Response to Reviewer #1,

We thank reviewer for the detailed reviews, and we made all suggested corrections. In this response, the reviewer's comments are in black standard font. Our response is in standard blue font and the modifications to the manuscript are in blue bold font.

Major comments:

1) In Sect. 3.2, the UAV coverage appears to be limited to a small terminal portion of a single glacier within the study area. Please clarify to what extent this training domain is representative in both space and time (seasonal/full-year conditions), and justify the applicability of UAV calibrated parameters/weights when extrapolated to the entire region and the full annual cycle.

Thank you for this comment. We have clarified in Sect. 3.2 that although the UAV coverage is limited to ~30 km² near the Yanong Glacier terminus, the UAV acquisitions span (i) seasonal conditions from the ablation season to the early accumulation period (June–November 2023), (ii) a clear spatial velocity gradient from the fast-flowing trunk to slower marginal areas, and (iii) diverse surface facies including debris-covered ice, clean ice, and crevassed zones. These characteristics provide heterogeneous yet relevant conditions, giving the UAV subset potential transferability for evaluating the performance of Landsat, Sentinel-1, and Sentinel-2 velocity retrievals and for calibrating fusion weights used in this study.

"In addition, we acquired UAV photogrammetry for reference and evaluation using a DJI Matrice 300 RTK equipped with a M6 Pro (M6P) metric mapping camera. Six orthomosaics were collected in June-November 2023, forming five image pairs. Although coverage is limited to ~30 km² near the Yanong Glacier terminus, **the UAV surveys span heterogeneous conditions: (i) the period covers the ablation season through the transition to the early accumulation period, (ii) the mapped area includes both fast-flowing trunk regions and slower glacier margins, and (iii) surface types include debris-covered ice, clean ice, and crevassed areas. Therefore, this UAV subset has potential transferability in space and time for benchmarking the performance of Landsat, Sentinel-1, and Sentinel-2 glacier velocity retrievals and for calibrating fusion weights in the regional fusion**."

2) Although the manuscript states that velocities from the three data sources are harmonized to a 30 m spatial resolution, the underlying native grids of the different sensors/products are not necessarily co-registered. For pixel-wise WLS fitting and subsequent fusion to be valid and reproducible, the pre-fusion co-registration/harmonization steps should be described explicitly, including the choice of the reference

grid, reprojection and resampling methods (and interpolation scheme), grid alignment strategy, NoData/mask propagation rules, and whether these operations are performed prior to the WLS fitting.

Thank you for this helpful suggestion. We agree that explicit description of the pre-fusion grid harmonization is necessary for ensuring the validity and reproducibility of the pixel-wise WLS fitting and subsequent fusion. We have therefore added a brief statement in Sect. 3.3 clarifying (i) the reference grid used for harmonization, (ii) the resampling method adopted for co-registration, and (iii) that these operations are performed prior to the WLS fitting.

"To support pixel-wise WLS fitting and the subsequent fusion, all velocity products were harmonized to a common 30 m reference grid prior to WLS. We used the Sentinel-2 COSI-Corr velocity image as the reference geometry, and co-registered the Landsat-, Sentinel-1-, and UAV-derived velocity image to this grid using nearest-neighbour resampling, so that the original velocity values are preserved without interpolation smoothing."

3) For the WLS-based fusion weights, please specify the exact objective function, the interpretation ofthe weights within the WLS formulation, and any constraints imposed during optimization. In particular, clarify whether non-negativity and/or a normalization constraint are enforced. If constraints are used, briefly describe the corresponding solution strategy/implementation to ensure reproducibility.

Thank you for this comment. We have clarified the WLS formulation by explicitly stating the constraints imposed on the fusion weights. In our implementation, the weights are constrained to be non-negative and to sum to one.

"Let $V_{\text{uav}}(i,j)$ denote the UAV-derived velocity at pixel (i,j), and $L(i,j)$,$S_1(i,j)$,$S_2(i,j)$ denote the Landsat, Sentinel-1, and Sentinel-2 velocities, with corresponding fusion weights w_L,w_{S1},w_{S2}. The WLS objective is:

$$f(w) = \sum_{i=1,j=1}^{n} \left[V_{\text{uav}}(i,j) - \left(w_L \cdot L(i,j) + w_{S1} \cdot S_1(i,j) + w_{S2} \cdot S_2(i,j)\right)\right]^2 \quad (1)$$

After defining the objective in Eq. (1), we estimate the fusion weights by minimizing $f(w)$ subject to the constraints $w_L \geq 0$, $w_{S1} \geq 0$, $w_{S2} \geq 0$ and $w_L + w_{S1} + w_{S2} = 1$. This simplex constraint ensures that the weights are directly interpretable as non-negative mixing fractions. The resulting optimal weight triplet is then applied to fuse the remaining monthly velocity maps."

