Response to Reviewer #1

This study presents a valuable precipitation dataset over the Third Pole region, with the unique capability of providing precipitation amount, precipitation phase, and associated uncertainty estimates. Given the critical importance of accurate precipitation information for understanding hydrological processes and climate variability over the Tibetan Plateau, this work is highly relevant and potentially useful for the community. Overall, I support publication of this manuscript after revision. However, several aspects should be further improved to enhance the scientific rigor and impact of this study.

Response:

We sincerely thank the reviewer for the positive and encouraging assessment of our manuscript. We appreciate the recognition that the CRISP dataset provides precipitation amount, precipitation phase, and associated uncertainty estimates over the Third Pole region, where reliable precipitation information is particularly important yet challenging to obtain because of complex terrain, sparse observations, and the large proportion of solid precipitation.

We agree that several aspects of the manuscript should be further strengthened to improve its scientific rigor and impact. In the revised manuscript, we will provide a more comprehensive comparison with existing precipitation datasets, strengthen the independent validation of precipitation phase information, and improve the presentation of the methodology, results, and figures. We will also clarify the specific advantages of CRISP, particularly its integrated provision of precipitation amount, precipitation phase, and uncertainty information.

In addition, we will discuss the temporal resolution of CRISP in relation to existing hourly and sub-hourly products and clarify the potential for extending the current framework to higher temporal resolutions in future work. These revisions will help better demonstrate the applicability and distinctive value of CRISP for hydrological, cryospheric, and climate studies over the Third Pole region.

We thank the reviewer again for the constructive suggestions, which will help us substantially improve the manuscript.

1. A more comprehensive comparison with existing precipitation datasets is needed.

The authors should provide a more thorough review and systematic comparison with existing precipitation datasets developed for the Tibetan Plateau and surrounding regions. Several recently developed datasets should be considered, including AIMERG, a new Asian precipitation dataset (0.1°, half-hourly, 2000–2015) generated by calibrating GPM-era IMERG using APHRODITE observations; AERA5-Asia, a long-term Asian precipitation dataset (0.1°, hourly, 1951–2015) developed by constraining ERA5-Land precipitation with APHRODITE-based total volume control; and GMCP, a fully global multisource merging and calibration precipitation dataset (1-hourly, 0.1°, 2000–present).

These datasets have achieved relatively high temporal resolution (e.g., hourly or sub-hourly), and a comprehensive assessment against these products would help better demonstrate the advantages and uniqueness of the proposed CRISP dataset. In addition, the authors should discuss whether the current daily temporal resolution could be further improved to hourly resolution, which would significantly enhance its applicability for hydrological and cryospheric studies.

Response:

Done. In the revised manuscript, we have added a new subsection entitled "Comparison with existing precipitation datasets and temporal resolution" and expanded Table 3 and Table 4 to include AIMERG, AERA5-Asia, and GMCP. The revised subsection summarizes the temporal coverage, spatial and temporal resolutions, input data sources, calibration strategies, and the availability of precipitation-phase and uncertainty information for the compared products.

We have also clarified the procedures used for inter-product evaluation. All hourly and half-hourly datasets were accumulated into daily precipitation totals using the same UTC-based daily definition. The datasets were then transformed to the common 0.1° CRISP grid before comparison.

The new comparison clarifies that the main added value of CRISP is not its temporal resolution, since AIMERG, AERA5-Asia, and GMCP provide hourly or sub-hourly estimates. **Instead, CRISP is designed to provide precipitation amount, precipitation phase, and uncertainty estimates within a unified daily dataset specifically developed for the complex terrain and cold-climate conditions of the Third Pole region.**

We added a concise performance comparison in Section 4.2. The 2011–2015 period was selected because it provides a common period of availability across the evaluated products and sufficient gauge coverage for a consistent comparison. The results are presented in Table 4 and Figure X shown below. **The comparison shows that CRISP provides competitive daily precipitation estimates among the recent developed datasets.**

We further added a discussion of the application value of hourly precipitation data. Hourly or sub-hourly data are particularly useful for flash-flood forecasting, event-scale rainfall-runoff simulation, small-catchment hydrological modeling, short-duration extreme precipitation analysis, and the representation of snow accumulation and melt timing. However, the higher temporal resolution of these datasets may also increase sensitivity to satellite sampling errors, sparse gauge calibration, precipitation intermittency, and phase-dependent measurement errors over the Tibetan Plateau.

