

Referee comment on essd-2026-35

Manuscript: Sommer, C., Groos, A. R., Fürst, J. J., and Braun, M.: *Transient snow line altitudes of glaciers in the European Alps from multi-mission remote sensing data (2000–2025)*, Earth Syst. Sci. Data Discuss. [preprint], <https://doi.org/10.5194/essd-2026-35>, 2026.

Recommendation: reconsider after major revisions.

Competing interests. I work on snowline retrieval from Landsat time series in another mountain range and have a manuscript in preparation on that topic. I declare this as a potential competing interest. Nothing in this comment asks the authors to cite unpublished work of mine.

Note on this review. In addition to the manuscript, I downloaded and analysed the published data product (TABLES_transient_snow_line_altitudes_of_glaciers_europ_alps.zip: 409 per-glacier tables, 355,752 observations, plus the three end-of-summer tables), the README.pdf of the Zenodo record, and the full source code of the snowy-glaciers repository. Several comments below are therefore measurements on the authors' own product rather than opinions, and I state explicitly which. I did not download the 898 MB SHAPES archive. I am happy to make the analysis scripts available to the authors through the handling editor, or to post them as a supplement to this comment, whichever the editor prefers.

1. Summary and overall assessment

This is a valuable dataset and a well-built processing chain. The glacier- and scene-specific Otsu thresholding, the vertical correction for surface lowering, and the explicit terrain-shadow masking are sound choices, better documented than in most of the antecedents. The Zenodo record is clean, openly licensed (CC BY 4.0) and its README documents the delivered fields with units. The product (the first regionally consistent, glacier-specific SLA dataset for the Alps) is clearly within the scope of ESSD and clearly worth publishing.

My concerns are not about the per-scene retrieval, which I do not question, but about the **annual aggregator**, the **cross-sensor verification**, the **uncertainty budget**, and the **completeness of the data description**. The central one is quantitative: reproducing the authors' own pipeline on their own published tables, **roughly 45 % of the reported trend is attributable to the growth of the satellite constellation rather than to a climatic signal** (General comment 1).

I want to be equally clear about what that does *not* mean. The trend survives. After controlling for sampling density it is still $+2.70 \text{ m yr}^{-1}$ and 326–348 of the 398 glaciers remain positive. The Alpine snowline is rising; what is not defensible is the published magnitude. All of the tests I ask for can be run on the data the authors have already generated; none requires reprocessing the archive. If they come out as I expect, the paper will be considerably stronger than it is now, because the trend will have been defended rather than merely reported.

2. General comments

G1. The end-of-summer estimator is a sample maximum, and sampling density explains ~45 % of the reported trend

Sections 3.6 (l. 329–332), 4.1 (l. 357–365), Figs. 2 and 8.

The annual EoS SLA is defined as "the highest SLA observed on each glacier during the mid- to late ablation season" (15 July – 30 September). The expected value of a sample maximum increases monotonically with sample size. Consequently, even for a perfectly stationary true SLA, the estimated EoS SLA would rise through the record purely because acquisition density rises. The direction of the artefact coincides with the direction of the published result.

The replication is faithful. Testing all combinations of window, field and QA cut-off against the published EoS table, the one that reproduces it is the **15 July – 30 September window, the 'SLA_CORR' field, and QA = 1**, with 99.4 % exact agreement over 9,573 glacier-years. (Incidentally, the manuscript never states which of the two QA cut-offs governs the main result; it is the strict one.) Adding the eligibility criteria (i)–(iii) and the ± 2 SD outlier filter of Sect. 3.6:

	This replication	Manuscript
Glaciers with a trend	398	398
Mean of glacier trends	+4.93 m yr⁻¹	+5.27
Positive trends	362	363

The 0.34 m yr⁻¹ difference comes from unpublished details of the outlier filter and is immaterial, because the comparison that follows is internal: the same chain with and without homogenised density.

