Reviewer 2

Comments to the Author

Overall, this dataset fills an important gap in publicly accessible river discharge observations across China and will likely be of broad interest to the hydrological and remote sensing communities. My main concerns relate to the clarity and reproducibility of the methodology, the validation strategy, and the dataset uncertainty.

Response: We sincerely appreciate the reviewer's constructive feedback and are greatly encouraged by the reviewer's recognition of the value of the Chinese Daily River Discharge Record (CDR^2) dataset in filling an important observational gap. In response to the reviewer's concerns, we have revised the manuscript to provide more complete methodological details to ensure reproducibility, clarify our validation strategy, and incorporate a systematic evaluation of dataset uncertainty. Below please find our detailed responses addressing your comments.

Major comments

1. The in situ discharge observations used to establish the width-discharge relationships are primarily from 2014-2024, whereas the reconstructed discharge records span 1990-2024. This reconstruction implicitly assumes that the width-discharge relationship remains stable over the entire reconstruction period. However, channel morphology and hydraulic conditions may have changed substantially over the past three decades due to dam construction, reservoir regulation, channel engineering, sediment erosion/deposition, sand mining, or river course modifications. At a minimum, I recommend explicitly acknowledging this assumption and discussing its implications in the Limitations section. In addition, given that Landsat observations are available before 1990, could the authors clarify why 1990 was selected as the starting year of the reconstruction? Extending the dataset further back in time (e.g., to the 1980s, or earlier if feasible) could potentially provide additional value for long-term trend analyses.

Response 1: We sincerely thank the reviewer for raising these critical points. We fully agree that the implicit assumption of a stable width–discharge relationship over a 35-year period is a significant source of uncertainty, given the potential for morphological and hydraulic changes induced by both natural processes and anthropogenic activities. Accordingly, we have added a dedicated subsection "4.3 Limitations, uncertainties and future perspectives" to the revised manuscript (Page 29, Line 618), which provides a thorough discussion of this issue:

4.3 Limitations, uncertainties, and future perspectives

Although the CDR^2 dataset substantially extends the observational basis for long-term global gauge-scale river discharge monitoring, it remains subject to inherent limitations and uncertainties arising from both remote sensing retrieval and hydrological assumptions. First, a critical source of uncertainty lies in the implicit assumption of temporal stationarity in the width–discharge relationship. Model calibration relies on in situ observations acquired during 2014–2024 and is subsequently applied retrospectively to reconstruct discharge over 1990–2024. The conversion of short-term observational hydraulic relationships into multi-decadal discharge estimates is a common simplification strategy in current satellite-based discharge retrieval and fundamentally assumes temporal invariance of the AHG relationship (Pavelsky, 2014; Paris et al., 2016). To evaluate uncertainty associated with temporal extrapolation, we quantified the extrapolation ratio for each gauge, defined as the proportion of remotely sensed river widths during 1990–2024 that fall outside the calibration width range (Fig. 12). Extrapolation ratios were generally low across the 310 reconstructed gauges, with a median of 0.81% and a mean of 2.62%. Among them, 86.5% of gauges exhibit extrapolation rates below 5%, and 55.2% below 1%. These results indicate that historical discharge reconstruction at most gauges remained within the calibrated width range, substantially mitigating uncertainties associated with extrapolating AHG relationships beyond observed river states. Nevertheless, river hydraulic responses may undergo significant changes over the 35-year study period due to both anthropogenic and natural disturbances, including dam construction, reservoir regulation, channel engineering, sediment erosion and deposition, and sand mining (Mansanarez et al., 2019; Zhang et al., 2015), introducing additional uncertainty into long-term discharge reconstruction. For example, downstream channel incision caused by sediment retention in upstream reservoirs may substantially increase water depth and flow velocity without corresponding changes in channel width.

Second, using river width as the sole remote-sensing hydraulic variable for discharge retrieval is another important limitation. Although river width exhibits strong correlations with discharge across most selected gauges (median ρ = 0.88), discharge is jointly governed by water depth, flow velocity, channel roughness, and bed slope (Leopold and Maddock, 1953; Singh et al., 2003). These hydraulic controls are not directly constrained in the current framework and may be particularly important in rivers where width responds weakly to discharge variability. This limitation is supported by the sensitivity analysis of b, which shows progressively improved reconstruction performance with increasing hydraulic sensitivity (Fig. 11). Additionally, the inherent limitations of optical remote sensing introduce further uncertainty. The temporal continuity of CDR^2 is inevitably affected by cloud cover, atmospheric aerosols, and seasonal snow and ice (Feng et al., 2022; Langhorst et al., 2024), resulting in substantial differences in observation frequency among basins (Fig. 9b). More critically, intense rainfall events are frequently accompanied by persistent cloud cover, causing optical sensors to miss peak-flow observations and hindering the detailed characterization of rapid flood hydrographs.

Future advances will rely on integrating multi-source satellite observations with multi-dimensional hydraulic information. Incorporating channel morphology, sediment dynamics, and anthropogenic regulation will help identify temporal variations in width–discharge relationships and improve the robustness of historical discharge reconstruction. Concurrently, combining optical imagery with synthetic aperture radar (SAR) observations can substantially improve temporal sampling under persistent cloud cover while enhancing the detection of extreme flood events. In addition, further integration of river width with other hydraulic observations (e.g., water surface elevation and slope) from emerging satellite missions, such as the Surface Water and Ocean Topography (SWOT) mission, through the development of nonlinear inversion models that exploit multivariate information holds promise for achieving tighter constraints and more accurate estimation of river discharge.

We agree that extending the time series back to the 1980s could theoretically expand the observational window and provide additional insights for long-term trend analysis. However, given the pronounced limitations of the early Landsat archive in spatial coverage, temporal continuity, and data quality, we designated 1990 as the starting year to safeguard the reliability of river width retrieval and subsequent long-term discharge analysis. During its 1980s commercial phase (the EOSAT era), Landsat suffered from limited onboard storage and relied heavily on line-of-sight downlinks to a limited number of international ground stations, resulting in highly fragmented data acquisition outside the United States (Wulder et al., 2016). To clarify this, we have appended the following text to Section 2.2.2 of the manuscript: “Considering the limited spatiotemporal coverage, high geolocation errors, and low signal-to-noise ratios in the early Landsat-5 archive (Roy et al., 2014; Vogelmann et al., 2001; Wulder et al., 2016), which compromise high-frequency and high-precision river width retrieval, we set 1990 as the starting point of the time series.” (Page 10, Line 219 and Page 11, Lines 220–222)

Statistical analysis of the Global Long-term River Width (GLOW) dataset (Feng et al., 2022) further corroborates the severe scarcity of early data. As the first global Landsat-based river width record spanning decades (1984–2020), GLOW contains more than one billion observations from approximately 11.6 million cross-sections. To quantify the sparseness of early data, we calculated for each cross-section the proportion of observations before 1990 relative to its total observations (Fig. R2). Under the ideal assumption of uniform sampling, the theoretical proportion for 1984–1989 (6 years) relative to the full study period (37 years) is 16.22%. However, the median proportion across cross-sections is only 7.20%, substantially lower than the expected value. Notably, approximately 83% of cross-sections fall below this theoretical benchmark, and 33.15% have no observations prior to 1990 at all. The 1984–1989 period spans only six years, whereas the 1990–2024 period used in this study covers 35 years, which is sufficient to capture long-term hydrological trends. Critically, low-quality and low-frequency early river width observations introduce large errors that undermine the reliability of trend analyses based on reconstructed discharge.