Finally, we clarified that an hourly extension of CRISP is technically feasible because the current framework already uses hourly atmospheric variables and high-frequency satellite information. However, reliable hourly production would require hourly training observations, an hourly-specific calibration strategy, and independent validation of precipitation timing, intensity, occurrence, and phase. Direct disaggregation of daily estimates would not adequately represent short-duration precipitation events. We therefore identify hourly CRISP development as a future direction rather than claiming that the current daily product can directly provide validated hourly estimates.

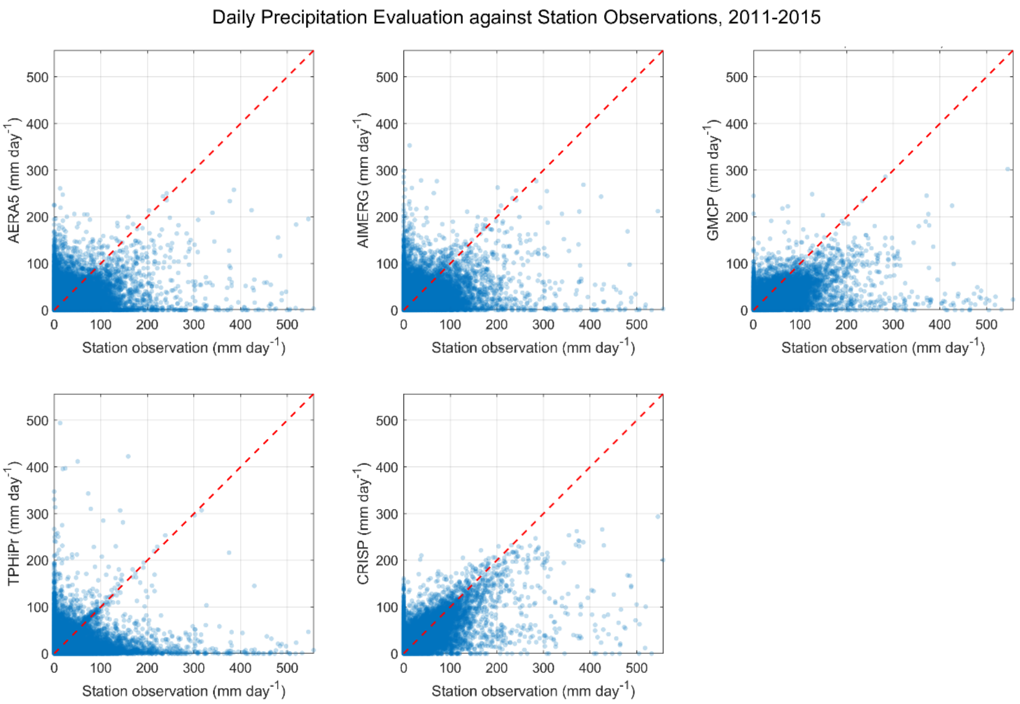

Figure X. The evaluation results of recent precipitation datasets against station observations.

****

Please see Section 4.2 in detail:

4.2 Comparison with existing precipitation datasets and temporal resolution

Ground-based precipitation gauges are useful references for precipitation evaluation, but they are not error-free observations. In particular, wind-induced undercatch can lead to substantial underestimation of solid and mixed precipitation, with the magnitude of the error depending on precipitation type, wind speed, gauge design, and site exposure. These issues are especially relevant over the Tibetan Plateau, where snowfall and strong winds are common. Following the discussions in Kochendorfer et al. (2017), gauge observations used in this study are therefore regarded as available observational references rather than unbiased measurements of true precipitation.

This limitation means that agreement with gauge observations alone cannot fully demonstrate the absolute accuracy of a precipitation product. The gauge-based results are consequently interpreted together with the comparisons against other precipitation products and the independent hydrological and cryospheric consistency analyses. Thus, we compared it with AIMERG, AERA5-Asia and GMCP, in addition to IMERG, GSMaP, ERA5, and TPHiPr. The main characteristics of these datasets are summarized in Table 3. AIMERG, AERA5-Asia, and GMCP provide hourly or sub-hourly precipitation estimates at 0.1° resolution, whereas CRISP provides daily estimates over 1998–2024. The products also differ in their input datasets and calibration strategies. AIMERG and AERA5-Asia use APHRODITE-based information for regional calibration, whereas GMCP is produced through global multi-source precipitation merging and calibration. CRISP integrates satellite retrievals, reanalysis variables, station observations, and terrain information, and additionally provides precipitation-phase classification and prediction intervals.