4) In Sect. 3.4, $(i,j)$is defined as the pixel location in the image. However, in Eq. (1) the weight estimation would normally iterate over all pixels of the 2-D grid, whereas the summation appears to run only over

i. Sentinel-1–derived velocities are used as the information source for filling gaps in the fused product (via the enhancement-coefficient field), but the rationale for selecting Sentinel-1 as the gap-filling reference is not sufficiently articulated. Also, the first two paragraph of Section 3.4 is a bit repetitive with Section 3.1.

Thank you for these helpful comments. We have revised Sect. 3.4 accordingly.

(1) We adjusted Eq. (1) and its notation to clearly indicate that the WLS objective is summed over the set of overlapping valid pixels on the 2-D grid.

(2) Before introducing the enhancement-coefficient infilling, we added a short rationale explaining why Sentinel-1 is used to guide gap repair: optical (Landsat/Sentinel-2) velocities are often missing due to cloud/snow/illumination limitations, whereas Sentinel-1 provides a more complete auxiliary field, helping preserve glacier-motion patterns while improving continuity.

(3) We rewrote the first two paragraphs of Sect. 3.4 to reduce redundancy with Sect. 3.1 and to focus the section on the weight-estimation formulation and subsequent steps.

"In this section, we describe how fusion weights are estimated using co-temporal UAV-derived velocities. Monthly velocity maps from Landsat, Sentinel-1, and Sentinel-2 are generated independently, and UAV orthomosaics are processed to provide a high-accuracy reference for the same periods. A WLS formulation is then used to estimate the optimal weights for the three satellite products, which are applied in the subsequent fusion.

Let $V_{\text{uav}}(i,j)$ denote the UAV-derived velocity at pixel (i,j), and $L(i,j)$,$S_1(i,j)$,$S_2(i,j)$ denote the Landsat, Sentinel-1, and Sentinel-2 velocities, with corresponding fusion weights w_L,w_{S1},w_{S2}. The WLS objective is:

$$f(w) = \sum_{i=1,j=1}^{n} \left[V_{\text{uav}}(i,j) - \left(w_L \cdot L(i,j) + w_{S1} \cdot S_1(i,j) + w_{S2} \cdot S_2(i,j)\right)\right]^2 \tag{1}$$

After defining the objective in Eq. (1), we estimate the fusion weights by minimizing $f(w)$ subject to the constraints $w_L \geq 0$, $w_{S1} \geq 0$, $w_{S2} \geq 0$ and $w_L + w_{S1} + w_{S2} = 1$. This simplex constraint ensures that the weights are directly interpretable as non-negative mixing fractions. The resulting optimal weight triplet is then applied to fuse the remaining monthly velocity maps.

In the fusion stage, to address pixels with missing values (NoData), we introduce a binary mask $M_k(i,j)$ to locally renormalize the weights, so that the weighted average at (i,j) is computed only over sources that are valid there (Landsat or Sentinel-2). The preliminary fused velocity is:

$$F(i,j) = \frac{w_{S1} \cdot S_1(i,j) \cdot M_{S1}(i,j) + w_{S2} \cdot S_2(i,j) \cdot M_{S2}(i,j) + w_L \cdot L(i,j) \cdot M_L(i,j)}{w_{S1} \cdot M_{S1}(i,j) + w_{S2} \cdot M_{S2}(i,j) + w_L \cdot M_L(i,j)} \tag{2}$$

Although weighted fusion can effectively integrate multi-source information, some areas may still contain NoData. **To further fill these gaps and enhance data continuity, this study introduces a sliding-window enhancement-coefficient infilling method. Because Landsat- and Sentinel-2–derived velocity maps often exhibit large gaps in this cloud- and snow-prone region and simple interpolation can be unreliable, we use Sentinel-1 SAR velocities—less sensitive to clouds and illumination—as a more complete auxiliary field to guide gap repair. Specifically, Sentinel-1 is used to infer the local spatial variation of the fused field and fill NoData accordingly: for each pixel (i,j), we consider a 10×10 sliding window $\Omega(i,j)$ (stride = 1 pixel) and estimate a local enhancement coefficient $\hat{a}(i,j)$ that describes the response of the fused value to Sentinel-1,:**

$$\hat{a}(i,j) = \frac{\sum_{(u,v)\in\Omega(i,j)} F(\mathrm{u},\mathrm{v}) \cdot S_1(\mathrm{u},\mathrm{v})}{\sum_{(\mathrm{u},\mathrm{v})\in\Omega(i,j)} S_1(\mathrm{u},\mathrm{v})^2} \quad (3)$$

”

5) In Eqs. (3) and (4), the enhancement coefficient is denoted as aΩ, which naturally reads as a single constant/parameter associated with a domain Ω. However, based on the stated computation and definition (window-based estimation followed by Gaussian smoothing to form a spatial field), the enhancement coefficient should vary spatially and thus be a gridded quantity.