The measurement. The median number of scenes per glacier and year inside the EoS window, at QA = 1, is **3.0 (2000–2012), 4.0 (2013–2016) and 14.0 (2017–2025)**. It correlates with the regional annual EoS anomaly at $r = +0.81$ over 26 years. Subsampling every glacier-year to a common density and repeating the entire chain:

Imposed density	Mean of glacier trends	90 % CI	Reduction
Actual (3 → 14)	+4.93 m yr ⁻¹	n/a	n/a
n = 5 per glacier-year	+3.27 m yr ⁻¹	[+3.20, +3.34]	34 %
n = 3 per glacier-year	+2.70 m yr⁻¹	[+2.59, +2.78]	45 %

(100 replicates per level; the CI is on the mean of the 398 trends.) The effect is systematic rather than noise: glacier by glacier, nearly every point falls below the 1:1 line. The attached figure shows all three panels.

The anomaly approach does not solve it. The alternative in Sect. 3.6 (l. 345–347) carries the same bias, because each year's anomaly is computed from a maximum whose bias depends on that year's n. That Fig. 8 colour-codes the mean DOY suggests the authors sense the problem; it needs to be quantified rather than hinted at.

I should add that this is not a peculiarity of the present workflow: Davaze et al. (2020) adopt the same annual aggregator ("the highest transient SLA") for the Alps, and explicitly note that the Landsat-7 SLC failure "drastically" reduced the number of usable images for SLA detection between 2003 and 2008. The concern is generic to the approach, which is precisely why a dataset of this ambition is the right place to quantify and correct it, and why the comparison proposed in G3(d) would be informative in both directions.

Requests. (i) Report a density-controlled trend as the headline figure (subsampling to a common n with bootstrap), or (ii) replace the annual maximum with an estimator that is insensitive to sample size (the value of a fitted seasonal curve on a fixed date, the median of the three highest observations, or a high percentile of the window); (iii) include n as a covariate in the regression and report its coefficient; (iv) add the **number of scenes used per glacier and year** to the dataset (the DOY is already delivered), without which this test is not reproducible by third parties; (v) state which QA cut-off governs the published results.

G2. The uncertainty budget is built in the wrong direction, and the delivered field is effectively a constant

Section 3.5 (l. 297–327), Eq. 5, Sect. 4.1 (l. 371–372).

Three points, in order of severity.

(a) Direction of the experiment. The terms $\delta^{\text{MASK_glacier}} = 88.0$ m, $\delta^{\text{MASK_extreme}} = 41.7$ m and $\delta^{\text{MASK_t.shadow}} = 37.9$ m are obtained by *deactivating* each filter and computing the RMSE against the fully filtered estimate. That measures the **sensitivity** of the estimate to the filter, not the **error** of the filtered estimate. What is needed is the residual error that survives the filter: undetected cloud, unmasked shadow, unexcluded rock. That residual cannot be obtained by switching the filter off, because the unfiltered run is not truth. The 23 reference glaciers already provide a route: label cloud/shadow/rock on those same scenes and measure the error that survives the mask.

(b) Independence. Equation 5 sums in quadrature assuming independence, but the first four terms act on the *same* NSIR histogram of the *same* scene: a scene with partial coverage is also the one most prone to undetected cloud and to extreme-value bias. Please justify the independence assumption, estimate the covariance, or derive an empirical global uncertainty from the reference comparison, stratified by coverage.

(c) Verified in the delivered data: the field has a floor and barely varies. In `snow_line_export_filter.R` the three terms above enter as constants (`ndsi_error <- 88.03`, `range_error <- 41.73`, `shade_error <- 37.88`), so the minimum possible value of `SLA_ERROR` is $\sqrt{88.03^2 + 41.73^2 + 37.88^2} = \mathbf{104.53}$ m. Across the 355,752 published observations the minimum is exactly **104.53 m**, the **median is 105.9 m**, and the interquartile range is 105.4–110.1 m. The terms that actually vary, coverage and the dh/dt correction, therefore contribute **less than 6 m for half the dataset**: the "per-observation" uncertainty is in practice a constant. Moreover, **no observation in the dataset carries an uncertainty below the 95.7 m RMSE** of the manual validation, or below the 89.7 m same-day cross-sensor RMSE.

An uncertainty field whose minimum exceeds the validated error, and whose median sits 1.4 m above that minimum, cannot be propagated by a downstream user, which is precisely the use proposed in Sect. 5.3. It needs to be rebuilt or rescaled.

