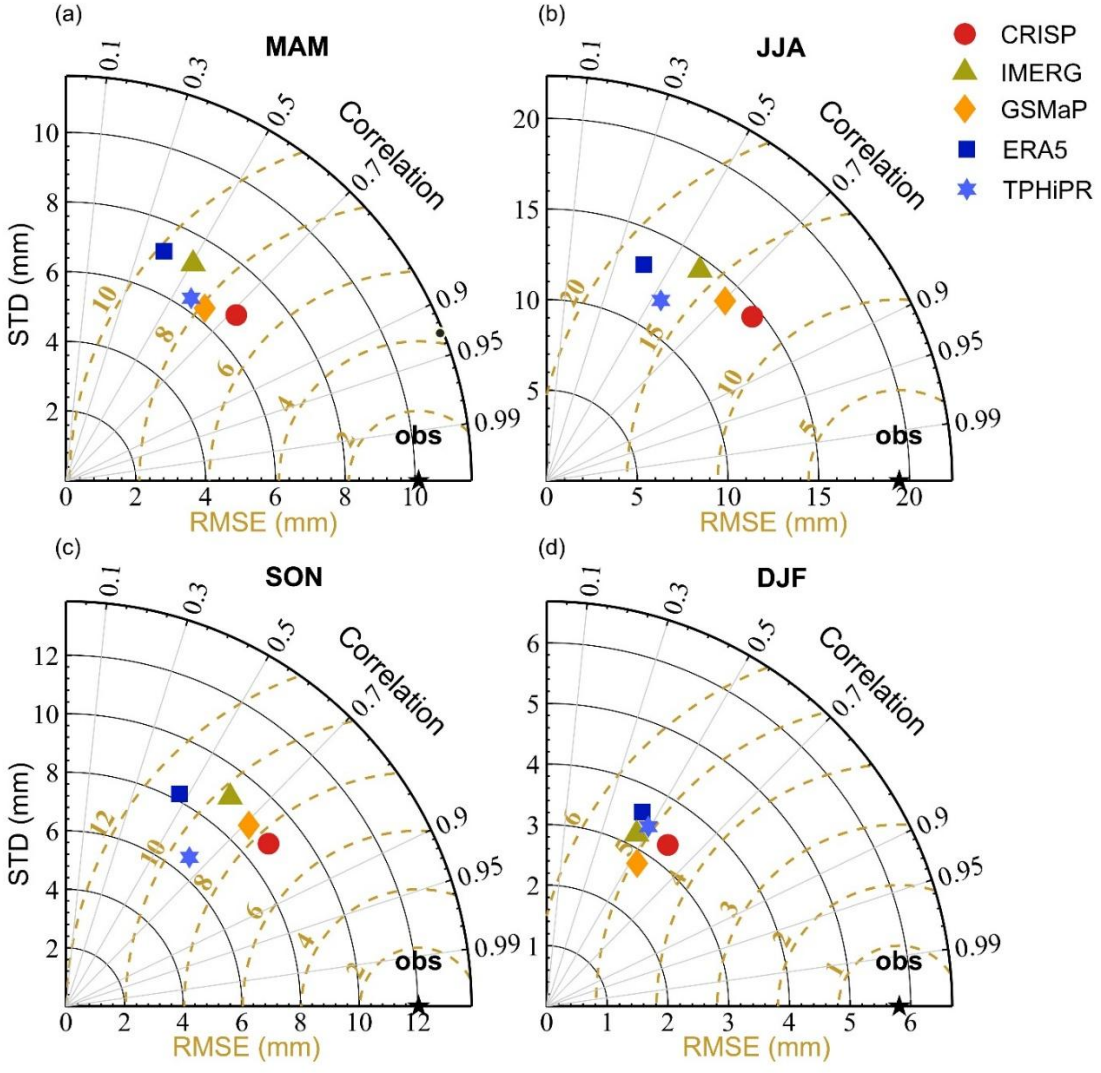


Figure 8. Seasonal multidimensional performance evaluated using Taylor diagrams during the validation period (2016-2020). The diagrams summarize the statistical agreement between estimated datasets and ground observations across four seasons: **(a)** MAM (Spring), **(b)** JJA (Summer), **(c)** SON (Autumn), and **(d)** DJF (Winter). The distance from the origin represents the standard deviation, the azimuthal angle denotes the Pearson correlation coefficient, and the distance from the "obs" reference point along the x-axis signifies the centered root-mean-square error.

The most critical differentiation happens during the transition and winter seasons (SON, MAM, and especially DJF). In the DJF (Winter) period (Figure 8d), precipitation mainly occurs as trace snowfall under strong background winds, conditions that difficult for traditional estimation datasets. However, CRISP dataset maintains a high performance near the observation benchmark. This directly validates the efficiency of the CRISP dataset, which preserve the efficiency of solid precipitation estimates while filtering out winter clear-sky noise.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Open Access Earth System Science Data Discussions

4.4 Uncertainty bounds and phase discrimination

A primary limitation of deterministic precipitation products is their inability to communicate the inherent unreliability of their estimates. For instance, the year 2022, dominated by a rare "triple-dip" La Niña, exhibited extreme spatial heterogeneity, introduced profound atmospheric non-stationarity that severely challenges traditional single-value estimations. The CRISP dataset resolves this by coupling deterministic values with rigorous, dynamically calibrated estimation intervals, ensuring reliable error quantification even under such extreme contemporary climate anomalies.