To ensure a consistent comparison, all datasets were converted to daily precipitation totals using the same UTC-based daily definition and regridded to the 0.1° CRISP grid before evaluation. The spatial comparison was conducted within the geographic boundary of the Tibetan Plateau rather than over the full rectangular input domain. For each product, only grid cells with valid data within this common TP mask were included in the statistical analysis. Therefore, areas outside the TP boundary and regions not covered by a specific dataset were excluded from the corresponding evaluation rather than treated as zero precipitation. TPHiPr has a different effective spatial coverage near the southern margin of the study domain; this difference was considered by applying the common TP mask and by calculating the statistics only over the overlapping valid area.

Table 3. Recent developed precipitation datasets.

Dataset	Coverage	Resolution	Main calibration/source	Phase information	Uncertainty information
CRISP	1998–2024	0.1°; Daily	Satellite, reanalysis, gauges, terrain	Yes	Yes
AIMERG	2000–2015	0.1°; Half-hourly	IMERG calibrated with APHRODITE	No	No
AERA5-Asia	1951–2015	0.1°; Hourly	ERA5-Land constrained by APHRODITE	No	No
GMCP	2000–present	0.1°; Hourly	Global multi-source merging and calibration	No	No
TPHiPr	1979–2020	1/30°; Daily	ERA5_CNN and gauge observations	No	No

The higher temporal resolution of AIMERG, AERA5-Asia, and GMCP is advantageous for applications that depend on the timing and intensity of short-duration precipitation events, such as flash-flood forecasting, event-scale rainfall-runoff simulation, urban or small-catchment hydrological modeling, and sub-daily extreme precipitation analysis. Hourly precipitation information may also improve the representation of snow accumulation and melt timing in cryospheric models, particularly when combined with an energy-balance model or an hourly snowpack model. However, higher temporal resolution does not necessarily ensure higher accuracy or temporal stability over the Tibetan Plateau. At hourly and sub-hourly scales, the estimates are more sensitive to satellite sampling limitations, retrieval errors, sparse gauge calibration, precipitation intermittency, and wind-induced undercatch during snowfall. These uncertainties are particularly important for weak, solid, and mixed-phase precipitation events.

The current daily resolution of CRISP was selected because the primary training observations are daily precipitation records, while reliable and spatially extensive hourly observations remain limited over the Third Pole. Daily aggregation provides a relatively stable basis for multi-source calibration and is appropriate for long-term climatological analysis, daily water-balance assessment, snow accumulation, glacier mass-balance studies, and daily or sub-daily hydrological applications in which daily precipitation is used as the forcing input. The CRISP framework could be extended to hourly resolution because it already uses hourly atmospheric forcing and high-frequency satellite information. Nevertheless, an hourly version would require hourly training observations, an hourly-specific bias-correction strategy, and independent validation of precipitation occurrence, timing, intensity, and phase. Direct temporal disaggregation of daily CRISP estimates would not adequately reproduce the actual timing or short-duration intensity of precipitation events. Therefore, an hourly CRISP product is technically feasible but is left for future development.

Table 4. Evaluation results of recent developed precipitation datasets during the period of 2011-2015.

Dataset	**CC**	**RB (%)**	**RMSE (mm/day)**	**POD**	**FAR**	**CSI**
CRISP	0.75	11.49	8.53	0.90	0.36	0.60
AIMERG	0.49	-10.75	11.31	0.78	0.40	0.51
AERA5-ASIA	0.46	2.66	11.55	0.86	0.49	0.46
GMCP	0.68	-0.37	9.71	0.89	0.45	0.52
TPHiPr	0.60	-4.45	12.86	0.79	0.50	0.43