Thank you for pointing this out. We agree that the enhancement coefficient is a spatially varying gridded field rather than a single constant associated with a window domain. We have therefore revised the notation in Sect. 3.4.

“**Although weighted fusion can effectively integrate multi-source information, some areas may still contain NoData. To further fill these gaps and enhance data continuity, this study introduces a sliding-window enhancement-coefficient infilling method. Because Landsat- and Sentinel-2–derived velocity maps often exhibit large gaps in this cloud- and snow-prone region and simple interpolation can be unreliable, we use Sentinel-1 SAR velocities—less sensitive to clouds and illumination—as a more complete auxiliary field to guide gap repair. Specifically, Sentinel-1 is used to infer the local spatial variation of the fused field and fill NoData accordingly: for each pixel (i,j), we consider a 10×10 sliding window $\Omega(i,j)$ (stride = 1 pixel) and estimate a local enhancement coefficient $\hat{a}(i,j)$ that describes the response of the fused value to Sentinel-1,:**

$$\hat{a}(i,j) = \frac{\sum_{(u,v)\in\Omega(i,j)} F(\mathrm{u},\mathrm{v}) \cdot S_1(\mathrm{u},\mathrm{v})}{\sum_{(\mathrm{u},\mathrm{v})\in\Omega(i,j)} S_1(\mathrm{u},\mathrm{v})^2} \quad (3)$$

Only pixels where both $F(u,v)$ and $S_1(u,v)$ are valid are included in the summation. Next, **$\hat{a}(i,j)$ is smoothed with a Gaussian filter to form a spatially varying enhancement-coefficient field $a(i,j)$ over the entire image.** Finally, for missing pixels (i,j) in the fused map, infilling is performed as

$$F_{filled}(i,j) = \begin{cases} F(i,j), if\ F(i,j)\ is\ valid \\ a(i,j) \cdot S_1(i,j), if\ F(i,j) = Nodata \end{cases} \quad (4)$$

Through this enhancement-based infilling, the spatial completeness of the fused image is markedly improved and gaps are reasonably reconstructed, providing more stable and continuous data support for subsequent glacier-change analyses and time-series modeling."

6) In Sect. 4.1, the manuscript reports an "~70% / ~30% improvement" in valid pixels, but it is unclear whether this refers to an absolute increase (percentage points) or a relative increase with respect to a baseline. To avoid ambiguity, please provide the explicit definition/formula used to compute the improvement.

Thank you for this helpful comment. We agree that the phrase "~70% / ~30% improvement" is ambiguous. To avoid confusion, we have revised Sect. 4.1 to report the improvement in percentage points (pp) instead of relative percentages.

"Figure 3 presents the monthly fraction of valid pixels in velocity images from 2015–2024 for the two representative glaciers, Xirinongpu and Yanong. The results show that, in most months, the fused product attains a substantially higher valid-pixel ratio than either single-source dataset. **Over Xirinongpu, the mean valid-pixel ratios for Landsat and Sentinel-2 are 57.4% and 54.7%, respectively, whereas the fused result reaches 97.9%, corresponding to absolute increases of 40.5 and 43.2 percentage points, respectively. Over Yanong, Landsat and Sentinel-2 achieve 74.7% and 77.1%, while the fused image further increases to 99.2%, corresponding to absolute increases of 24.5 and 22.1 percentage points, respectively.** These findings indicate that the proposed fusion method effectively overcomes the limitations of individual sources and markedly enhances data availability and continuity."

7) In Sect. 3.5, "image smoothness" is defined as the pixelwise standard deviation within the mask for each month ("…we compute the pixelwise standard deviation…"). However, later (Fig. 5 and associated text) the manuscript refers to using variance to assess image smoothness.

Thank you for pointing out this inconsistency. We agree that the terminology should be consistent. In our analysis of image smoothness, the metric used is variance, and the "standard deviation" wording in Sect. 3.5 was imprecise. We have revised Sect. 3.5 to consistently use variance when describing the image-smoothness assessment.

"(3) Image smoothness: For the three velocity products and for GoLIVE and ITS_LIVE, we compute the pixelwise variance within the mask for each month. Changes in **variance** indicate noise suppression and smoothing, enabling a comprehensive assessment of fusion-driven gains in coverage, temporal continuity, and smoothness.”

8) In Sect. 4.4, the manuscript states that the observed bimodal intra-annual velocity pattern “accords with” the subglacial hydrology-dynamics evolution framework for maritime glaciers. While this interpretation is plausible, the current discussion remains largely qualitative and lacks a clear evidential link to the data presented in this study.

Thank you for this important comment. We agree that the original wording could be interpreted as an overly strong attribution. In the revised manuscript, we have moderated the interpretation and reframed it as a plausible process-based explanation that is broadly consistent with the seasonal pattern observed in our data and with previous studies. We also added a citation to Nanni et al. (2023) to support that similar seasonal acceleration behaviour, including autumn acceleration, has been reported in mountain glaciers in other regions.