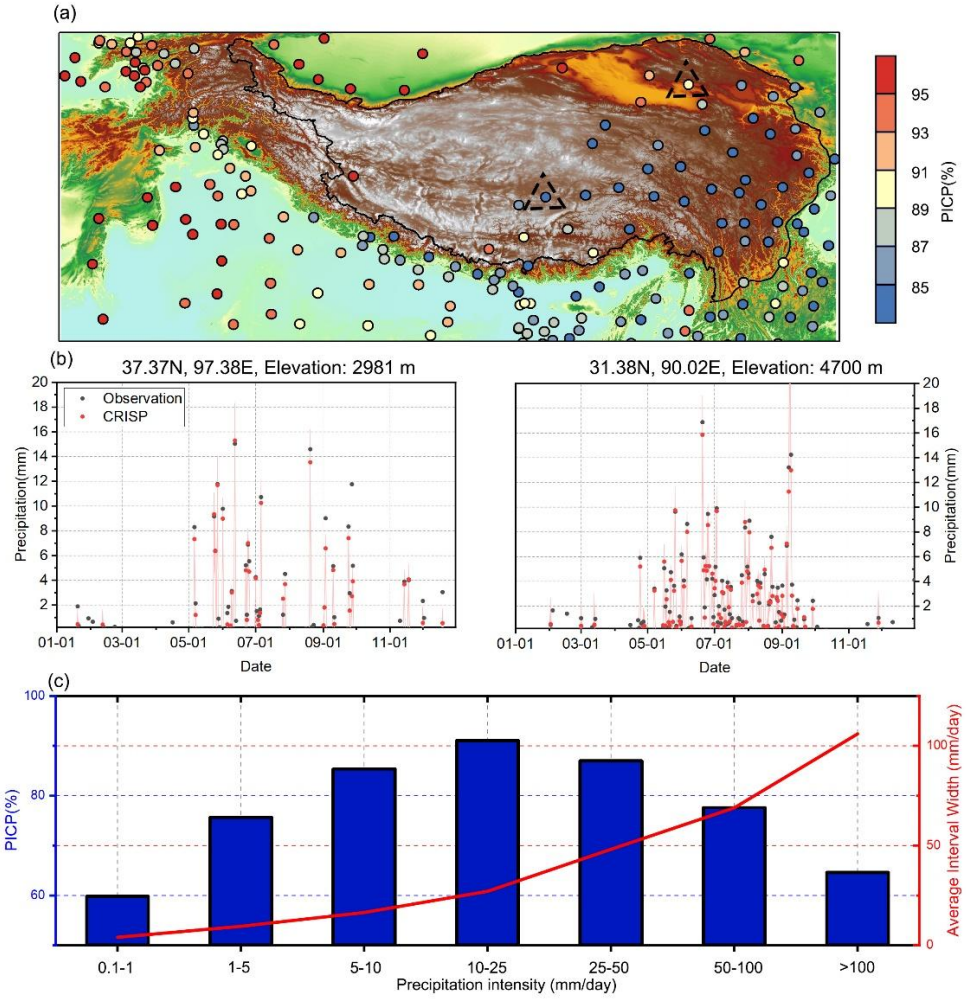


Figure 9. Evaluation of the explicit uncertainty bounds generated by the ACQR module during the validation period (2016-2020). (a) Spatial distribution of the Prediction Interval Coverage Probability (PICP) across the NMIC gauge network. **(b)** Daily precipitation time series at two representative high-altitude stations (2,981 m and 4,700 m) in 2022. Red vertical lines denote the dynamic 5th-95th percentile prediction intervals framing the best estimates. **(c)** PICP (blue bars) and average interval width (red line) stratified by precipitation intensity.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

The spatial distribution of the Prediction Interval Coverage Probability (PICP) (Figure 9a) demonstrates the good reliability of the ACQR module. Across the most observed region of the TP, the generated 90% prediction intervals successfully encompass the true ground observations with an over 85% PICP rate, ensuring spatial robustness across diverse topographic and climatic regions. This event-level reliability is confirmed at two representative high-altitude gauges (Figure 9b). The dynamic estimation intervals (red vertical lines) seamlessly envelope the observations, providing guaranteed error margins that are essential for downstream hydrological risk assessments.

Crucially, the CRISP uncertainty ranges strictly follow the physical nature of precipitation intensity distribution. As quantified in Figure 9c, the average interval width (red line) increases along with precipitation intensity. For low intensity rainfall, the model outputs narrow, highly confident bounds while during extreme heavy rainfall events (>50 mm/day), the intervals automatically widen to reflect the heightened atmospheric chaos and greater potential error. Concurrently, the PICP maintains high coverage (>80%) for moderate-to-heavy events, proving that the ACQR successfully calibrated the boundaries without failure under extreme values.