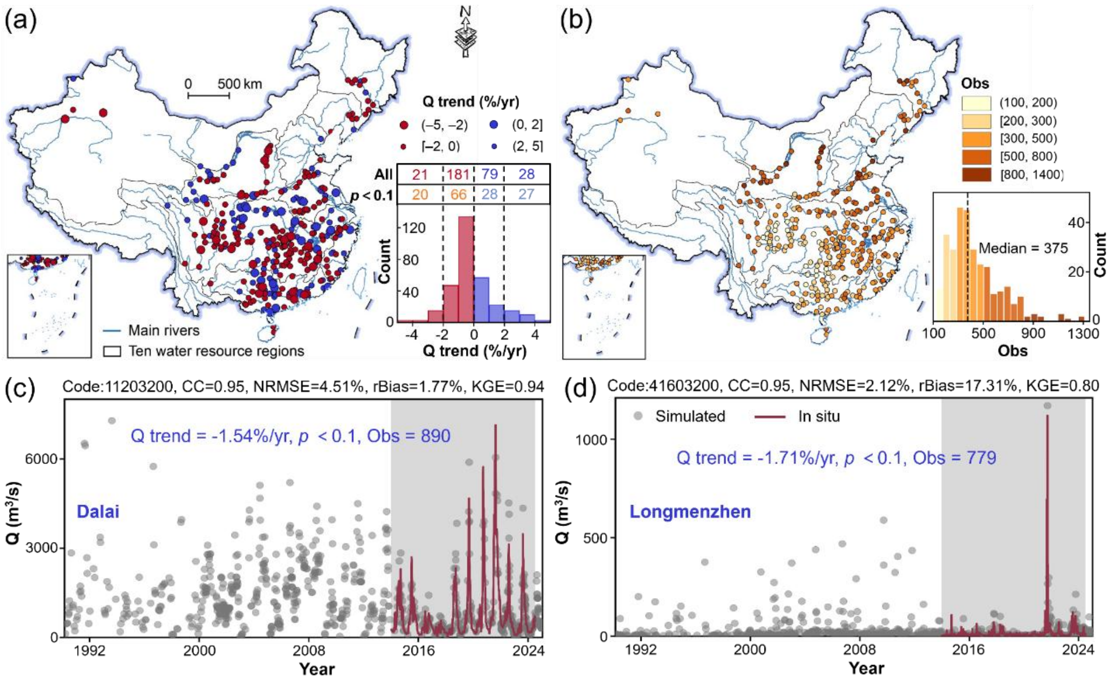

Figure 9. Spatial patterns and long-term trends of satellite-extended river discharge across 309 gauges in China (1990–2024). (**a**) Spatial distribution of relative discharge change rates (Q trend, %/yr), normalized by multi-year mean discharge, with inset histograms showing the trend distribution and corresponding significance levels (p) across all gauges. (**b**) Spatial distribution of additional observations (Obs) provided by the satellite-extended discharge records at each gauge. (**c–d**) Representative time series of reconstructed daily-scale discharge estimates at Dalai and Longmenzhen gauges,

respectively, illustrating long-term discharge variability and trends together with corresponding model performance metrics. (Page 24, Lines 516–524)

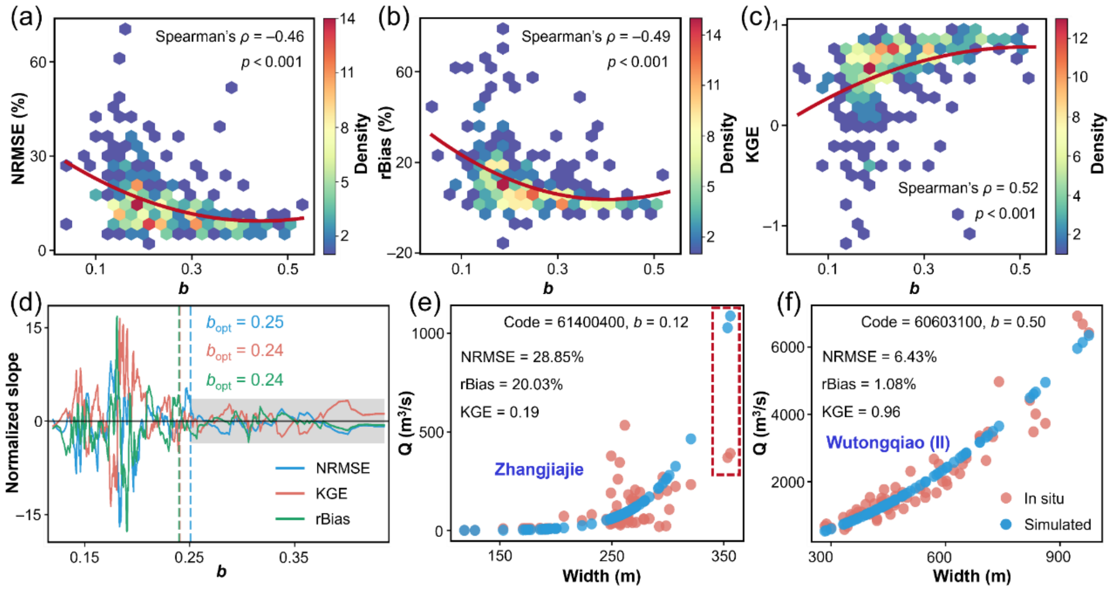

Figure 11. Influence of the width–discharge power-law exponent (b) on the performance of satellite-extended discharge reconstruction. (**a–c**) Relationships between the hydraulic geometry exponent (b) and the performance metrics NRMSE, rBias, and KGE, respectively. Red solid lines indicate fitted relationships, with the corresponding Spearman rank correlation coefficient (ρ) and significance level (p) indicated. (**d**) Variation in the normalized slope of performance metrics with respect to b, used to identify the optimal threshold (b_{opt}) beyond which reconstruction performance becomes stable and less sensitive to changes in hydraulic responsiveness. (**e–f**) Comparison between in situ and reconstructed discharge at the Zhangjiajie and Wutongqiao (II) gauges, representing examples of low-b and high-b hydraulic conditions, respectively. (Page 28, Lines 608–617)