As shown in Table 4, CRISP achieved the best overall performance among the evaluated datasets. It obtained the highest correlation coefficient (CC = 0.75), the lowest RMSE (8.53 mm/day), and the highest CSI (0.60), indicating better agreement with the gauge observations and a more balanced representation of precipitation occurrence. CRISP also produced the highest POD (0.90) and the lowest FAR (0.36) among the evaluated datasets. GMCP showed the smallest relative bias (RB = −0.37%) and relatively high POD (0.89), but its higher FAR resulted in a lower CSI (0.52). AIMERG had a moderate CSI (0.51) and the second-lowest FAR, but its lower CC and negative bias reduced its overall performance. AERA5-Asia and TPHiPr showed comparatively larger errors, with lower CC and higher RMSE, FAR, or both. These results suggest that CRISP provided the most consistent daily precipitation estimates during 2011–2015, while its positive relative bias (11.49%) indicates that the improvement in event detection and overall agreement was accompanied by a tendency toward precipitation overestimation.

****

2. The validation of precipitation phase information requires further strengthening.

One of the major strengths and unique contributions of this study is the explicit classification of precipitation phases, including rain, snow, and mixed precipitation. However, the current validation of precipitation phase information appears relatively limited. More comprehensive evaluations *using independent observations (e.g., ground-based precipitation phase observations, weather stations, snow observations, or available remote sensing products) are recommended. The authors should provide more quantitative evidence demonstrating the reliability of the precipitation phase classification, particularly over complex terrain and high-altitude regions where phase transitions are highly variable.*

Response:

Done. In the previous version, we evaluated the CRISP phase classification against the official Present Weather codes recorded by the National Meteorological Information Centre (NMIC) station network. These observations provide a direct ground-based reference for precipitation phase and are more appropriate for this purpose than snow-depth data or other indirect indicators. Data from individual flux stations are difficult to obtain, temporally discontinuous, and spatially unrepresentative of the broader Tibetan Plateau. We also did not use remote-sensing precipitation-phase products as independent validation data because CRISP incorporates DPR information, while the accuracy and independence of satellite-based phase retrievals remain uncertain over complex and high-altitude terrain.

We acknowledge that the number of available ground-based precipitation-phase observations is limited because of the sparse station distribution and incomplete temporal coverage. These limitations also prevent a sufficiently robust stratified assessment by elevation or other terrain-related factors, particularly for mixed-phase precipitation.

To further address the reviewer's concern, we added an additional observational benchmark from Wang et al. (2024). This dataset is also derived from the NMIC station network, but its precipitation-phase classification was further adjusted using observations from two OTT Parsivel^2 disdrometers. It therefore provides an additional evaluation based on a different observation-processing procedure, although it is not fully independent in terms of the underlying station network. We evaluated CRISP against this additional dataset using the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) for rain, mixed precipitation, and snow. We also compared the volumetric fractions of the three phases. **The new results are presented in Appendix Fig. A2 and described in the newly added Appendix section C.** CRISP maintains reasonable skill for rain and snow, whereas the performance for mixed precipitation is more variable. The volumetric fractions are broadly comparable between CRISP and the Wang et al. (2024) benchmark. Rain accounts for 76.14% in CRISP and 76.71% in the Wang et al. (2024) dataset, while snow accounts for 17.54% and 13.66%, respectively; the remaining fractions correspond to mixed precipitation. We have also revised the main text to report the complete phase fractions for the original NMIC comparison, including mixed precipitation, and to avoid overstating the agreement as "almost perfect." The revised text now describes the results as broadly consistent and explicitly identifies mixed-phase precipitation and the scarcity of ground-based observations as important sources of uncertainty. These changes provide additional quantitative support for the precipitation-phase classification while clarifying the remaining limitations of the validation.

****

Please see Appendix B in detail:

Appendix B: Additional validation of precipitation-phase classification

To further evaluate the reliability of the precipitation-phase classification in CRISP, we conducted a supplementary comparison with the observational dataset developed by Wang et al. (2024). This dataset was constructed using precipitation-phase observations from the National Meteorological Information Centre (NMIC) station network, with the phase classification further adjusted using measurements from two OTT Parsivel^2 disdrometers. It therefore provides a complementary observational benchmark based on a different phase-classification procedure.

The evaluation considered three precipitation phases: rain, mixed precipitation, and snow. The CRISP classification and the Wang et al. (2024) dataset were matched in space and time for the available station-event samples. For each phase, the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) were calculated using a one-versus-rest contingency table. The distributions of these metrics across the matched station-event samples are shown in Appendix Figure A1a–c. The volumetric fractions of the three phases were also compared between CRISP and the Wang et al. (2024) dataset (Appendix Figure A1d).