Furthermore, to directly support cryospheric and snowmelt modeling, CRISP explicitly decomposes the total precipitation volume into distinct physical states (rain, snow, and mixed) through its elevation-dependent thermodynamic phase module. Assessing the categorical accuracy of phase classification (Figure 10a–c) reveals that CRISP has advantages in identifying liquid precipitation (rain), achieving median POD and CSI scores near 0.7. Moreover, the model demonstrates robust capability in capturing solid precipitation (snow), maintaining a median POD above 0.5, which is a notable achievement given the difficulty of detecting snowfall under severe wind conditions on the TP.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

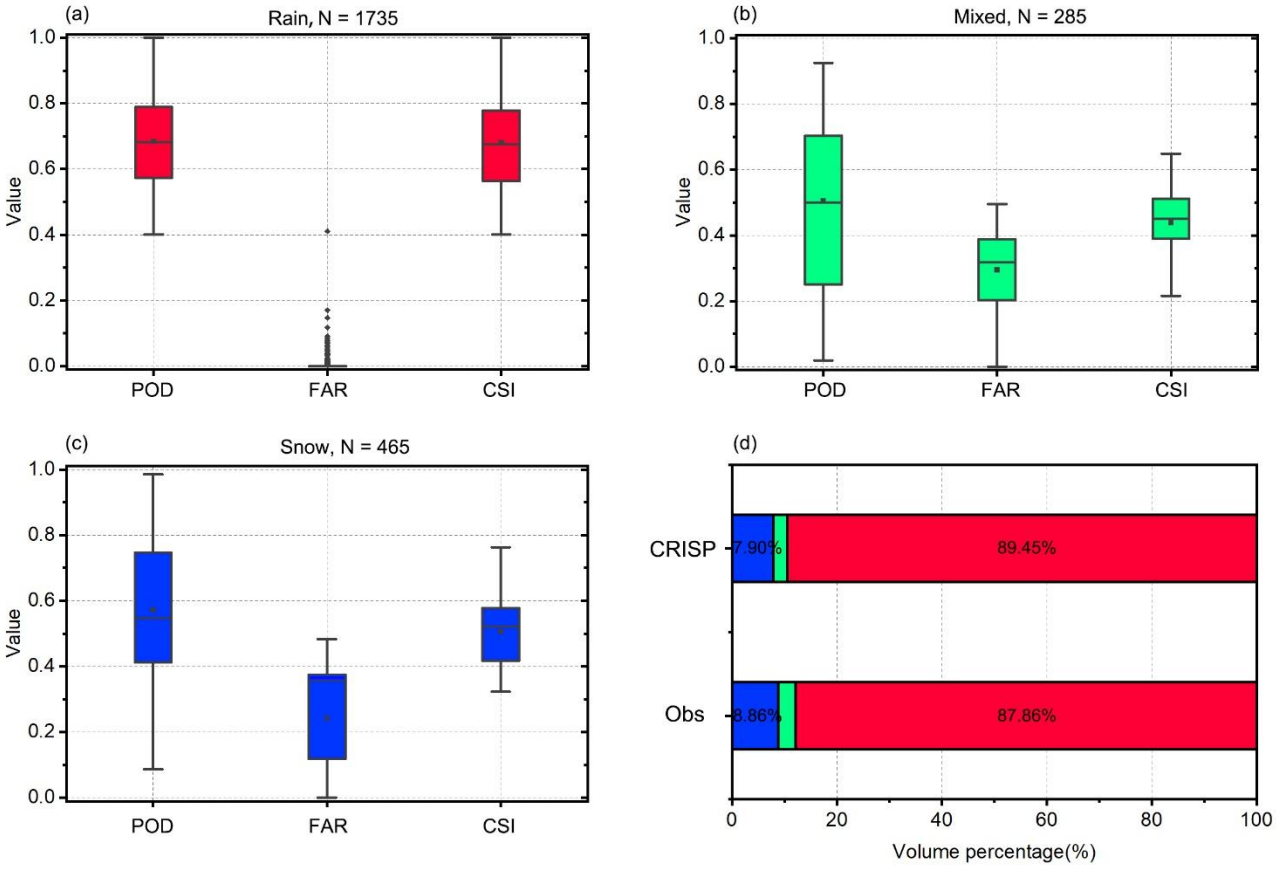


Figure 10. Performance evaluation of the thermodynamic phase discrimination module during the validation period (2016-2020). Boxplots illustrate the categorical metrics (Probability of Detection [POD], False Alarm Ratio [FAR], and Critical Success Index [CSI]) specifically evaluated for events classified as **(a)** Rain, **(b)** Mixed, and **(c)** Snow. **(d)** Volumetric percentage distribution of the three physical states.

In addition to precipitation detection skills, the regional hydro-climatological value of this discrimination lies in volumetric closure. As shown in the volume percentage analysis (Figure 10d), the CRISP framework accurately partitions the total precipitation mass. The classified proportions of Rain (89.45%), Snow (7.90%), and Mixed (2.65%) states almost perfectly align with the observed volumetric distribution (87.86% Rain, 8.86% Snow). This precise volumetric alignment ensures that researchers utilizing CRISP for glacier mass balance or snow-driven runoff modeling are fed with physically consistent, unbiased solid precipitation inputs.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

4.5 Independent generalization and water balance test

A risk of machine-learning based datasets in precipitation estimation is their susceptibility to overfitting. To validate the actual capabilities of the CRISP dataset, we conducted a generalization test using all available independent gauge samples that were not used during the training process due to the short period.