“To characterize intra-annual variability in the Kangri Karpo region, we extracted monthly velocities from the 2015–2024 fused product and averaged each calendar month across years to derive a climatological annual cycle (Fig. 11). **The resulting series shows a clear seasonal signal: velocities are relatively low in January-March, increase from March onward, reach a first peak in May, remain comparatively stable during summer, rise again to a second peak in October, and then decline. This indicates a bimodal intra-annual velocity pattern in the study region.**

This bimodal pattern is broadly consistent with process-based interpretations proposed for temperate/maritime glaciers, in which seasonal changes in subglacial drainage efficiency modulate basal sliding. In particular, early-season meltwater and rainfall inputs may enhance sliding under a relatively inefficient distributed drainage system, whereas subsequent drainage-system evolution toward more efficient channelized flow may reduce mean basal water pressure and limit velocities later in the melt season; partial re-closure or renewed pressurization may contribute to a later-season acceleration. Similar seasonal acceleration behaviour, including autumn acceleration, has also been reported for mountain glaciers in other regions (e.g., Nanni et al., 2023). However, we emphasize that the present study does not include direct observations of subglacial hydrology, and therefore this interpretation should be viewed as plausible rather than diagnostic.

The standard-deviation envelope in Fig. 11 further suggests month-dependent interannual variability. Standard deviations are larger in April-May, when seasonal acceleration initiates, indicating stronger year-to-year differences in the timing and magnitude of acceleration. By contrast, variability is comparatively smaller from June to the following March. Taken together, these results indicate a clear seasonal velocity response in the Kangri Karpo region, with a

recurrent bimodal intra-annual structure."

Nanni, U., Scherler, D., Ayoub, F., Millan, R., Herman, F., and Avouac, J.-P.: Climatic control on seasonal variations in mountain glacier surface velocity, The Cryosphere, 17, 1567–1583, https://doi.org/10.5194/tc-17-1567-2023, 2023.

9) Based on the velocity formulation described in Sect. 3.3, the primary unit of the derived glacier surface velocity should be m d-1. While, in several places (including Sect. 4.3) the manuscript reports velocities in m yr-1 or m/year. Please standardize the unit system throughout the manuscript.

Thank you for this helpful comment. We agree that the velocity unit should be standardized throughout the manuscript. The primary unit directly derived from the velocity formulation in Sect. 3.3 is m d^{-1}, and in the revised manuscript we have unified the velocity unit accordingly. Specifically, we revised the text, figure labels, and related descriptions (including Sect. 4.3 and associated figures) to consistently report glacier surface velocity in m d^{-1} and removed the previous mixed usage of m yr^{-1} / m/year.

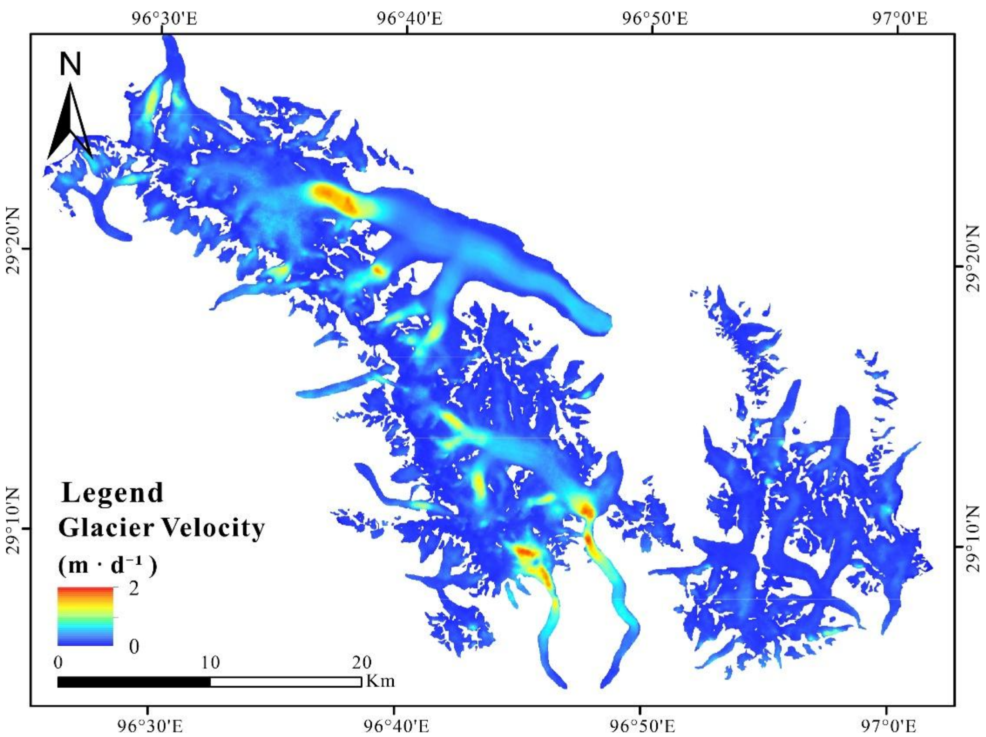


Figure 7: Mean glacier surface velocity in the Kangri Karpo region (2015–2024).