Overall, CRISP maintains reasonable skill in identifying rain and snow, whereas the evaluation of mixed precipitation is less robust. This result is expected because mixed precipitation is relatively infrequent and is intrinsically more difficult to classify using both automated instruments and gridded precipitation products. The volumetric fractions are broadly comparable between the two datasets. Rain accounts for 76.14% and 76.71% of the total precipitation volume in CRISP and Wang et al. (2024), respectively.

Snow accounts for 17.54% and 13.66%, while mixed precipitation accounts for 6.32% and 9.63%, respectively.

This additional comparison provides quantitative support for the CRISP phase classification under an alternative observation-based adjustment procedure. However, it should not be interpreted as fully independent validation because the CRISP evaluation and the Wang et al. (2024) dataset both rely on the NMIC station network. The sparse station distribution, incomplete temporal coverage, and limited availability of independent phase observations also prevent a robust stratified evaluation by elevation, terrain, or precipitation intensity.

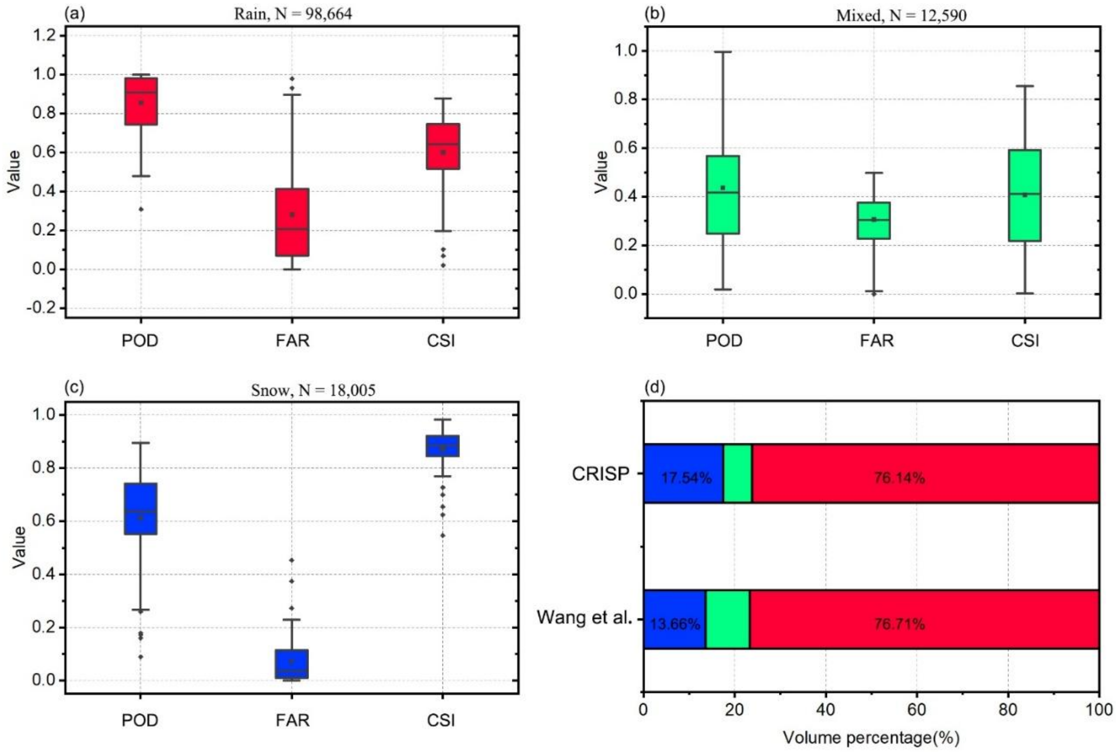

Figure A1. Additional validation of CRISP precipitation-phase classification against the observational dataset of Wang et al. (2024) during the period of 1998-2017. Boxplots show the station-event distributions of (a) probability of detection (POD), (b) false alarm ratio (FAR), and (c) critical success index (CSI) for rain, mixed precipitation, and snow. (d) compares the volumetric fractions of rain, snow, and mixed precipitation between CRISP and the Wang et al. (2024) dataset.