The evaluation metrics derived from these independent gauges (Figure 11) confirm the robust generalization of the dataset. While the overall performance shows a slightly wider spread compared to the cross-validation, CRISP consistently maintains the highest median CC value and the lowest median RMSE among all evaluated datasets. Most importantly, the precipitation detecting skill of the precipitation classifier remains highly stable in these unseen regions. CRISP achieves the highest median CSI by sustaining a high POD and low FAR values, generally avoiding the severe false alarm trouble observed in the ERA5 and TPHiPr products. This proves that the dataset successfully presented high applicability for ungauged regions.

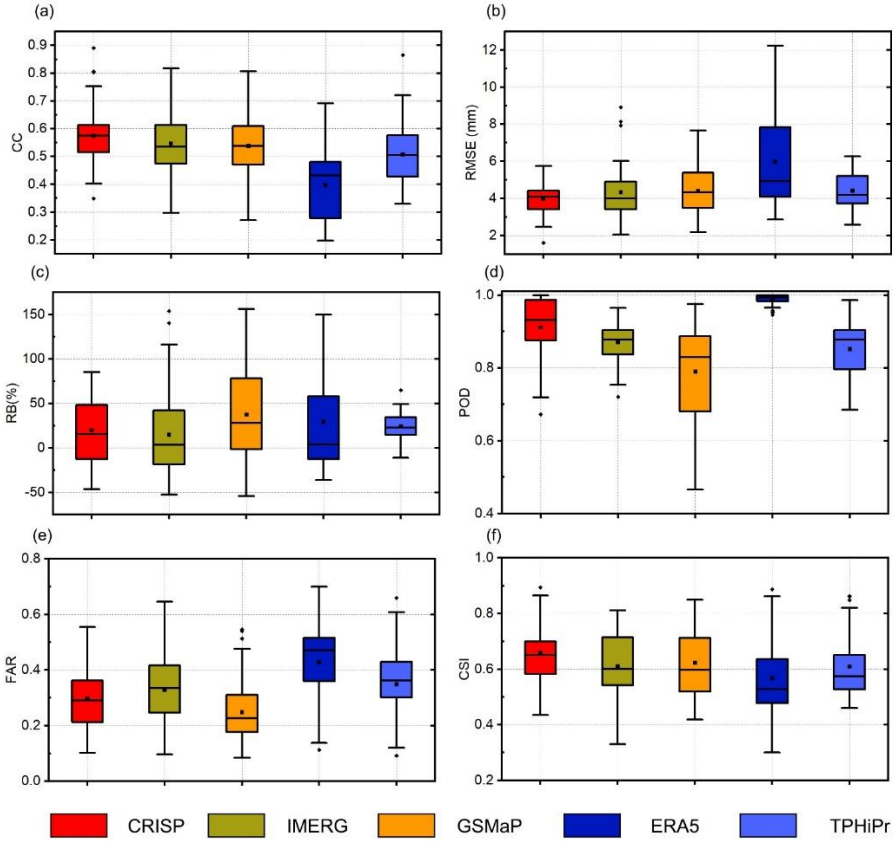
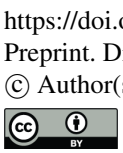


Figure 11. Generalization performance evaluated using all available independent gauge samples. Boxplots display the distribution of continuous and categorical metrics: **(a)** CC, **(b)** RMSE, **(c)** RB, **(d)** POD, **(e)** FAR, and **(f)** CSI. The central horizontal line indicates the median, the box spans the interquartile range (25th to 75th percentiles), and the whiskers represent the 5th and 95th percentiles.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026



Open Access Earth System Science Data Discussions

To further assess physical consistency, we evaluated the dataset against the "Third Pole Precipitation Paradox" (Miao et al., 2024). Standard gridded products often violate thermodynamic limits by overestimating precipitation beyond regional water and energy boundaries, while raw ground gauges typically underestimate solid precipitation due to wind-induced undercatch. Given the limited availability of reliable basin-scale runoff data, we utilized gauge-specific Potential Evapotranspiration (PET) and Snow Water Equivalent (SWE) as physical constraints to examine gridded water balance closure (Figure 12).