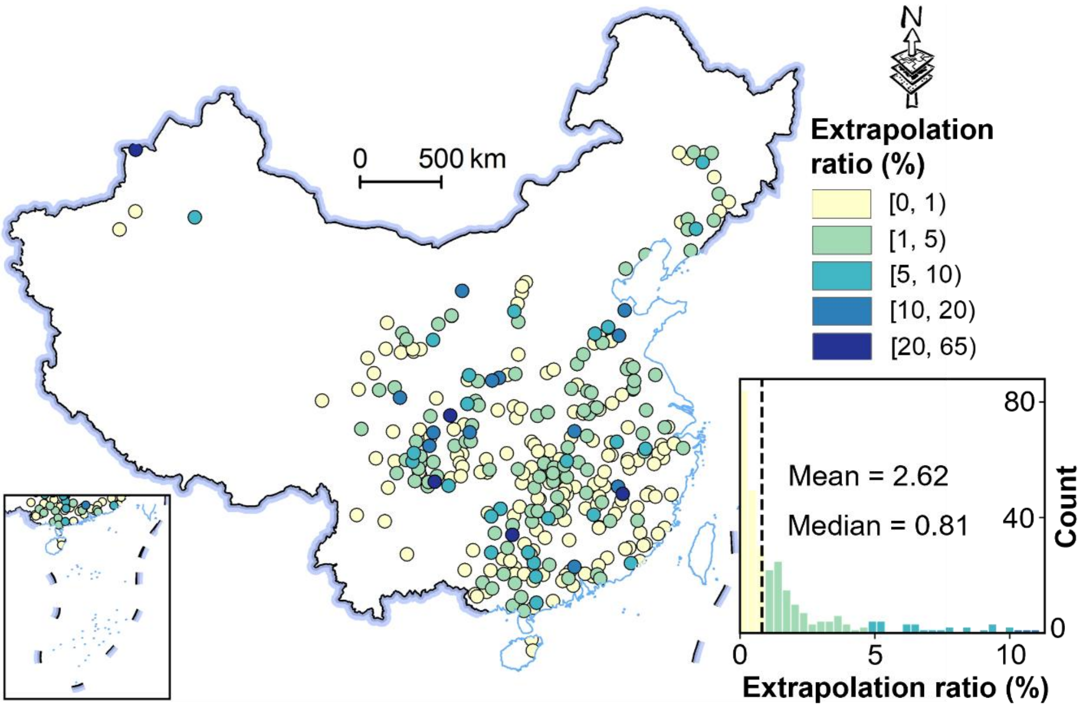

Figure 12. Spatial distribution of extrapolation ratios (%) for the 310 gauges included in the CDR^2 dataset. The inset shows the frequency distribution of these extrapolation ratios across all gauges (mean = 2.62, median = 0.81), truncated to the 0–95th percentile for improved visualization. (Page 31, Lines 667–670)

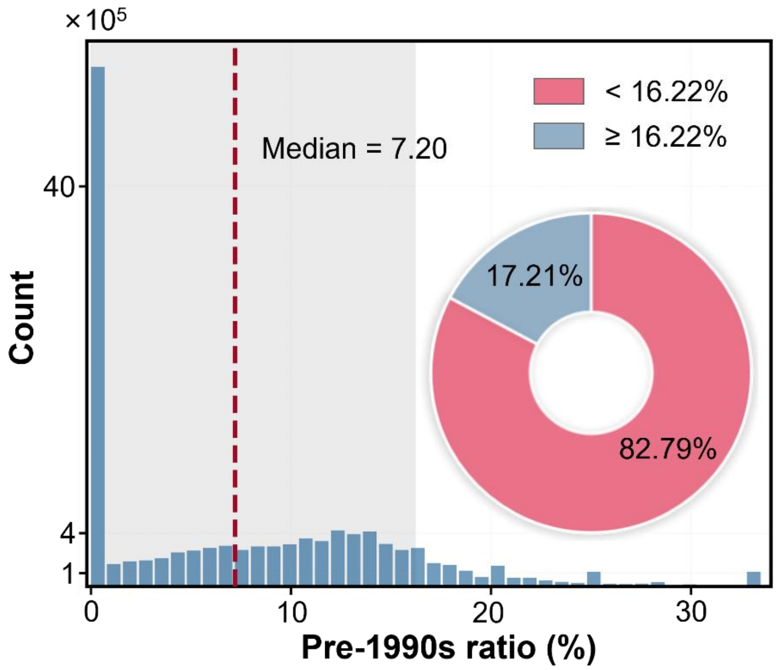

Figure R2. Frequency distribution of pre-1990s observation ratios (%) across river reaches in the Global Long-term River Width (GLOW) dataset (Feng et al., 2022), with a median of 7.20%. The ratio represents the proportion of river width observations collected before 1990 relative to the total observations for each reach. The analysis includes ~11.56 million river reaches. The shaded region denotes reaches with ratios below the theoretical expectation (16.22%), while the inset donut chart summarizes the corresponding proportions. Values above the 95th percentile were truncated for improved visualization.

References

Feng, D., Gleason, C. J., Yang, X., Allen, G. H., and Pavelsky, T. M.: How have global river widths changed over time?, Water Resour. Res., 58, e2021WR031712, https://doi.org/10.1029/2021WR031712, 2022.

Langhorst, T., Andreadis, K. M., and Allen, G. H.: Global cloud biases in optical satellite remote sensing of rivers, Geophys. Res. Lett., 51, e2024GL110085, https://doi.org/10.1029/2024GL110085, 2024.

Leopold, L. B. and Maddock, T.: The hydraulic geometry of stream channels and some physiographic implications, U.S. Government Printing Office, Washington, DC, https://pubs.usgs.gov/pp/0252/report.pdf (last access: 12 March 2026), 1953.

Mansanarez, V., Renard, B., Le Coz, J., Lang, M., and Darienzo, M.: Shift happens! Adjusting stage-discharge rating curves to morphological changes at known times, Water Resour. Res., 55, 2876–2899, https://doi.org/10.1029/2018WR023389, 2019.

Pavelsky, T. M.: Using width-based rating curves from spatially discontinuous satellite imagery to monitor river discharge, Hydrol. Process., 28, 3035–3040, https://doi.org/10.1002/hyp.10157, 2014.

Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P.-A., Collischonn, W., Bonnet, M.-P., and Seyler, F.: Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin, Water Resour. Res., 52, 3787–3814, https://doi.org/10.1002/2014WR016618, 2016.

Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., Hipple, J. D., Hostert, P., Huntington, J., Justice, C. O., Kilic, A., Kovalskyy, V., Lee, Z. P., Lymburner, L., Masek, J. G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R. H., and Zhu, Z.: Landsat-8: Science and product vision for terrestrial global change research, Remote Sens. Environ., 145, 154–172, https://doi.org/10.1016/j.rse.2014.02.001, 2014.

Singh, V. P., Yang, C. T., and Deng, Z. Q.: Downstream hydraulic geometry relations: 1. Theoretical development, Water Resour. Res., 39, 1337, https://doi.org/10.1029/2003WR002484, 2003.