Please see Section 3.3.2 in detail:

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℃ at lower elevations to -1.0℃ for regions above 4,000 m, while pure rainfall is defined strictly above 2.0℃. 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. To evaluate the phase-classification results, we used ground-based precipitation-phase observations that were not used for CRISP training. The primary observational reference was the official Present Weather classification from the National Meteorological Information Centre (NMIC) station network. This reference provides direct station-based information on precipitation phase, including rain, mixed precipitation, and snow.

We further conducted a supplementary evaluation using the dataset developed by Wang et al. (2024). This dataset was also derived from the NMIC station network, but its phase classification was adjusted using measurements from two OTT Parsivel² disdrometers. It therefore provides a complementary benchmark based on a different observation-processing procedure. Because the two datasets share the same underlying NMIC station network, the Wang et al. (2024) comparison is considered supplementary rather than fully independent validation.

For both observational references, precipitation-phase classifications were matched in space and time between the gridded CRISP product and the available station observations. The categorical performance of each phase was evaluated using the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI). For the multiclass phase classification, each phase was evaluated using a one-versus-rest contingency table. In addition, the volumetric fraction of each phase was calculated as the precipitation amount assigned to that phase divided by the total precipitation amount across all matched samples.

Please see Section 4.5 in detail:

In addition to categorical detection skill, the regional hydroclimatic value of phase discrimination can also be assessed by comparing phase-specific precipitation fractions. As shown in Figure 10d, the phase fractions estimated by CRISP are broadly consistent with those obtained from the NMIC observations. CRISP identifies 89.45% rain, 7.90% snow, and 2.65% mixed precipitation, whereas the corresponding observed fractions are 87.86%, 8.86%, and 3.28%, respectively. Thus, CRISP reproduces the dominant rain fraction and provides a comparable estimate of the contribution from solid precipitation, although the mixed-precipitation fraction is somewhat underestimated.

As a supplementary assessment, we compared CRISP with the phase dataset developed by Wang et al. (2024), which uses a different observation-processing procedure involving OTT Parsivel^2 disdrometer measurements. The results are presented in Appendix Figure A1. CRISP maintains reasonable categorical skill for rain and snow, whereas the results for mixed precipitation are more variable. The volumetric fractions are also broadly comparable: rain accounts for 76.14% in CRISP and 76.71% in the Wang et al. (2024) dataset, while snow accounts for 17.54% and 13.66%, respectively. The corresponding mixed-precipitation fractions are 6.32% and 9.63%. These results provide additional support for the phase-classification capability of CRISP, while also indicating that mixed-phase precipitation remains the largest source of uncertainty.

****

3.The manuscript presentation and visualization should be further improved.

The writing, structure, and figures require additional refinement. The authors should further emphasize the unique characteristics and scientific contributions of the CRISP dataset, particularly its advantages in simultaneously providing precipitation estimates, uncertainty information, and precipitation phase classification. Clearer comparisons with existing datasets and more effective visualization of the dataset's distinctive features would substantially improve the readability and impact of the manuscript.

Response:

Done. We have carefully revised the manuscript to improve its overall presentation, structure, and visualization. The unique characteristics of the CRISP dataset are now emphasized more clearly, particularly its ability to provide precipitation estimates together with uncertainty information and precipitation-phase classification within a unified framework.

We also revised the descriptions of the dataset and added clearer comparisons with existing precipitation products. The relevant figures and captions were improved to better illustrate the spatial distribution, temporal variation, uncertainty characteristics, and precipitation-phase information of CRISP. **In addition, redundant descriptions were reduced, and the organization and wording of several sections were refined to improve readability and highlight the scientific**

contributions of the dataset more effectively. These revisions clarify the advantages and potential applications of CRISP, particularly in data-scarce and high-altitude regions.

****

Please see Introduction as example:

Regional precipitation datasets have subsequently been developed by integrating available gauge observations with high-resolution atmospheric simulations or gridded background fields. For example, the Tibetan Plateau High-resolution Precipitation (TPHiPr) dataset improves precipitation estimates through dense observation-based correction (Jiang et al., 2023). Several recent developed products, including AIMERG, AERA5-Asia, and GMCP, have further improved precipitation estimation over Asia and the TP (Ma et al., 2020, 2022, 2025). These products provide enhanced temporal resolution and regional coverage, but they differ in their calibration strategies and in whether they provide information on precipitation phase and estimation uncertainty. Moreover, many regional products rely on global reanalysis datasets as background fields. This dependence can propagate systematic errors from the parent models, including excessive light-precipitation frequency and insufficient representation of extreme rainfall intensity (Ward et al., 2011).