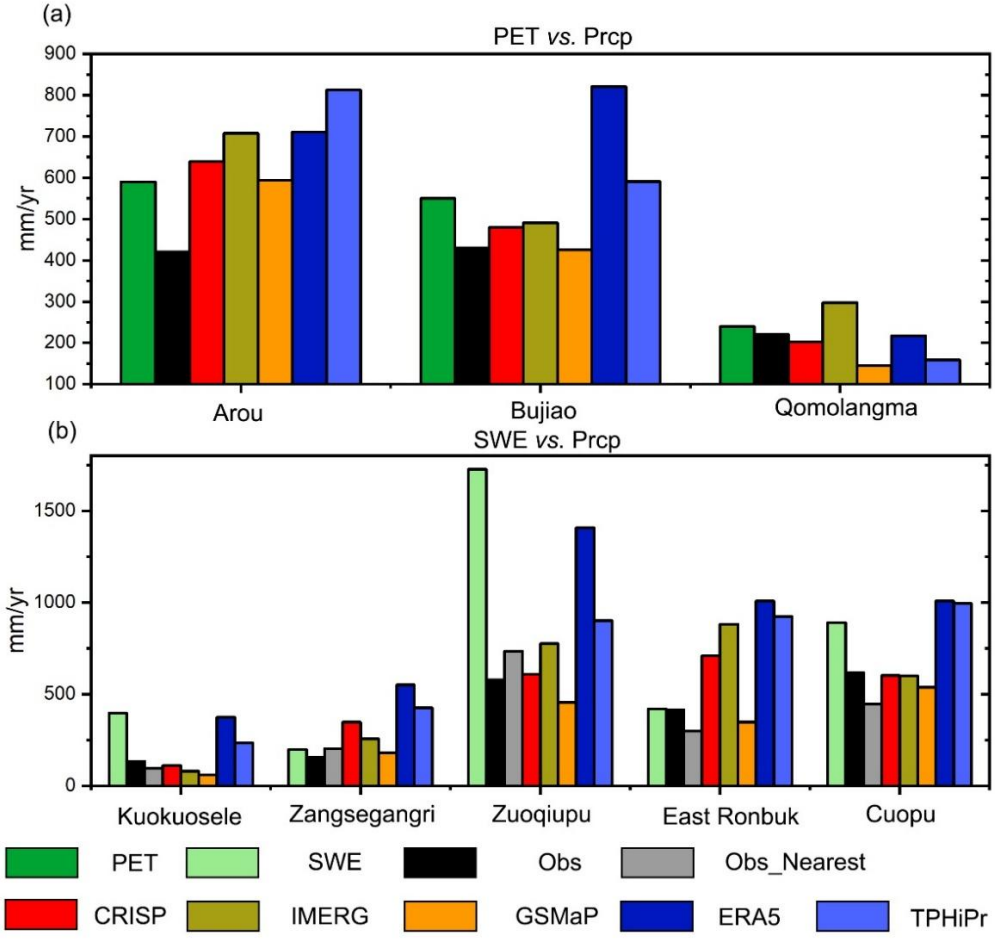


Figure 12. Validation of gauge-based water balance closure. Empirical data are derived from the Third Pole precipitation paradox study (Miao et al., 2024). **(a)** Comparison of estimated precipitation against Potential Evapotranspiration (PET) and gauge observations at alpine stations. **(b)** Evaluation of solid precipitation undercatch and mass conservation using Snow Water Equivalent (SWE) as a physical lower bound.

In alpine environments (Figure 12a), the precipitation paradox is primarily evidenced by local atmospheric demand (PET) substantially exceeding raw ground observations at stations such as Arou and Bujiao. Climatologically, long-term precipitation must sufficiently sustain regional evapotranspiration. Thus, this persistent deficit (Obs < PET) reveals the severe measurement undercatch inherent in conventional estimation. CRISP successfully addresses this paradox. By incorporating atmospheric and

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

topographic factors to implicitly compensate for wind-induced undercatch, CRISP systematically elevates precipitation estimates above the raw observational baseline, offering estimates that more realistically align with the PET constraints. Consequently, CRISP can provide a physically consistent moisture input which is beneficial to the observational underestimation.

The physical performance of the datasets is further examined in glacierized zones utilizing the SWE constraint (Figure 12b). From a mass conservation perspective, annual accumulated precipitation must objectively equal or exceed localized SWE. However, raw observations frequently report precipitation significantly lower than the ground-measured SWE (e.g., at Kuokuosele and Zuoqiupu), providing solid evidence of severe aerodynamic undercatch in solid precipitation measurements. Satellite algorithms (IMERG, GSMaP) largely perform this negative bias, consistently failing to satisfy the SWE baseline. Conversely, the ERA5 reanalysis generally produces substantially higher estimates, sometimes approximating extreme SWE values (e.g., at Zuoqiupu) but frequently overshooting into physically inconsistent volumes elsewhere.

CRISP is evidenced to improve the situation of underestimation and has potential to achieve a realistic mass balance. Concurrently, it avoids the excessive volumetric overestimations typical of reanalysis products. By integrating wind speed, temperature, and terrain roughness, the dataset was adjusted to compensate for the wind-induced loss of solid precipitation.

5 Code and data availability

The CRISP precipitation dataset in NetCDF format is available at the National Tibetan Plateau Data Center, which can be accessed at https://doi.org/10.11888/Atmos.tpdc.303469 (Yong & Lyu, 2026). The codes used for producing this dataset are available from https://doi.org/10.5281/zenodo.19567996. The input data can be found in https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGDF_07；https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5；http://sharaku.eorc.jaxa.jp/GSMaP/index.htm ； https://www.earthdata.nasa.gov/data/instruments/srtm. The gauge data is available at https://data.cma.cn/ai/#/detail?id=12.

6 Conclusion

Accurate quantification of precipitation across the Tibetan Plateau remains challenging due to severe orographic effects, wind-induced gauge undercatch, and the limitations of satellite and reanalysis algorithms. To address this, this study introduces the Conformal Regression and Integrated Stacking for Precipitation (CRISP) dataset, offering a continuous, high-resolution (0.1°, daily) multidimensional precipitation dataset covering 27 years (1998–2024) over the Tibetan Plateau.

Rather than relying on purely data-driven statistical interpolation, the CRISP incorporates physics-aware machine learning correction methods. By integrating geomorphometric factors and dynamic atmospheric forcing, the ensemble model captures the aerodynamic relationships to compensate for localized orographic lifting and wind-induced undercatch. Consequently, CRISP dataset maintains stable performance across extreme topographic gradients. It shows reliable performance in high-altitude environments (>4,000 m) and during the winter snowfall season. Furthermore, the application of an asymmetric class-

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

weighting mechanism effectively reduces the "drizzle effect" phenomenon, achieving a good balance between precipitation detection and false-alarm rejection.