G3. The validation is not independent, and the more independent comparison is under-weighted

(a) The 890 reference snow lines were digitised by the same group, and their elevations are extracted from the **same NASADEM** with the **same dh/dt correction**. DEM error, correction error and georeferencing error are common to both sides of the comparison and cancel. The 95.7 m RMSE correctly measures *classification* error, but not the total error of the delivered elevation, which is what Sect. 3.4's title and the abstract's "root mean square error of ~ 100 m vertical deviation" imply. Please say so explicitly in the abstract and in Sect. 4.1.

(b) Definitional mismatch. The automated SLA is the 10th percentile of the elevations of snow-classified pixels (l. 239–240); the manual reference is a drawn line. These are different quantities whose difference depends on glacier hypsometry and snowpack fragmentation. This probably explains the spread of per-glacier biases in Table A2, from –142 m (Corbassière) to +99 m (Fernerstube). That is a 240 m range, far larger than the global RMSE. Please add (i) a sensitivity of the SLA to the percentile choice (5/10/20/25) and an empirical justification of 10; (ii) a testable hypothesis for the spread of per-glacier bias (area? debris cover? elevation range? Table A2 already has the columns); (iii) if the per-glacier bias is stable in time, deliver it as a column.

(c) The ELA comparison deserves more weight, not less. Figure A2 gives RMSE = 212.9 m, with EoS SLAs *hundreds of metres below* the ELA above 3,200 m a.s.l. This is the only genuinely independent comparison in the paper (WGMS glaciological data, an entirely different measurement chain) and the one that matters most for the application the paper puts forward as its main justification. It is currently settled in five lines, relegated to an appendix, and absent from the abstract. Please move it into the main text; report the *signed* bias as a function of elevation, not only the RMSE; give the figure in the abstract alongside the ~100 m; and assess whether the product is usable for surface mass balance calibration on the highest, coldest glaciers, where the paper itself acknowledges that firn is classified as snow (l. 479–483). If it is not, that is a limitation of use and should be stated as such.

(d) A directly comparable, independent SLA dataset for the same mountain range exists and is not used. Davaze, L., Rabatel, A., Dufour, A., Hugonnet, R., and Arnaud, Y.: Region-Wide Annual Glacier Surface Mass Balance for the European Alps From 2000 to 2016, *Front. Earth Sci.*, 8, 149, <https://doi.org/10.3389/feart.2020.00149>, 2020. It is cited once in the introduction and never engaged with, and it derives **annual end-of-summer snowline altitudes for 239 Alpine glaciers over 2000–2016** from multi-mission optical imagery (Landsat-5/7/8, Sentinel-2, ASTER) using a NIR/SWIR band ratio, i.e. the same spectral basis as NSIR. Their annual SLA is defined, as here, as **the highest transient SLA of the year**, and their annual ELA product is publicly available through the THEIA portal, with the mass-balance series as CSV supplementary material.

This is the independent validation the paper is missing, it is like-for-like (same region, same quantity, same annual aggregator, overlapping 2000–2016), and it requires no new processing. I would regard a glacier-by-glacier comparison over the common period and the common glaciers as the single most valuable addition the revision could make: agreement would settle much of G1 and G4 at once, and disagreement would be equally informative. Note also that their reported SLA uncertainty is ± 123 m, of the same order as the value obtained here, which is reassuring for the magnitude even though it does not resolve the internal inconsistency raised in G2.

G4. Cross-sensor homogeneity is not demonstrated

Sections 3.4 (l. 294–296) and 4.1 (l. 379–384), Table A3.

The text states that there are "no systematic differences between image acquisitions of the individual optical satellite missions" (l. 379–381). **Table A3 cannot support that claim:** it reports mean absolute deviation, RMSE and percentiles of *absolute* differences, and a signed systematic difference is exactly what those statistics remove by construction. Please report the *signed* mean for each sensor pair with a confidence interval.