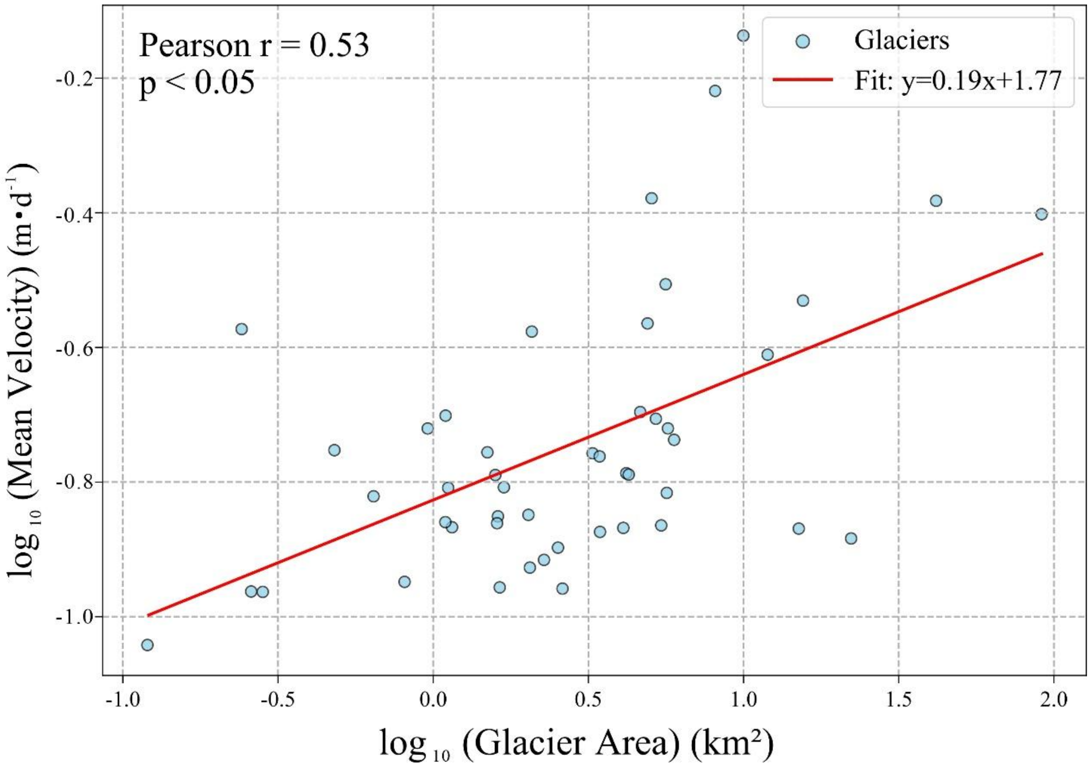


Figure 8: Correlation between glacier area and velocity. Blue points denote glaciers in the study area; the red solid line is the fitted regression; axes are logarithmic.