Besides high accuracy deterministic estimates, a key scientific contribution of the CRISP dataset is its multidimensional structure. Compared to traditional single-value precipitation mapping, the CRISP dataset can offer reliable 90% estimation intervals using an Asymmetric Conformalized Quantile Regression module. These dynamically calibrated bounds reflect the intensity-dependent variations of precipitation, providing users with explicit uncertainty quantification. Concurrently, an elevation-dependent module separates the total precipitation into rain, snow, and mixed phases, offering solid precipitation information that closely match observed ground proportions.

Finally, CRISP dataset bridges the gap between daily meteorological estimates and basin hydrological closure. Through climatological calibration by CDF mapping, the dataset aligns with the regional climatological baseline. When evaluated against the physical constraints of the Third Pole precipitation paradox, CRISP adjusts estimates above uncorrected gauge measurements to correspond to the observed other variables (PET and SWE), while avoiding the overestimations frequently observed in reanalysis models. By offering these key variables, CRISP dataset provides a physically consistent and uncertainty-quantified observational foundation for future cryospheric and hydrological research.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Open Access Earth System Science Data Discussions

Appendix A: Statistical and Categorical Evaluation Metrics

To comprehensively and objectively quantify the performance of the estimated precipitation datasets against ground observations, a suite of deterministic, categorical, and probabilistic metrics was employed in this study.

For continuous volumetric evaluation, we utilized the Correlation Coefficient (CC), Relative Bias (RB), and Root Mean Square Error (RMSE). To evaluate the categorical skill of the models in correctly identifying precipitation events, the Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) were computed based on a standard precipitation threshold of 0.1 mm/day. The mathematical formulations, variable definitions, and optimal values for these primary metrics are systematically summarized in Table A1.

Table A1 Statistical metrics used for evaluating precipitation estimation in this study.

Statistic metrics	Unit	Equation	Description	Optimal Value
CC	NA	$CC = \frac{\sum_{i=1}^{n}(G_i - \bar{G})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2} \times \sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}}$	Measures the linear correlation and phase agreement between the estimated and observed spatial/temporal patterns.	1
RMSE	mm/day	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)^2}$	Evaluates the absolute magnitude of the estimation error, heavily penalizing extreme outliers.	0
RB	%	$RB = \frac{\sum_{i=1}^{n}(S_i - G_i)}{\sum_{i=1}^{n} G_i} \times 100\%$	Quantifies the systematic volumetric deviation. Positive values indicate overestimation; negative values indicate underestimation.	0
POD	NA	$POD = \frac{H}{H+M}$	Represents the fraction of actual observed precipitation events that were successfully predicted (Hit Rate).	1
FAR	NA	$FAR = \frac{F}{H+F}$	Represents the fraction of predicted precipitation events that did not actually occur (False Alarm).	0
CSI	NA	$CSI = \frac{H}{H+F+M}$	Provides a comprehensive assessment of categorical accuracy by penalizing both misses and false alarms.	1

Notation: G_i refers to the baseline and $\bar{G}_i$ is the average of the baseline, S_i and $\bar{S}$ mean the tested precipitation estimates and their average, n refers to the number of samples. H represents the number of precipitation events hit by tested precipitation products; M represents the number of precipitation events that were missed by tested precipitation products; F represents the number of precipitation events that were misreported by tested precipitation products.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Beyond deterministic evaluations, the CRISP dataset offers explicit uncertainty bounds via Conformalized Quantile Regression. To evaluate the mathematical reliability and sharpness of these dynamically generated 90% prediction intervals, two specific probabilistic metrics were utilized:

1. Prediction Interval Coverage Probability (PICP) quantifies the empirical reliability of the generated uncertainty bounds by calculating the percentage of true ground observations that successfully fall within the predicted upper and lower limits. For a mathematically rigorously calibrated 90% prediction interval, the ideal PICP should be approximately 90%. It is defined as:

$$PICP = \frac{1}{n}\sum_{i=1}^{n} c_i \times 100\% \tag{5}$$

where c_i is a boolean condition defined as $c_i = 1$ if $L_i \leq O_i \leq U_i$, and $c_i = 0$ otherwise. Here, L_i and U_i represent the estimated 5th and 95th percentiles (lower and upper bounds) for the i-th sample, respectively.

2. While a 100% PICP could theoretically be achieved by predicting infinitely wide bounds (spanning from zero to infinity), such intervals would provide zero practical information. Therefore, the Average Interval Width is computed to assess the sharpness and informational value of the bounds. A narrower AIW indicates higher model confidence. It is calculated as:

$$AIW = \frac{1}{n}\sum_{i=1}^{n} (U_i - L_i) \tag{6}$$

In this study, the dynamic variation of AIW is explicitly analyzed across different precipitation intensities to demonstrate that the CRISP dataset respects the heteroscedasticity of physical precipitation (i.e., narrower intervals for trace drizzle, wider margins for extreme rainfall).