This is not a presentational detail. The composition of the archive changes monotonically through the record, and Sentinel-2 contributes 7,736 of the 13,454 scenes (Table A1). Three specific concerns:

1. **The fixed NDSI threshold of 0.7 is the sensor-sensitive link.** The paper's argument is that Otsu makes the method robust to radiometric differences, which is true for the snow–ice separation. But *before* Otsu, a **fixed** $\text{NDSI} \geq 0.7$ threshold is applied to GREEN and SWIR bands whose bandwidths differ between TM, ETM+, OLI and MSI (as the paper itself notes at l. 116–117, citing Chastain et al., 2019). That step determines which pixels enter the histogram and can therefore shift the Otsu threshold in a sensor-dependent way. The robustness attributed to Otsu does not protect it.
2. **Spatial resolution.** Sentinel-2 delivers 10–20 m against Landsat's 30 m, and an estimator defined as the 10th percentile of snow-covered elevations is resolution-sensitive: small snow patches resolved at 10 m are diluted into mixed pixels at 30 m. The effect has a sign and coincides in time with the second half of the record. Please degrade a subset of S2 scenes to 30 m, reprocess, and report the signed offset.
3. **The static NSIR valid range.** The 10th/90th percentiles of the per-pixel minimum and maximum are computed over the full 25 years of August–September imagery and applied to all scenes (l. 216–219). If bare ice darkens through the period, as documented in the Alps by Naegeli et al. (2019), this static clipping acts differently at the beginning and end of the record. This is a third route to instrumental drift and is not evaluated.

Requests. Signed means per sensor pair with CIs, stratified by elevation and month, resolved by **platform** rather than by collection, since the delivered Sensor field lumps Sentinel-2A, 2B and 2C under one label and the constellation changed composition in 2025 (see l. 116); the regional trend recomputed on a **sensor-homogeneous subset** (Landsat-only 2000–2025 is the informative one, with OLI+MSI 2017–2025 as a control); a discussion of the sensor sensitivity of the NDSI step; and, if a systematic offset appears, band-pass adjustment to a common sensor *before* the indices are formed, rather than a post-hoc correction of the SLA. Offsets of a few tens of metres between TM/ETM+ and OLI/MSI would be enough to matter at the magnitude at stake here, so a null result would itself be a valuable addition to the paper.

One caveat on how that adjustment can be done, since the four sensors are not equally served. Published surface-reflectance transformation coefficients exist for ETM+ and OLI (Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., and Egorov, A.: Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, *Remote Sens. Environ.*, 185, 57–70, <https://doi.org/10.1016/j.rse.2015.12.024>, 2016), and Chastain et al. (2019), already cited at l. 116–117, covers ETM+, OLI and MSI. Neither covers **TM**, which contributes the entire early part of the record and is therefore the sensor on which the trend most depends. The tandem cross-calibration of Teillet et al. (2001, *Remote Sens. Environ.*, 78, 39–54) is radiometric rather than a band-pass transformation of surface reflectance, and its agreement is weakest in bands 5 and 7, which are precisely the SWIR bands that NSIR and NDSI rely on. For TM, then, I am not asking for coefficients that do not exist: an empirical same-day LT05–LE07 comparison over the 2000–2011 overlap, on the glaciers themselves, would establish whether an offset is present and would be the natural basis for a correction. Table A3 already reports that pair.

G5. Fixed thresholds without empirical distributions or sensitivity analysis

The workflow depends on several constants presented without the empirical distribution that would justify them: $\text{NDSI} \geq 0.7$ (l. 197–198), $\text{NDWI} \leq 0.1$ (l. 224–226), the 10th percentile that defines the SLA (l. 239–240), and the 10th/90th percentiles of the NSIR valid range (l. 216–219). Notably, three of the six QA criteria *are* derived as the 10th percentile of the observed distribution, which is exactly the right approach; please apply the same standard to the rest. A single appendix figure with the histograms and the shift in median SLA when each threshold is moved one step in either direction would settle this.

There is also an internal inconsistency in the NDWI reasoning: at l. 181–183 partially melted bare ice is said to show *high* NDWI relative to the snow-covered accumulation area, while at l. 224–226 pixels with NDWI > 0.1 are rejected to avoid detecting snow where refrozen ice is present. The same index is used as the indicator of two different surfaces in consecutive paragraphs. Please rewrite, and quantify what fraction of pixels the filter removes and at which elevations.

Finally, a reproducibility point that the code resolves but the paper does not: since NSIR is computed on physical reflectance and is quadratic in NIR (see Technical corrections), the values "SD > 3" and "Otsu = 7" are index values in physical units, not scaled integers. As written, the reader cannot verify either threshold. Please state the scaling explicitly.