Our previous work introduced a double machine-learning framework for precipitation estimation over the TP and its subregions (Lyu & Yong, 2024, 2025). The framework combined an initial wet or dry estimation step with a subsequent gauge-based error-correction step, thereby improving the representation of topographic and atmospheric effects on precipitation. Although this approach outperformed several benchmark products under specific conditions, including MSP and TPHiPr, it primarily focused on deterministic precipitation accuracy. CRISP extends this line of work in three ways. First, it integrates multiple precipitation, topographic, and atmospheric data sources within a unified stacking framework, allowing the model to use spatially continuous environmental information in areas with limited observations. Second, it applies distribution-free conformalized quantile regression to produce calibrated prediction intervals and precipitation probabilities rather than a single deterministic estimate. The intervals reflect the predictive uncertainty learned from the available predictors and are not intended to explicitly decompose measurement errors, sampling biases, retrieval uncertainties, or other individual sources of uncertainty. Third, it includes a thermodynamic phase-discrimination module to classify precipitation events as dry, snow, mixed, or rain. The framework also accounts for measurement-related effects, particularly potential wind-induced undercatch, by incorporating atmospheric and surface predictors that influence gauge exposure and precipitation measurements. These predictors help the models learn systematic variations in catch efficiency associated with near-surface wind conditions, precipitation phase, and the surrounding surface environment, thereby reducing the potential bias in precipitation estimates caused by wind-driven undercatch.

…

The primary contribution of CRISP is therefore not the generation of completely independent precipitation information, but the integration of multiple complementary data sources within a unified framework and the provision of information that is not commonly available in existing products. In particular, CRISP provides precipitation amount together with calibrated uncertainty information, precipitation occurrence probability, and precipitation-phase classification. These variables are important for hydrological, cryospheric, and atmospheric applications, especially over regions where precipitation phase, measurement uncertainty, and unresolved small-scale processes strongly affect the interpretation of precipitation estimates.

Please see Section 2.3 as example:

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. Although the gauge-adjusted satellite products evaluated in this study (e.g., IMERG calibrated via GPCC and GSMaP adjusted via CPC) also incorporate gauge observations, the GTS dataset provides additional station-based measurements with local information that may not be fully retained after large-scale satellite calibration and merging. These observations were therefore included as an additional, spatially detailed constraint for model training. We acknowledge that GTS measurements, particularly under snowfall and complex terrain conditions, may also be affected by undercatch and representativeness errors; thus, they were not treated as an error-free reference.

Prior to serving as the training target, the raw gauge observations were corrected by the authors for potential wind-induced undercatch. The correction factor, referred to as the catch ratio (CR), was empirically calculated from concurrent near-surface air temperature (T, °C) and 10-m wind speed (WS). These variables were obtained from the NMIC station streams when available and were supplemented with co-located ERA5 reanalysis data when station observations were missing. The dataset included 484 stations, although the number of available stations varied among years; approximately 30% of the station-day records required ERA5 supplementation. We also derived the gauge-height wind speed (WS_g, constrained to a minimum of 0.1m/s) via an empirical scaling factor of 0.7. The aerodynamic CR was dynamically calculated based on precipitation phase thresholds:

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

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

For mixed-phase precipitation (-2°C≤T≤2°C), *CR* was determined through linear interpolation:

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

To prevent unrealistically large precipitation corrections 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. The corrected gauge observations were then used as the training targets for the CRISP framework, allowing the model to learn precipitation patterns and residual uncertainties associated with solid precipitation in alpine environments.

For independent validation, we incorporated supplementary datasets that were completely excluded from model training because of their limited temporal coverage. These include hydrological observations from the Yellow River Conservancy Commission (Lyu & Yong, 2025) and observations from a specialized 53-station high-altitude network (Zhan et al., 2022). The independent stations differ from the meteorological stations in their deployment purpose, instrumentation, temporal coverage, and environmental representation. The hydrological stations were primarily established to monitor key catchments or specific river sections rather than for routine meteorological monitoring, and they generally operate only during the flood or warm season. The specialized high-altitude network consists of spatial transects equipped with tipping-bucket rain gauges with a 0.2 mm resolution and provides warm-season hourly observations from 2017 to 2020. Hydrological observations also commonly have a 0.2 mm gauge resolution, whereas the meteorological stations generally use gauges with a 0.1 mm resolution. The locations of the independent stations are shown in Figure 2. These stations sample high-elevation, remote, and topographically complex environments that are relatively underrepresented by the meteorological network. Consequently, their validation statistics are not expected to be directly equivalent to those derived from the meteorological stations. Instead, they provide an independent and relatively stringent assessment of model transferability beyond the environments represented by the training network.