Author contributions

Y.L. designed the study, developed the methodology, processed the datasets, performed the analyses, and prepared the manuscript. B.Y. conceived and supervised the research, contributed to the interpretation of the results, and revised the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors gratefully acknowledge the organizations and research groups that provided the datasets essential for this research. We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Climate Change Service

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

(C3S) for providing the ERA5 reanalysis data. We acknowledge the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) Global Precipitation Measurement (GPM) mission for the IMERG dataset. The GSMaP precipitation data were kindly provided by the Earth Observation Research Center (EORC), JAXA. We thank NASA and the United States Geological Survey (USGS) for the provision of the SRTM Digital Elevation Model (DEM). Furthermore, we extend our gratitude to the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA) for providing the station observation data. We are also deeply grateful to the Yellow River Conservancy Commission (YRCC) and Prof. Kun Yang's research group for generously providing the independent observation datasets used in this study.

Financial support

This work was supported by National Key Research and Development Program of China (2024YFC3214801), National Natural Science Foundation of China (U2243229), Fundamental Research Funds for the Central Universities (B250205002).

References

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Cheng, Y., Zhang, X., Wang, K., Zhang, Y., Guo, Y., and Shen, Y. J.: Multidimensional evaluation of satellite-based and reanalysis-based precipitation datasets in the Tibetan Plateau, *J. Hydrol.*, 660, 133364, https://doi.org/10.1016/j.jhydrol.2025.133364, 2025.

Cui, X. and Graf, H. F.: Recent land cover changes on the Tibetan Plateau: a review, *Clim. Change*, 94, 47–61, https://doi.org/10.1007/s10584-009-9556-8, 2009.

Das, S.: Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique, *Int. J. Climatol.*, 39, 407–423, https://doi.org/10.1002/joc.5819, 2019.

Frazier, P. I.: Bayesian optimization, in: Recent Advances in Optimization and Modeling, *INFORMS*, 255–278, https://doi.org/10.1287/educ.2018.0188, 2018.

Fushiki, T.: Estimation of prediction error by using K-fold cross-validation, *Stat. Comput.*, 21, 137–146, https://doi.org/10.1007/s11222-009-9153-8, 2011.

Gao, Y. C. and Liu, M.: Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau, *Hydrol. Earth Syst. Sci.*, 17, 837–849, https://doi.org/10.5194/hess-17-837-2013, 2013.

Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., and Francés, F.: Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020), *Atmos. Res.*, 284, 106606, https://doi.org/10.1016/j.atmosres.2023.106606, 2023.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., and Thépaut, J. N.: The ERA5 global reanalysis, *Q. J. Roy. Meteor. Soc.*, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., and Yoo, S. H.: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD), https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf, 2015.

Jiang, Y., Yang, K., Qi, Y., Zhou, X., He, J., Lu, H., and Zhou, J.: TPHiPr: a long-term (1979–2020) high-accuracy precipitation dataset (1/30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations, *Earth Syst. Sci. Data*, 15, 621–638, https://doi.org/10.5194/essd-15-621-2023, 2023.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., and Liu, T. Y.: LightGBM: A highly efficient gradient boosting decision tree, *Adv. Neural Inf. Process. Syst.*, 30, 2017.

Kidd, C. and Levizzani, V.: Status of satellite precipitation retrievals, *Hydrol. Earth Syst. Sci.*, 15, 1109–1116, https://doi.org/10.5194/hess-15-1109-2011, 2011.

Kirkland, E. J.: Bilinear interpolation, in: *Advanced Computing in Electron Microscopy*, Springer, Boston, 261–263, https://doi.org/10.1007/978-1-4419-6533-2_12, 2010.

Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., and Steinbrecher, M.: Multi-layer perceptrons, in: *Computational Intelligence*, Springer, Cham, 53–124, https://doi.org/10.1007/978-3-030-42227-1_5, 2022.

Kubota, T., Aonashi, K., Ushio, T., Shige, S., Takayabu, Y. N., Kachi, M., and Oki, R.: Global Satellite Mapping of Precipitation (GSMaP) products in the GPM era, in: *Satellite Precipitation Measurement*, Springer, Cham, 355–373, https://doi.org/10.1007/978-3-030-24568-9_20, 2020.

Liu, X. and Chen, B.: Climatic warming in the Tibetan Plateau during recent decades, *Int. J. Climatol.*, 20, 1729–1742, https://doi.org/10.1002/1097-0088(20001130)20:14<1729::AID-JOC556>3.0.CO;2-Y, 2000.

Lyu, Y. and Yong, B.: A novel Double Machine Learning strategy for producing high-precision multi-source merging precipitation estimates over the Tibetan Plateau, *Water Resour. Res.*, 60, e2023WR035643, https://doi.org/10.1029/2023WR035643, 2024.

Lyu, Y. and Yong, B.: Using an explainable machine learning approach to produce high-resolution hourly precipitation estimates for a typical data-deficiency basin, *J. Geophys. Res. Mach. Learn. Comput.*, 2, e2024JH000489, https://doi.org/10.1029/2024JH000489, 2025.

Lyu, Y., Yong, B., Huang, F., Qi, W., Tian, F., Wang, G., and Zhang, J.: Investigating twelve mainstream global precipitation datasets: Which one performs better on the Tibetan Plateau?, *J. Hydrol.*, 633, 130947, https://doi.org/10.1016/j.jhydrol.2024.130947, 2024.