G6. Trend statistics

1. **The mean of only the positive trends is not a regional estimator.** Lines 415–418 report $5.27 \pm 4.54 \text{ m yr}^{-1}$ (all, $n = 398$), 5.99 ± 4.00 (positive only, $n = 363$) and 6.61 ± 4.15 (positive and significant, $n = 306$). The latter two are conditioned on sign and on the outcome of the test and therefore overestimate by construction; they should not be offered as alternatives to the first. The abstract and conclusions adopt " $\sim 6 \text{ m yr}^{-1}$ ", which is closer to the conditioned figures than to the unbiased one.
2. **$\pm$ SD is not an uncertainty.** The ± 4.54 is the spread across glaciers, not the error of the regional mean. The latter requires an interval that accounts for spatial correlation between neighbouring glaciers (evident in Fig. 5) and for temporal autocorrelation; a spatial block bootstrap is the minimum.
3. **A t-test on the OLS slope is not appropriate here.** Annual residuals are heteroscedastic by design, since the number of scenes per year varies by a factor of five (now measured, G1), and they are autocorrelated. Theil–Sen with Mann–Kendall, or OLS with robust errors and weighting by n , would be defensible.
4. **No correction for multiplicity.** 306 significant trends out of 398 tests; at $\alpha = 0.05$ some 20 false positives are expected. The count will probably survive, but a false-discovery-rate procedure (Benjamini–Hochberg) should be applied and reported, and likewise for the six bivariate comparisons of Fig. 6.
5. **The outlier filter is circular and asymmetric.** Discarding EoS SLAs more than ± 2 SD from the glacier's multiannual mean preferentially removes values at the ends of a trending series, thereby attenuating the trend being measured, and it removes exactly the 2022 and 2023 values later highlighted as signal (l. 411–412). The paragraph is also internally contradictory: it opens by describing a procedure for years with *fewer than three scenes* and continues with rules based on the number of valid *years*. Please rewrite it, and report how many values are discarded, in which years, and the trend with and without the filter.

G7. Over-interpretation of the bivariate analysis (Fig. 6)

The $r^2 = 0.89$ correlation with mean glacier elevation is largely structurally forced: the SLA is confined to the glacier's own elevation range, so the correlation mostly measures that constraint rather than a climatic property. Normalising (SLA minus mean elevation, or the SLA's relative position in the glacier hypsometry) would likely reveal the spatial patterns that Figs. 5 and 6a,b hint at and that elevation currently masks; the conclusions (l. 584–585) present this correlation as a finding and should be reformulated. Aspect is a circular variable and does not admit a linear bivariate regression in degrees (Fig. 6f); circular statistics or sector-wise non-parametric tests are needed. The west–east and south–north gradients are described as "present but statistically insignificant" (l. 407–409) without reporting the statistic or the p-value. Finally, all six RGI

variables are geometric, while the most interesting result, the sharp contrasts across the divide in the Bernese Alps (l. 404–406), is climatic; adding one precipitation variable from a gridded Alpine product would turn that observation into a result.

G8. Physically meaningless SLAs are delivered as numbers and are not flagged

The paper honestly acknowledges three situations in which the SLA has no physical meaning: a fully snow-covered glacier, where Otsu forces a split and biases upward (l. 472–476); a fully snow-free glacier, where isolated residual patches produce "a physically meaningless SLA estimate" (l. 476–478); and exposed firn classified as snow (l. 479–483).

The difficulty is that these observations enter the dataset as numerical values indistinguishable from valid ones. The QA criteria capture some of the cases, but indirectly and aggregated into a scalar. For a product intended for model calibration, explicit categorical flags are needed: `snow_covered_full`, `snow_free`, `unimodal` (the Bhattacharyya distance is already computed and delivered; exposing it as a flag would suffice) and `firn_likely`. Plotting SLA against the snow-class fraction (`Cov_C1s`, already delivered) and against the Otsu threshold should separate the degenerate regimes visually and give the flagging criterion.