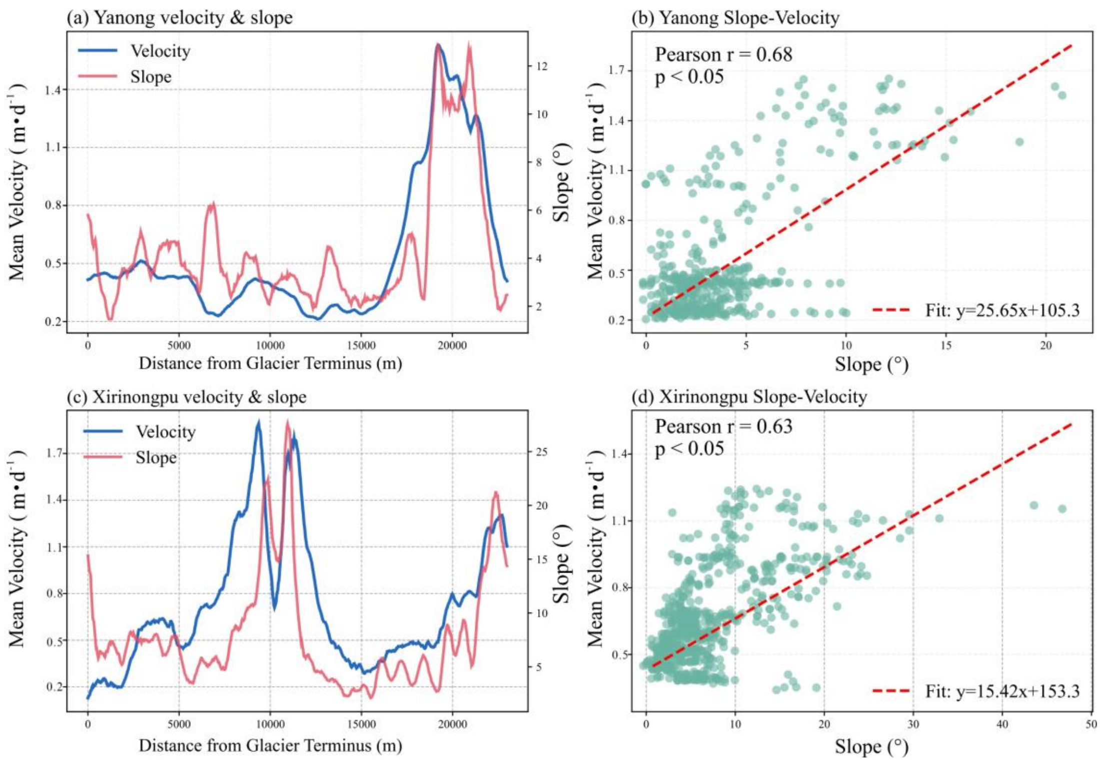


Figure 9: (a), (c) Variations of glacier surface velocity and slope along the centerlines of the Yanong and Xirinongpu glaciers. Blue curve: velocity; red curve: slope. (b), (d) Scatterplots of velocity versus slope for all pixels within the Yanong and Xirinongpu glacier masks; red dashed line: linear fit of the slope-velocity relationship.

Specific comments

Title: "High resolution" could mean high spatial resolution and temporal resolution. As you indicated "monthly glacier surface velocity", how about using "30-meter" to replace "highresolution"?

Thank you for this helpful suggestion. We agree that the term "high resolution" may be ambiguous because it can refer to either spatial or temporal resolution. To make the title more specific, we have revised it by replacing "high-resolution" with "30-meter".

Revised title

30-meter monthly glacier surface velocity mapping in the Kangri Karpo region (2015–2024) using multi-source remote sensing data fusion

Line 17: This sentence confused me the fusion method is for whether high spatial resolution or high spatial and temporal resolution?

Thank you for this comment. We agree that the original wording was ambiguous and could be interpreted as referring to both spatial and temporal resolution. In the revised manuscript, we clarified this point by replacing "high-resolution" with "high-spatial-resolution" in the relevant sentence (Line 17) and throughout the manuscript where appropriate.

Line 25: It would be better to state as "with similar area and slope, south-facing glaciers are slightly faster than northern ones".

Modified accordingly.

"…,within individual glaciers, **velocity responds more strongly to slope; and, with similar area and slope, south-facing glaciers are slightly faster than north-facing ones.**"

Line 96: Please add detail precipitation of this area to illustrate this area is "among the wettest sectors". You also mentioned in Line 99 "high annual precipitation and humidity", which may be repetition in meaning.

Thank you for this comment. We agree that the expression "among the Plateau's wettest sectors" would be stronger if supported by explicit precipitation values. As the cited reference (Wu et al., 2021) provides a qualitative description rather than site-specific precipitation numbers for our study area, we avoided introducing unsupported quantitative values. Instead, we revised the sentence to a more cautious qualitative wording ("humid, monsoon-influenced sector") and removed the potentially repetitive phrasing elsewhere in the manuscript. And by replacing the repeated climate description with a more specific statement on monsoon influence, seasonal hydroclimatic variability, and frequent cloud cover.

"Located at the eastern terminus of the Nyainqêntanglha Range on the southeastern Qinghai–Tibet Plateau, **the Kangri Karpo region is a relatively wet sector of the Plateau** (Wu et al., 2021)."

"Climatically, the area lies within the monsoonal temperate glacier region and is strongly influenced by the Indian monsoon (Yang et al., 2008), **with pronounced seasonal hydroclimatic variability and frequent cloud cover.**"

Figure 1: "KM" or "km", but not "Km".

Modified accordingly.