McDonald, G. C.: Ridge regression, *WIREs Comput. Stat.*, 1, 93–100, https://doi.org/10.1002/wics.14, 2009.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, *WIREs Water*, 5, e1319, https://doi.org/10.1002/wat2.1319, 2018.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Miao, C., Immerzeel, W. W., Xu, B., Yang, K., Duan, Q., and Li, X.: Understanding the Asian water tower requires a redesigned precipitation observation strategy, *Proc. Natl. Acad. Sci. USA*, 121, e2403557121, https://doi.org/10.1073/pnas.2403557121, 2024.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., and Thépaut, J. N.: 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.

Popat, E. and Hartmann, A.: Exploring cumulative probability functions for streamflow drought magnitude, *J. Hydrol.*, 637, 131426, https://doi.org/10.1016/j.jhydrol.2024.131426, 2024.

Romano, Y., Patterson, E., and Candes, E.: Conformalized quantile regression, *Adv. Neural Inf. Process. Syst.*, 32, 2019.

Shukla, T. and Sen, I. S.: Preparing for floods on the Third Pole, *Science*, 372, 232–234, https://doi.org/10.1126/science.abh3558, 2021.

Su, F., Pritchard, H. D., Yao, T., Huang, J., Ou, T., Meng, F., and Chen, D.: Contrasting fate of western Third Pole's water resources under 21st century climate change, *Earth's Future*, 10, e2022EF002776, https://doi.org/10.1029/2022EF002776, 2022.

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., and Hong, Y.: Have satellite precipitation products improved over last two decades?, *Remote Sens. Environ.*, 240, 111697, https://doi.org/10.1016/j.rse.2020.111697, 2020.

Tang, G., Long, D., Hong, Y., Gao, J., and Wan, W.: Documentation of multifactorial relationships between precipitation and topography of the Tibetan Plateau using spaceborne precipitation radars, *Remote Sens. Environ.*, 208, 82–96, https://doi.org/10.1016/j.rse.2018.02.007, 2018.

Tong, K., Su, F., Yang, D., and Hao, Z.: Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau, *J. Hydrol.*, 519, 423–437, https://doi.org/10.1016/j.jhydrol.2014.07.044, 2014.

Ward, E., Buytaert, W., Peaver, L., and Wheater, H.: Evaluation of precipitation products over complex mountainous terrain: A water resources perspective, *Adv. Water Resour.*, 34, 1222–1231, https://doi.org/10.1016/j.advwatres.2011.05.007, 2011.

Wu, G., Duan, A., Liu, Y., Mao, J., Ren, R., Bao, Q., and Hu, W.: Tibetan Plateau climate dynamics: recent research progress and outlook, *Natl. Sci. Rev.*, 2, 100–116, https://doi.org/10.1093/nsr/nwu045, 2015.

Yang, S., Jones, P. D., Jiang, H., and Zhou, Z.: Development of a near-real-time global in situ daily precipitation dataset, *Int. J. Climatol.*, 40, 2795–2810, https://doi.org/10.1002/joc.6367, 2020.

Yao, T., Thompson, L., Chen, D., and Piao, S.: Reflections and future strategies for Third Pole Environment, *Nat. Rev. Earth Environ.*, 3, 608–610, https://doi.org/10.1038/s43017-022-00342-4, 2022.

Yong, B., Qi, W., Lu, D., Lyu, Y., Liao, A., Wang, G., and Zhang, J.: Streamflow change of major rivers over the Tibetan Plateau during the last half century and its possible causes, *Hydrol. Process.*, 38, e15066, https://doi.org/10.1002/hyp.15066, 2024.

https://doi.org/10.5194/essd-2026-377
Preprint. Discussion started: 19 June 2026

Yong, B., Lyu, Y. (2026). A multidimensional daily precipitation dataset over the Third Pole (1998-2024) featuring physics-aware corrections and conformalized uncertainty quantification. National Tibetan Plateau / Third Pole Environment Data Center. https://doi.org/10.11888/Atmos.tpdc.303469

You, Q., Cai, Z., Pepin, N., Chen, D., Ahrens, B., Jiang, Z., and Zhang, Y.: Warming amplification over the Arctic Pole and Third Pole: Trends, mechanisms and consequences, *Earth-Sci. Rev.*, 217, 103625, https://doi.org/10.1016/j.earscirev.2021.103625, 2021.

Zhan, C. H., Chen, Y. Y., Yang, K., Lazhu, Zhou, X., Jiang, Y. Z., Ling, X. Y., Tian, J. X., Wang, Y., Li, X., and Yang, H.: First evaluation of GPM-Era satellite precipitation products with new observations on the western Tibetan Plateau, *Atmos. Res.*, https://doi.org/10.1016/j.atmosres.2022.106559, 2022.

Zhang, C., Tang, Q., and Chen, D.: Recent changes in the moisture source of precipitation over the Tibetan Plateau, *J. Climate*, 30, 1807–1819, https://doi.org/10.1175/JCLI-D-15-0842.1, 2017.

Zhang, G., Yao, T., Xie, H., Wang, W., and Yang, W.: An inventory of glacial lakes in the Third Pole region and their changes in response to global warming, *Glob. Planet. Change*, 131, 148–157, https://doi.org/10.1016/j.gloplacha.2015.05.013, 2015.