G9. Elevation-change correction, shadow mask and window length

(a) dh/dt correction. The mean correction is ~30 m (l. 493), about 20 % of the reported total rise; a term of that magnitude deserves its own sensitivity analysis rather than a passing mention. Constant 2000–2019 rates are extrapolated to 2020–2025, a period with exceptional losses that the paper itself highlights; if the extrapolation underestimates surface lowering, the correction is insufficient and the trend is overestimated. Please recompute the regional trend with post-2019 rates scaled by, say, 1.5× and 2× (or using the GLaMBIE product already cited at l. 442) and report the range. The factor of 2 applied to the post-2019 *uncertainty* (l. 326–327) is also unjustified.

(b) Static terrain-shadow mask (l. 200–206). A mask computed every 10° of azimuth at the *mean* solar elevation and then merged is time-invariant: it removes valid pixels at midsummer and lets shadow through at the season's ends, exactly where the paper reports its largest biases (l. 592–595). More importantly, terrain shadow is not randomly distributed in hypsometry; it concentrates in high cirques and north-facing slopes, so a static mask systematically clips the upper part of the histogram, which is what defines the snow class. Please report the elevation distribution of masked pixels and the SLA shift relative to a scene-specific mask, separated by month.

(c) Window length. Images are processed to 30 November (l. 159–160) but the EoS SLA is defined only to 30 September. If the ablation season is lengthening, the true maximum increasingly falls *outside* the window, biasing the trend downward. That is opposite in sign to G1 and therefore equally worth quantifying. Please repeat the computation with the window extended to 15 and 31 October, and report the distribution of the selected maximum's DOY through the record; a shift would be direct evidence of a lengthening melt season, and the data are already published in `Alps_Glacier_EoS_SLA_DOY_2000-2025.csv`.

(d) Glacier area turnover. Section 5.1 (l. 447–458) acknowledges that retreat causes a "progressive reduction in the effective glacier area in later years". That mechanism ties spatial coverage to year, and coverage is both the dominant error term (RMSE 162.6 m, $r^2 = 0.68$) and the principal QA criterion, so the quality filter becomes progressively stricter through the record, which interacts with G1. Please report the annual fraction of observations passing each QA cut-off.

3. Data and code availability (ESSD-specific)

The Zenodo record is in good shape and I want to say so plainly: DOI, CC BY 4.0, and a README that documents the eighteen fields of the time series with units, including the per-observation uncertainty, the applied elevation correction, the Otsu threshold and the Bhattacharyya distance. The vertical datum (EGM96) is stated there. All Earth Engine inputs are *public* catalogue assets (projects/sat-io/open-datasets/RGI/RGI_VECTOR_MERGED_V7 and the Hugonnet et al. elevation-change collection), so the workflow is reproducible without private permissions; the manuscript's vague reference to the "awesome-gee-community-catalog" (l. 101–102) should simply name these two asset IDs.

However:

6. **The QA-FLAG rule in the code contradicts the manuscript.** In `snow_line_export_filter.R` the flag is the unweighted mean of six booleans divided by six, with thresholds `flag_cov_glb <- 30` (per cent), 0.5 km², 0.09 km², 3, 7 and 0.2. Five match Sect. 3.3; the first does not. The manuscript states "> 50 %" coverage (l. 256–257) while the code applies 30 %. Recomputing the flag from the delivered columns reproduces the published QA_FLAG for **100.0000 %** of the 355,752 rows with the 30 % threshold, and for only **81.4 %** with 50 %. The code value governs version 1.0 of the record; Sect. 3.3 needs correcting, and the aggregation rule should be given as an equation.
7. **The flag is unweighted, and that matters.** Coverage, the term that governs the error, carries the same weight as the Bhattacharyya distance, so an observation that fails *only* coverage scores $5/6 = 0.833$ and passes the $QA \geq 0.75$ cut-off used for the headline count. Measured: of the 235,707 observations with $QA \geq 0.75$, **56,958 (16.0 %)** have coverage below the 50 % stated in the paper, and **12,894 (3.6 %)** fail even the code's 30 % criterion, with a median coverage of 22.7 %. Please either weight the flag, make coverage a disqualifying rather than a scored criterion, or at minimum report these numbers.
8. **"QA ≥ 0.75 " is exactly "five of six criteria".** The flag can only take the values $k/6$, as verified in the data. Stating it that way avoids sending users to look for observations between 0.75 and 0.833, of which there are none.
9. **The paper is not self-contained.** ESSD expects the article to describe the product rather than to refer implicitly to a README that a reader only finds after downloading the record. Please move the field table with units, the vertical datum and the licence into the paper or an appendix.
10. **Missing from the dataset:** the number of scenes used per glacier and year (G1). Missing from the README: the coordinate reference system of the geopackages, record counts per file, and documentation of the statistics file's columns (units of `eos_sla_trend`, the method behind `eos_sla_sig`).
11. **The code does not meet the availability requirement.** The repository has **no licence** (the GitHub `license` field is null), so by default it is all-rights-reserved and not legally reusable; and there are **no releases**, so there is no tagged version and no code DOI. The last push (2026-01-14) predates the discussion phase, so there is no frozen state corresponding to the preprint either. Please enable the Zenodo–GitHub integration, tag a release, cite that DOI in Sect. 7, and add a licence. The chain is also bilingual, Python for processing and R for post-processing, which the paper does not mention; an environment file (`requirements.txt` / `renv.lock`) and stated versions of `earthengine-api`, Python and R would make the product genuinely reproducible. The README additionally still contains the placeholder `ADD-PREPRINT-DOI`.