Vogelmann, J. E., Helder, D., Morfitt, R., Choate, M. J., Merchant, J. W., and Bulley, H.: Effects of Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper Plus radiometric and geometric calibrations and corrections on landscape characterization, Remote Sens. Environ., 78, 55–70, https://doi.org/10.1016/S0034-4257(01)00249-8, 2001.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., and Roy, D. P.: The global Landsat archive: Status, consolidation, and direction, Remote Sens. Environ., 185, 271–283, https://doi.org/10.1016/j.rse.2015.11.032, 2016.

Zhang, W., Wang, W.-G., Zheng, J.-H., Wang, H.-G., Wang, G., and Zhang, J.-S.: Reconstruction of stage–discharge relationships and analysis of hydraulic geometry variations: The case study of the Pearl River Delta, China, Glob. Planet. Change, 125, 60–70, https://doi.org/10.1016/j.gloplacha.2014.12.004, 2015.

2. The manuscript reports validation statistics based on concurrent width-discharge pairs. However, it is currently unclear whether the calibration and validation datasets are independent. Specifically, readers would benefit from knowing:

- how many observations were used for model calibration versus validation;
- whether an independent testing dataset, cross-validation, or leave-one-year-out validation was employed; or
- whether the same width-discharge pairs were used for both model fitting and performance evaluation.

Response 2: We sincerely thank the reviewer for this insightful and highly constructive comment. The reviewer raised a critical point regarding the independence of our model validation. In the original manuscript, the same concurrent width–discharge pairs at each gauge were used for both model fitting and performance evaluation—essentially an in-sample fitting (ISF) approach. We completely agree that clarifying the independent predictive capability of the models is essential for the readers.

To rigorously address your concern and test the predictive independence of our models, we have conducted a leave-one-out cross-validation (LOOCV) analysis for all gauges, and the following description has been added to Section 3.2 "Given the limited sample size at individual gauges, and because the discharge predictive capability of

AHG relationships is not the focus of this study, the in-sample fitting (ISF) approach was adopted for final parameterization to maximize the calibration sample size and thereby ensure the robustness of the long-term reconstruction." (Page 17, Lines 360–363) and "To test the robustness of the performance evaluation, a leave-one-out cross-validation (LOOCV) analysis was conducted (Kim et al., 2014). Compared with the ISF results, LOOCV exhibited only marginal performance degradation, with median decreases in CC and KGE ranging from 0.01 to 0.04, while differences in NRMSE and rBias remained below 0.7% (Fig. A1). These minor differences indicate that the ISF approach does not introduce substantial overfitting and that the reconstructed discharge estimates are robust to the choice of calibration strategy." (Page 17, Lines 381–386)

In addition, we have added related statements in Section 4.3 to evaluate the uncertainty associated with temporal extrapolation beyond the calibration range: "To evaluate uncertainty associated with temporal extrapolation, we quantified the extrapolation ratio for each gauge, defined as the proportion of remotely sensed river widths during 1990–2024 that fall outside the calibration width range (Fig. 12). Extrapolation ratios were generally low across the 310 reconstructed gauges, with a median of 0.81% and a mean of 2.62%. Among them, 86.5% of gauges exhibit extrapolation rates below 5% and 55.2% below 1%. These results indicate that historical discharge reconstruction at most gauges remained within the calibrated width range, substantially mitigating uncertainties associated with extrapolating AHG relationships beyond observed river states." (Page 29, Lines 627–635)

Overall, these additional analyses demonstrate that the reported reconstruction performance is robust to the validation strategy and that the long-term discharge reconstruction involves only limited temporal extrapolation beyond the calibrated hydraulic range.

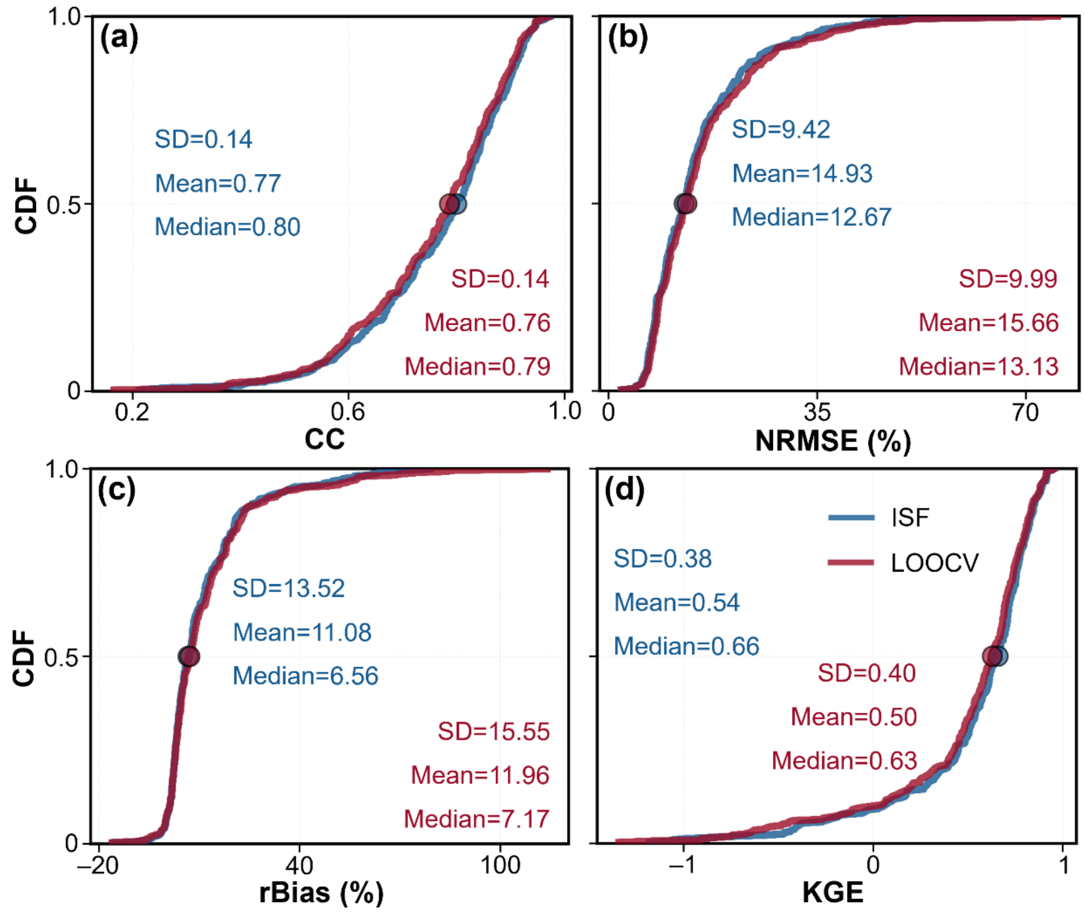

Figure A1. Comparison of performance evaluation for satellite-extended river discharge across 310 gauges using in-sample fitting (ISF) and leave-one-out cross-validation (LOOCV) based on concurrent width–discharge pairs. (**a**–**d**) Cumulative distribution functions (CDF) of the Pearson correlation coefficient (CC), normalized root mean square error (NRMSE, %), relative bias (rBias, %), and Kling–Gupta efficiency (KGE), respectively. For each metric, the mean, median, and standard deviation (SD) are annotated for both ISF and LOOCV. (Page 33, Lines 721–724 and Page 34, Lines 725–727)