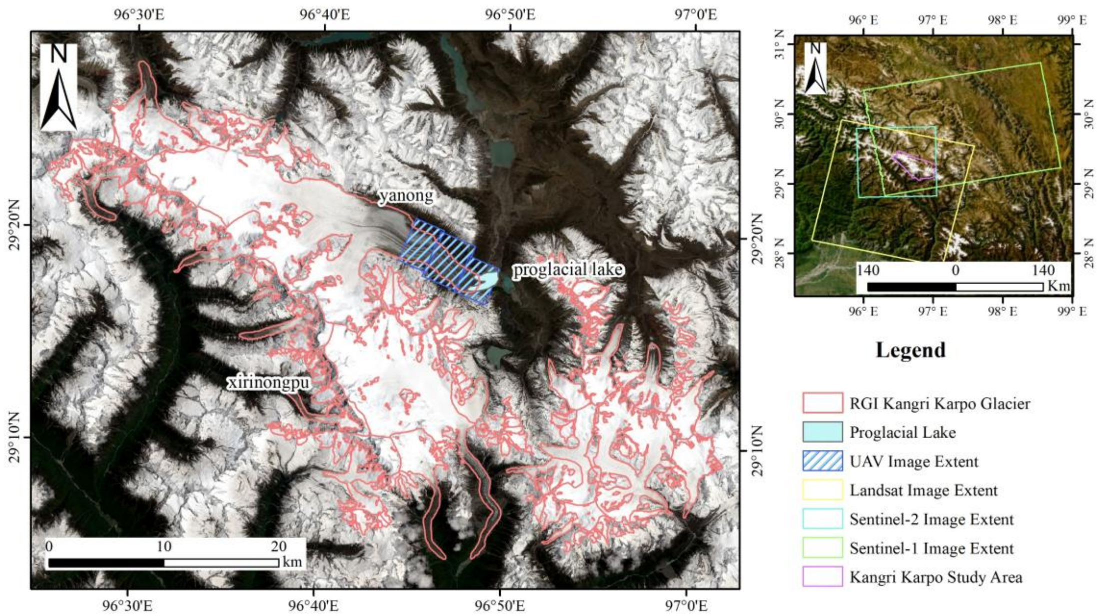
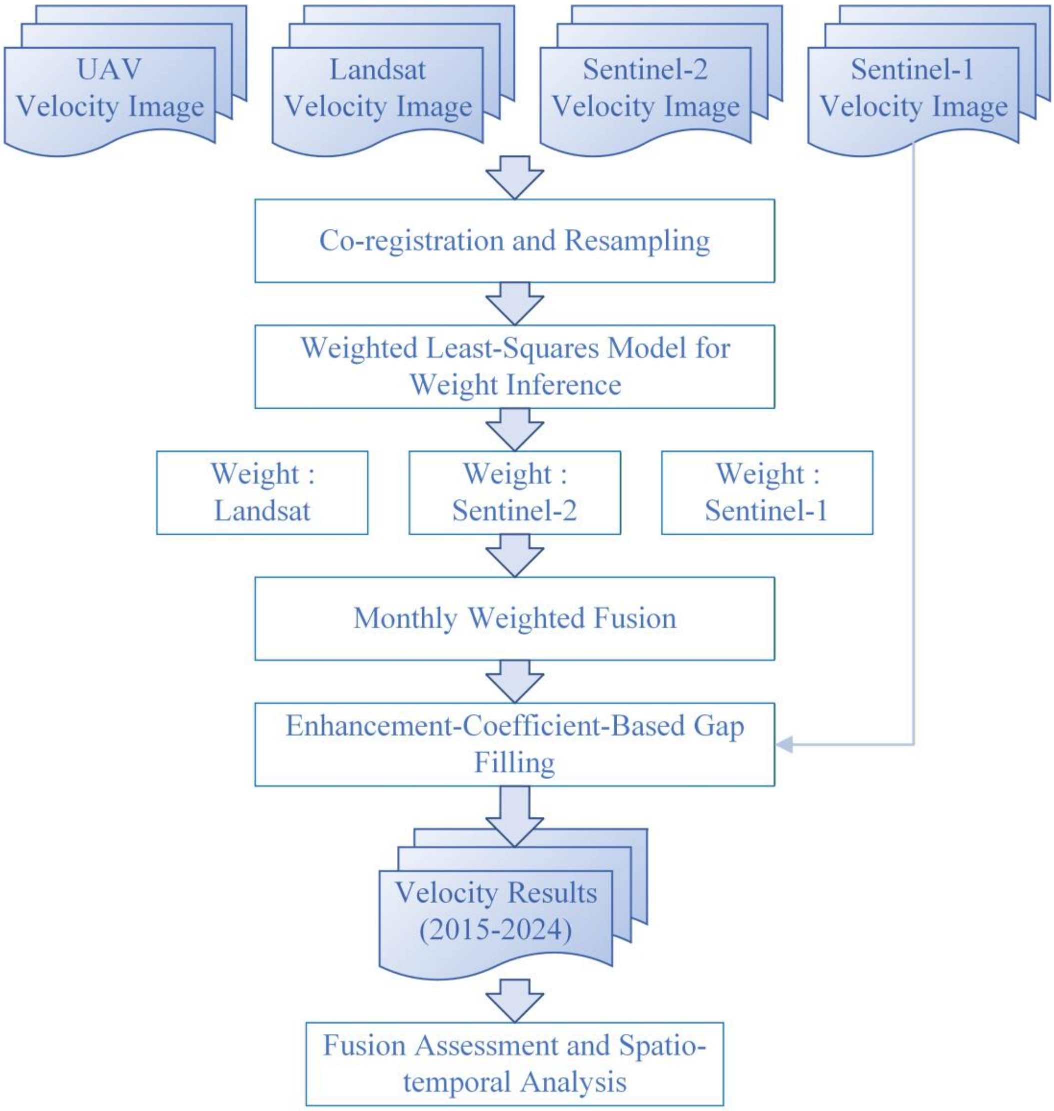


Figure 2: "Weight" not "Weigth". Also, the use of blue and white in the figure is a bit misleading. UAV velocity image and Landsat velocity image are in white, which means they are processing and analysis steps as you indicate in the caption. What are the input datasets of them? Why the final outputs of the velocity results (2015-2024) are not in blue?