Please see Section 4.6 as example:

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. The persistent deficit between raw gauge precipitation and PET is consistent with substantial undercatch, particularly for solid precipitation, although it may also reflect uncertainties in PET, storage changes, runoff, and spatial representativeness. CRISP produces estimates that are higher than the raw gauge observations and reduces this apparent deficit. This behavior is consistent with a partial correction of gauge undercatch, but the PET comparison alone cannot establish the absolute accuracy of the CRISP estimates. By incorporating atmospheric and topographic factors to implicitly compensate for wind-induced undercatch, CRISP systematically elevates precipitation estimates above the raw observational baseline, producing estimates that are more consistent with the broad hydrological plausibility constraint provided by PET. Consequently, CRISP can provide a physically consistent moisture input thereby reducing the apparent precipitation deficit relative to the raw gauge observations. The water-balance analysis is designed to provide a physical consistency constraint that is complementary to direct gauge-based validation (Figure 12b). Since both gauge precipitation and SWE data contain measurement or processing uncertainties, the purpose of this analysis is not to establish an exact closure of the water budget. Rather, it examines whether the precipitation estimates exhibit a physically plausible relationship with the accumulation and evolution of snow water storage. 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), suggesting substantial aerodynamic undercatch in solid precipitation measurements. Satellite algorithms (IMERG, GSMaP) largely exhibit 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 reduces the apparent underestimation relative to the raw gauge observations and shows potential for improving the plausibility of the mass-balance relationship. Concurrently, it avoids the excessive volumetric overestimations typical of reanalysis products. By integrating wind speed, temperature, and terrain roughness, CRISP was adjusted to compensate for the wind-induced loss of solid precipitation.

****

Overall, this study provides an important contribution by developing a multidimensional precipitation dataset for the Third Pole region. With strengthened literature comparisons, more comprehensive validation of precipitation phase information, and improved presentation, the manuscript would become a valuable resource for hydrological, cryospheric, and climate research communities.

Response:

We sincerely thank the reviewer for the positive assessment of our revised manuscript. We appreciate the reviewer’s constructive comments, which have helped us further improve the clarity, scientific framing, and reproducibility of the study.

References

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

Ma, Z., Xu, J., Dong, B., Hu, X., Hu, H., Yan, S., ... & Weng, F. (2025). GMCP: a fully global multisource merging-and-calibration precipitation dataset (1-hourly, 0.1°, global, 2000–the present). *Bulletin of the American Meteorological Society*, *106*(4), E596-E624. https://doi.org/10.1175/BAMS-D-24-0051.1

Ma, Z., Xu, J., Ma, Y., Zhu, S., He, K., Zhang, S., ... & Xu, X. (2022). AERA5-ASIA-Asia: a long-term asian precipitation dataset (0.1°, 1-hourly, 1951–2015, Asia) anchoring the ERA5-land under the total volume control by APHRODITE. *Bulletin of the American Meteorological Society*, *103*(4), E1146-E1171. https://doi.org/10.1175/BAMS-D-20-0328.1

Ma, Z., Xu, J., Zhu, S., Yang, J., Tang, G., Yang, Y., ... & Hong, Y. (2020). AIMERG: a new Asian precipitation dataset (0.1∘/half-hourly, 2000–2015) by calibrating the GPM-era IMERG at a daily scale using APHRODITE. Earth System Science Data, 12(3), 1525-1544. https://doi.org/10.5194/essd-12-1525-2020

Wang, C., Zhao, L., Ma, L., Hu, G., Zhang, L., Zou, D., ... & Zhao, J. (2024). Precipitation adjustment by the OTT Parsivel2 in the central Qinghai–Tibet Plateau. *Earth Surface Processes and Landforms*, *49*(8), 2424-2441. https://doi.org/10.1002/esp.5836