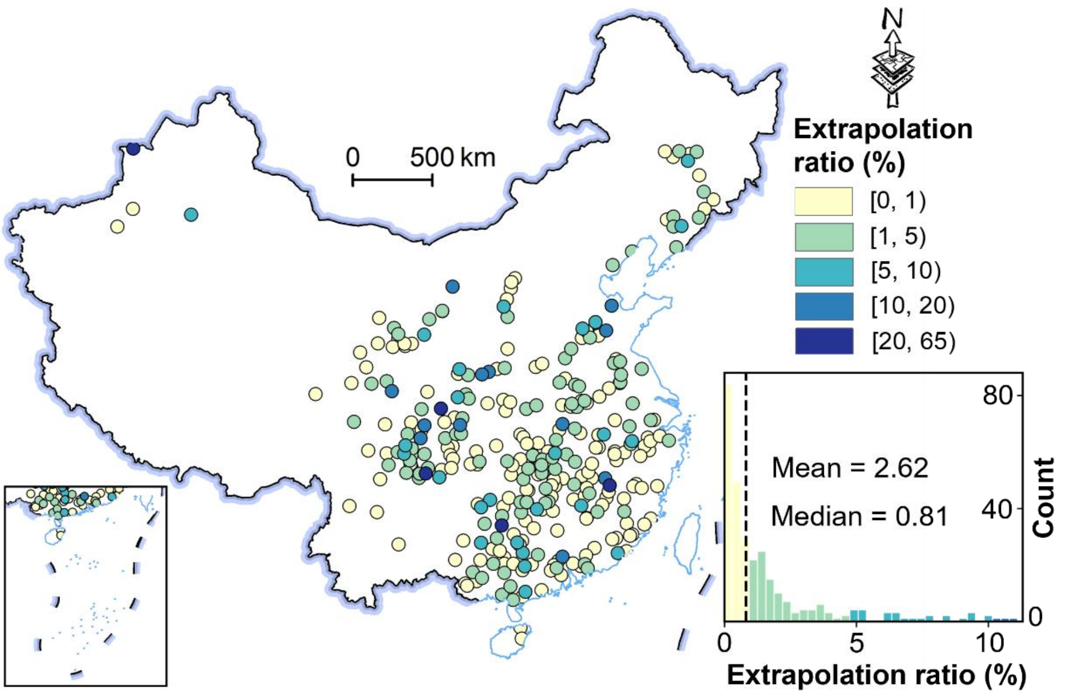

Figure 12. Spatial distribution of extrapolation ratios (%) for the 310 gauges included in the CDR^2 dataset. The inset shows the frequency distribution of these extrapolation ratios across all gauges (mean = 2.62, median = 0.81), truncated to the 0–95th percentile for improved visualization. (Page 31, Lines 667–670)

References

Kim, Y. H., Im, J., Ha, H. K., Choi, J.-K., and Ha, S.: Machine learning approaches to coastal water quality monitoring using GOCI satellite data, GIScience Remote Sens., 51, 158–174, https://doi.org/10.1080/15481603.2014.900983, 2014.

3. I think the dataset would benefit from a more comprehensive assessment of uncertainty. In particular:

- Are uncertainty estimates or confidence intervals available for the reconstructed discharge values?
- What is the uncertainty associated with the fitted AHG parameters?

Response 3: We sincerely thank the reviewer for this valuable comment. We agree that a comprehensive uncertainty assessment is essential for evaluating the reliability of satellite-based discharge reconstruction. In the original manuscript, uncertainties associated with the fitted AHG parameters and their propagation into discharge estimates were not explicitly quantified. To address this concern, we have added a new uncertainty quantification framework in Section 2.4 (Page 14, Line 298) and corresponding results in Section 3.3 (Page 20, Line 423) based on a nonparametric bootstrap resampling approach.

2.4 Uncertainty evaluation

To quantify the uncertainty in fitting the AHG relationship parameters and reconstructing the discharge series, a nonparametric bootstrap resampling was implemented for each gauge. This method is particularly robust for hydrological applications as it requires no prior assumptions regarding the error distribution of remote sensing observations (Hesterberg et al., 2011; Rodrigues et al., 2015). For each bootstrap realization, N pairs of width–discharge observations were sampled with replacement from the original observations (N), and the AHG relation in Eq. (4) was refitted using nonlinear least squares. Repeating this procedure 1,000 times yielded empirical distributions of the coefficient *a* and exponent *b*. Their relative uncertainties, RU_a and RU_b, were defined as the ratio of the half-width of the 95% confidence interval to the median of the distribution:

$$\mathrm{RU_a} = \frac{(\mathrm{a}_{97.5} - \mathrm{a}_{2.5})}{2\mathrm{a}_{50}} \times 100\% \tag{10}$$

$$\mathrm{RU}_b = \frac{(b_{97.5} - b_{2.5})}{2b_{50}} \times 100\% \tag{11}$$

where $\mathrm{a}_{2.5}$, a_{50}, and $\mathrm{a}_{97.5}$ are the 2.5th, 50th, and 97.5th percentiles of the bootstrap distribution of coefficient *a*, with analogous definitions for $b_{2.5}$, b_{50}, and $b_{97.5}$ corresponding to exponent *b*. Subsequently, for a remotely sensed river width observation (W_t) at time step t, the 1,000 bootstrap parameter sets were used to derive the corresponding discharge distribution:

$$\widehat{\mathrm{Q_t}}^{(i)} = \left(\frac{\mathrm{W_t}}{\mathrm{a}^{(i)}}\right)^{1/b^{(i)}}, i = 1, 2, ..., 1000 \tag{12}$$

The 2.5th, 50th, and 97.5th percentiles of this distribution (denoted as $\mathrm{Q}_{t,2.5}$, $\mathrm{Q}_{t,2.5}$, and $\mathrm{Q}_{t,97.5}$) were then used to define the relative discharge uncertainty at time step t:

$$\mathrm{RU}_Q(\mathrm{t}) = \frac{(\mathrm{Q}_{t,97.5} - \mathrm{Q}_{t,2.5})}{2\mathrm{Q}_{t,50}} \times 100\% \tag{13}$$

Finally, the median of $\mathrm{RU}_Q(\mathrm{t})$ across all time steps was used to characterize the overall uncertainty of discharge reconstruction at each gauge. All uncertainty metrics are dimensionless percentages, enabling direct comparison across gauges of varying magnitude and geographic regions.