Modified accordingly.



Line 159: Please provide the detail acquisition time of the six UAV images, rather than listing a time span of June to November in 2023.

Thank you for this suggestion. We have revised the manuscript to report the exact acquisition dates of the six UAV orthomosaics (8 June, 4 July, 10 August, 1 September, 3 October, and 20 November 2023), instead of only giving the June–November 2023 time span.

"In addition, we acquired UAV photogrammetry for reference and evaluation using a DJI Matrice 300 RTK equipped with an M6 Pro (M6P) metric mapping camera. **Six orthomosaics were collected on 8 June, 4 July, 10 August, 1 September, 3 October, and 20 November 2023, forming five image pairs.**"

Line 166: "32×32 pixels" not "32*32 pixels". The results of "2 pixels for Landsat OLI" and "3 pixels for Sentinel-2 MSI" have the same spatial resolution. Why did you not set the sept size as 1 pixel for both Landsat and Sentinel-2, and resample the displacement result to 30m afterwards?

Thank you for this helpful comment. We have corrected the formatting from “32*32 pixels” to “32×32 pixels.”
Regarding the step-size setting, we intentionally used 2 pixels for Landsat OLI and 3 pixels for Sentinel-2 MSI so that the displacement output spacing is directly matched to the target ~30 m resolution. Using a 1-pixel step for both sensors and then resampling to 30 m would substantially increase computational cost and processing time. In addition, compared with first generating denser 10/15 m displacement fields and then resampling them to 30 m, directly retrieving results at ~30 m helps avoid an extra resampling/interpolation step and reduces potential smoothing effects, which is preferable for subsequent fusion.

Line 170: I suggest to introduce more about the stabilization of time series in this paragraph, including the description of the use of α-trimmed mean filter, the threshold standard of selecting 0.33, and the use of “*”, which may not be a formal use in the main text.

Thank you for this suggestion. We have clarified the rationale for setting α = 0.33 in the revised manuscript. This parameter was selected empirically based on the observed characteristics of glacier velocity time series in our study region. During the melt season (typically June-September), optical remote-sensing observations (Landsat and Sentinel-2) are frequently affected by cloud/fog contamination (and seasonal snow/cloud-shadow effects), which can introduce anomalously high or low mismatches in the derived monthly velocities. Such outliers may strongly contaminate the subsequent fusion if not suppressed. Based on observations across most years in the study region, the central portion of the annual monthly series provides a more stable reference, and we therefore adopted a symmetric trim with α = 0.33 (i.e., trimming the lower and upper thirds and averaging the middle ~34%) as a practical robust setting.
“For Landsat and Sentinel-2 imagery, velocities were derived with COSI-Corr using a frequency-domain cross-correlation algorithm. The search window was 32×32 pixels; the step size was 2 pixels for Landsat OLI and 3 pixels for Sentinel-2 MSI; the correlation threshold was 0.95. East-west (E-W) and north-south (N-S) displacements were combined to form total displacement, which was then divided by the pairwise time baseline to obtain monthly glacier surface speeds at 30 m resolution.
To improve reliability, we first removed mismatches with speeds > 3 m d^{-1} and low-confidence pixels with SNR < 0.9. For the optical-image-derived velocities (Landsat and Sentinel-2), additional stabilization is necessary because optical matching is more susceptible to cloud/cloud-shadow

contamination, seasonal snow cover, illumination changes, and surface-texture variations, which can produce anomalously large or small mismatches in some months. We therefore applied an intra-annual α-trimmed mean filter to stabilize each pixel's monthly time series and reduce the influence of residual outliers. Specifically, for each pixel, the 12 monthly speeds within a given year were ranked, and a symmetric trim with α = 0.33 was applied, i.e., approximately the lower 33% and upper 33% of values were discarded and the central ~34% was averaged to obtain a year-representative reference value. We selected α = 0.33 as a practical robust setting for a 12-month annual series, because it suppresses extreme values while retaining a central subset for stable estimation. Monthly values were then compared with this reference on a pixel-by-pixel basis; values greater than 1.5× the reference or smaller than 0.5× the reference were flagged as outliers and removed."

Line 311: "0.10, lower than" not "0.10-lower than"

Modified accordingly.

Line 321: "product" not "produc"

Modified accordingly.

Line 396: use en dash. "January–March" not "January-March"

Modified accordingly.