3.3 Uncertainty of river discharge estimates

Although the reconstructed discharge records agree well with in situ observations, uncertainties in the parameters fitted from the AHG relationships may accumulate and propagate through the discharge retrieval process. To quantify this effect, we employed bootstrap resampling to estimate the propagation of AHG parameter uncertainty in the reconstructed discharge series across 310 gauges (Fig. 8). The median relative uncertainties of the coefficient *a* and exponent *b* (RU_a and RU_b) were 14.36% and 13.24%, respectively (Fig. 8a, b). Approximately 68% of gauges exhibited RU_a below 20%, and about 80% showed RU_b below 20%, indicating robust parameter estimation for the majority of gauges, with the half-widths of their 95% confidence intervals all less than one-fifth of the median parameter estimates. The median relative uncertainty in reconstructed discharge attributable to parameter propagation (RU_Q) was 11.89%, and approximately 83% of gauges had RU_Q below 20% (Fig. 8c), suggesting that the influence of parameter estimation errors on discharge reconstruction is generally acceptable. Fewer than 2% of gauges exhibited RU_Q exceeding 40%, all of which had fewer than 25 width–discharge paired samples, implying that limited calibration data are the primary source of elevated discharge uncertainty. Spatially, gauges with higher parameter and discharge uncertainties were mainly located in the Yangtze and Pearl River basins (Fig. 8a–c). Frequent cloud cover in these regions may reduce the availability of effective optical observations and seasonal representativeness, thereby weakening the statistical constraints on AHG parameters and amplifying discharge uncertainty.

Moreover, RU_Q showed a significant positive correlation with the exponent *b* ($R^2 = 0.40$, $p < 0.001$), but no significant correlation with coefficient *a* ($R^2 = 0.01$, $p > 0.05$) (Fig. 8d). This divergence is consistent with the AHG equation, in which uncertainty in the exponent *b* is amplified through its reciprocal exponent. The exponent *b* therefore exerts a stronger control on uncertainty propagation, supporting the necessity of filtering out gauges with physically unreasonable *b* values, as described in Section 3.1. Overall, the CDR^2 dataset exhibits relatively constrained parameter-induced uncertainty for most gauges, providing a data foundation with quantitative confidence bounds for subsequent analyses of long-term discharge change across China. However, RU_Q only characterizes the propagation of AHG parameter uncertainty under given satellite-derived river widths and does not incorporate other sources of uncertainty, including remote sensing observation errors, model structural errors, or changes in external hydrological processes.

To further reflect the importance of uncertainty characterization for the CDR^2 dataset, we have also incorporated the key uncertainty result into the Abstract and Conclusions: “Compared with existing global satellite-derived discharge products, CDR^2 increases the number of available gauges in China by at least fivefold. It achieves improved performance, with a median Kling–Gupta efficiency of 0.66 during validation, while uncertainties propagated from model parameters remain limited, with a median relative uncertainty of 11.89% in discharge estimates." (Page 1, Lines 17–21) and "Notably, 91% of gauges exhibit positive KGE values, approximately 73% achieve CC values greater than 0.7, and 79% maintain NRMSE below 20%. Uncertainty analysis further indicates that uncertainties propagated from model parameters remain limited, with a median relative uncertainty of 11.89% in discharge estimates." (Page 32, Lines 704–708)

These revisions establish a quantitative uncertainty assessment framework for CDR^2, allowing users to better understand the confidence level of the reconstructed discharge records and the influence of AHG parameter variability on discharge estimation.

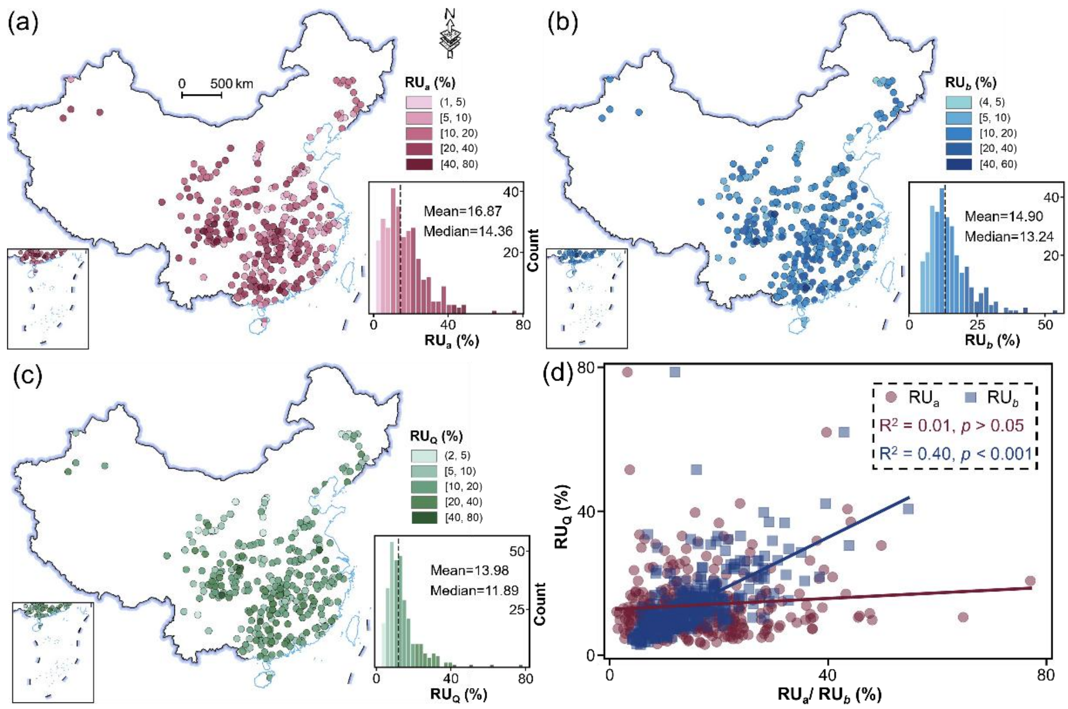

Figure 8. Uncertainty assessment of satellite-extended river discharge at 310 gauges based on the bootstrap method. (a–c) Spatial distributions of the relative uncertainties of the width–discharge power-law coefficient *a* (RU_a), exponent *b* (RU_b), and reconstructed discharge (RU_Q), respectively. The relative uncertainty is defined as the relative half-width of the 95% confidence interval, determined by the 2.5th and 97.5th percentiles of the bootstrap distribution. Insets in each panel show frequency distribution histograms, with mean and median values indicated. (d) Scatter plots and linear fits between RU_Q and RU_a/RU_b demonstrate that the uncertainty in discharge reconstruction is primarily controlled by exponent *b* during error propagation, with the

corresponding coefficients of determination (R^2) and statistical significance levels (p) indicated. (Page 21, Lines 455–464)

References

Hesterberg, T.: Bootstrap, WIREs Comput. Stat., 3, 497–526, https://doi.org/10.1002/wics.182, 2011.

Rodrigues, D. B. B., Gupta, H. V., Mendiondo, E. M., and Oliveira, P. T. S.: Assessing uncertainties in surface water security: An empirical multimodel approach, Water Resour. Res., 51, 9013–9028, https://doi.org/10.1002/2014WR016691, 2015.

Minor comments

1. Line 143: Regarding Fig. 2d, could the authors explain why CDR^2 contains substantially fewer gages in the Southwest Rivers Basin (Basin A) than the CHP dataset? Is this mainly due to the site-selection criteria adopted in this study?

Response 1: We sincerely thank the reviewer for this insightful comment. The substantially fewer gauges of the CDR^2 dataset in the Southwest Rivers Basin (Basin A) are not due to any basin-specific gauge-selection criteria adopted in this study. In fact, we directly compiled all available daily discharge observations from the National Hydrological and Rainfall Information System of the Ministry of Water Resources of China (http://xxfb.mwr.cn/), with the only processing step being the removal of spatially redundant gauges located within 100 m of each other.

The difference between CDR^2 and the CHP dataset mainly results from their different data sources and construction purposes. The CHP dataset was specifically developed for the Southwest Rivers Basin and therefore contains a relatively dense gauge network in this region. In contrast, no daily discharge records from the Southwest Rivers Basin were available from the source platform used for CDR^2 during the study period, resulting in the absence of gauges from this basin. To avoid potential misunderstanding, we have revised the corresponding description in the manuscript: "The resulting gauge network spans nearly all major river basins in China, with the Southwest Rivers Basin being the only exception due to the unavailability of discharge records from the source platform." (Page 5, Line 135 and Page 6, Line 136–137)

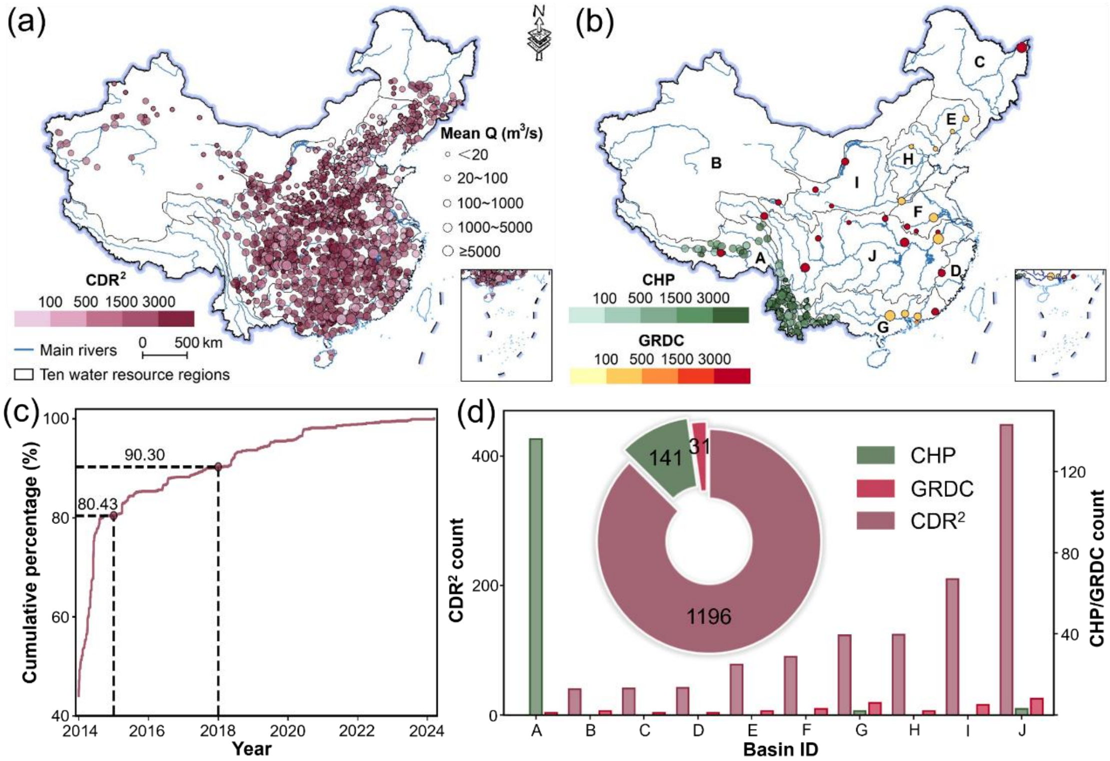

Figure 2. Spatial distribution and temporal coverage of in situ river discharge observations across China (2014–2024). (**a**) Spatial distribution of the 1,196 gauges included in the China Daily River Discharge Records (CDR^2) dataset, with symbol size representing mean discharge (Q, m³/s). (**b**) Spatial distribution of river gauges available from the Chinese Hydrology Project (CHP) and the Global Runoff Data Centre (GRDC) within China. The labeled regions (A–J) denote the ten major basins: Southwest Rivers (A), Northwest Rivers (B), Songhua River (C), Southeast Rivers (D), Liao River (E), Hai River (F), Pearl River (G), Huai River (H), Yellow River (I), and Yangtze River (J). (**c**) Cumulative probability distribution of gauge observation start years in the CDR^2, indicating that over 80% of gauges contain records beginning in or before 2015, while more than 90% begin in or before 2018. (**d**) Comparison of the number of gauges included in the CDR^2, CHP, and GRDC datasets across China. (Page 7, Lines 154–165)

2. Lines 206–207: Since both Landsat and Sentinel were used, did the authors perform any cross-sensor bias correction or consistency assessment before combining observations from the two sensor systems?

Response 2: We sincerely thank the reviewer for this valuable comment. In this study, we did not perform an additional cross-sensor bias correction before combining observations from Landsat and Sentinel-2. This is because river width was extracted independently from the Level-2 surface reflectance products of each sensor, which have already undergone standard radiometric calibration, atmospheric correction, and geometric correction to ensure a high level of consistency in radiometric properties and geolocation accuracy across sensor systems (Claverie et al., 2018). Moreover, the river width retrieval was based on MNDWI-derived water masks rather than direct spectral value fusion, thereby reducing the potential influence of inter-sensor radiometric differences on the final width estimates.

It should also be clarified that the median compositing procedure was applied only to multiple images acquired by the same sensor on the same day, rather than across different sensors. Therefore, no pixel-level co-registration or cross-sensor image fusion was involved in the processing workflow. When daily river width estimates from different sensors were available for the same date, the independently derived width measurements were averaged to obtain a single daily river width value. To avoid potential misunderstanding, we have revised the corresponding description in the manuscript: "For each sensor, images acquired on the same day were combined via median compositing to generate daily surface reflectance mosaics with minimized cloud effect. When multiple daily width estimates from different sensors were available for the same date, they were averaged to produce a single daily width value." (Page 11, Lines 226–230)

References

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, Remote Sens. Environ., 219, 145–161, https://doi.org/10.1016/j.rse.2018.09.002, 2018.

3. Lines 250–251: The site-selection procedure could be described more explicitly. It would be helpful to clearly state here that only 310 of the 401 candidate gages were ultimately retained because they satisfied the valid width-discharge pairing criteria.

Response 3: Thanks for this comment. We agree that the gauge-selection procedure should be described more explicitly. The criterion of at least ten valid width–discharge pairs (N≥10) represents only the initial screening step to ensure sufficient observations for width–discharge relationship assessment, rather than the final selection criterion. Specifically, among the 401 candidate gauges, 358 gauges satisfied the minimum requirement of N≥10 and were retained for further screening. Additional screening procedures were then applied to remove gauges with weak or physically unrealistic width–discharge relationships, including the exclusion of gauges with low Spearman correlation coefficients ($\rho < 0.25$) and physically implausible hydraulic geometry exponents ($b < 0$ or $b > 0.59$). After these quality-control procedures, a total of 310 gauges were finally retained for discharge reconstruction.

Although the final number of retained gauges (310) was previously reported in Section 3.1, we agree that explicitly stating this information in the Methods section improves the transparency of the gauge-selection procedure. Therefore, we have added the following description after the screening criteria: "Gauges with fitted exponents outside this range were removed to avoid incorporating unstable or physically implausible hydraulic relationships into the discharge reconstruction framework. After applying all screening criteria, a total of 310 gauges were retained for discharge reconstruction." (Page 13, Lines 279–282) In addition, the description of the initial screening step has been revised as follows: "Accordingly, the instantaneous river widths derived from satellite imagery (Section 2.2) were paired with concurrent in situ daily discharge observations at each gauge (Section 2.1). Among the 401 candidate gauges, only those with at least ten valid width–discharge pairs ($N \geq 10$) were retained, resulting in 358 gauges for subsequent width–discharge relationship assessment." (Page 12, Lines 263–267)

4. Line 349: "river dynamism" or "river dynamics"?

Response 4: Thanks for this comment. We agree that "river dynamics" is a more commonly used term to describe temporal variations in river systems. Although "river dynamism" was intended to represent the degree of river width variability quantified by the coefficient of variation of river width (CV_w), we have replaced it with "river dynamics" for clarity and consistency with the terminology commonly used in hydrological studies.

Accordingly, we have revised the corresponding description in the manuscript: "To further investigate how reconstruction performance depends on river dynamics, the selected gauges were stratified into five equal-frequency groups (G1–G5) according to the coefficient of variation of river width (CV_w), with group mean values ranging from 6% to 23%." (Page 18, Lines 393–395)

5. Line 383: The font of "with Sen's slope estimator" differs from the surrounding text. Similar formatting inconsistencies also appear around Lines 393 and 410.

Response 5: Thanks for pointing out this formatting issue. We have re-examined the formatting of the relevant sections and ensured that the font style and formatting of "with Sen's slope estimator" and the corresponding text around Lines 393 and 410 are consistent with the surrounding text in the revised manuscript.

6. Lines 412–413: The manuscript refers to reconstructed "daily discharge" from 1990-2024. However, the reconstructed time series are not temporally continuous on a daily basis. I suggest clarifying this point in the manuscript to avoid potential misunderstanding.

Response 6: We agree that the term "daily discharge" may potentially imply a temporally continuous daily record. In this study, the reconstructed discharge estimates are derived from satellite-observed river widths from Landsat and Sentinel-2 imagery. Therefore, each estimate corresponds to a specific date (daily temporal resolution), but the time series are not continuously available for every day due to satellite observation frequency and data availability. To avoid potential misunderstanding, we have revised the corresponding description in the manuscript: "Using the satellite-extended discharge estimates with daily temporal resolution from 310 gauges spanning 1990–2024, we quantified the long-term trends and observational characteristics of rivers across China." (Page 22, Lines 466–468)

In addition, we have also checked the manuscript and revised other potentially ambiguous uses of "daily discharge" and "continuous" to ensure that the reconstructed products are clearly described as date-specific discharge estimates rather than continuous daily records.

7. Line 439: I recommend adding a dedicated subsection discussing the limitations of the dataset and its appropriate scope of application.

Response 7: Thanks for this comment. We agree that the assumptions, limitations, and appropriate application scope of the CDR^2 dataset should be explicitly discussed. Accordingly, we have added a dedicated subsection to discuss the major limitations and potential uncertainties associated with the dataset, including the temporal stationarity assumption of width–discharge relationships, optical satellite observation constraints, and future improvement directions.

The newly added discussion is presented in Section 4.3 "Limitations, Uncertainties, and Future Perspectives" of the revised manuscript (Page 29, Line 618), which also addresses the related concerns raised in Major Comment 1.

8. Line 479 (Section 4.1 and Fig. 9): The comparison with existing products should be interpreted with greater caution. CDR^2 is derived from a screened subset of Chinese gages that satisfy the width-discharge relationship requirements, whereas the global datasets include a much broader range of river types, including hydraulically challenging sites. Therefore, statements such as "substantially outperforming" existing products may not represent a fully equivalent comparison and could be moderated accordingly.

Response 8: Thanks for this helpful comment. We agree that comparisons among satellite-based discharge datasets should be interpreted with caution, as different products may involve different gauge selection strategies and represent distinct ranges of river environments. In particular, CDR^2 was developed using a screened subset of Chinese gauges with stable width–discharge relationships, whereas global products aim to maximize spatial coverage and may include a broader range of hydraulically challenging gauges.

To avoid overinterpretation, we have moderated the corresponding statements in the manuscript: "Notably, CDR^2 achieves a median KGE of 0.66, exceeding the reported global median range of 0.33–0.48 among existing products (Table 2), although these comparisons should be interpreted considering differences in gauge selection criteria and river characteristics." (Page 25, Lines 545–548) and "Overall, CDR^2 provides an expanded and higher-resolution characterization of river discharge across China, a hydrologically critical yet globally underrepresented region." (Page 25, Lines 562–563)

In addition, we have revised related statements in the Abstract and Conclusions to avoid implying an absolute performance advantage of CDR^2 over existing global products. The corresponding revisions are as follows: "Compared with existing global satellite-derived discharge products, CDR^2 increases the number of available gauges in China by at least fivefold. It achieves improved performance, with a median Kling–Gupta efficiency of 0.66 during validation, while uncertainties propagated from model parameters remain limited, with a median relative uncertainty of 11.89% in discharge estimates." (Page 1, Lines 17–21) and "Compared with existing global satellite-based discharge datasets, CDR^2 substantially expands the spatial coverage of daily-scale discharge estimates across China by more than fivefold and demonstrates improved accuracy." (Page 32, Lines 709–711).