12. **Consider partitioning the 898 MB `SHAPES` archive** by glacier or region; as delivered, a user must download nearly a gigabyte to inspect one glacier.
-

4. Specific comments

- **I. 12–13 (abstract).** State that the ~100 m RMSE is agreement with the authors' own visual delineation on the same elevation basis, and give the 212.9 m RMSE against WGMS ELAs alongside it.
 - **I. 91.** Give the number of glaciers excluded by the 1 km² threshold, not only the 75 % of area; by count, the dataset covers a minority of Alpine glaciers, which conditions its use in regional models.
 - **I. 105–107.** Explain why the series begins in 2000 rather than 1985, given that Landsat-5 is available from 1984. If the reason is that the dh/dt correction begins in 2000, say so; it is a defensible choice.
 - **I. 112–113.** The 8-day parity of Landsat-5 and -7 is broken by the ETM+ SLC failure from May 2003, mentioned later (I. 428–430); the two passages should be linked.
 - **I. 116.** Clarify that Level-2A availability in GEE begins in 2017 although the mission operates from 2015, and state whether L1C + Sen2Cor was considered to recover 2015–2016. The delivered tables confirm the gap: the product contains no Sentinel-2 scene in 2015, one in 2016, and 69 in
13. Two further points belong here, because the Sentinel-2 constellation is not constant across the record. First, its composition changed inside the study period: Sentinel-2C was launched on 5 September 2024 and replaced Sentinel-2A operationally on 21 January 2025, and from March 2025 Sentinel-2A was returned to service in an extension campaign, so that for the whole of the 2025 ablation season the mission flew **three** satellites rather than the nominal two. The record ends on 23 November 2025, so the final season of the trend is acquired by a constellation unlike any earlier year. Measured on the delivered tables, 2025 carries 167 unique Sentinel-2 scenes against 146 in 2024, the highest count of the record; the product's yearly coverage begins in April, so the January and March transitions are not directly visible in it. Second, and this is the actionable part, the delivered Sensor field carries the single value S2_SR_HARMONIZED for all 147,745 Sentinel-2 observations, so neither Table A3 nor a downstream user can tell Sentinel-2A, 2B and 2C apart. The platform is available in the Earth Engine image properties (SPACECRAFT_NAME) and costs nothing to export. Please add it, declare which satellites contribute, and resolve Table A3 by platform rather than by collection (G4). Incidentally, the use of the *harmonized* collection is the right choice and worth stating in the text, since it absorbs the January 2022 processing-baseline offset that would otherwise introduce a step in reflectance in the middle of the record.
- **I. 131–133.** Give the nominal date of the RGI 7.0 outlines for the Central Europe region.
 - **I. 137–139.** Quantify the area change caused by simplifying outlines to a 60 m minimum vertex spacing, and its effect on the computed spatial coverage, which is the dominant QA criterion.
 - **I. 146–148.** The choice of NASADEM over SRTM v3 is well justified; one line on why not Copernicus GLO-30 (used in the products' own topographic correction, I. 125–126) would pre-empt the question.
 - **I. 158–159.** The < 75 % cloud filter uses whole-scene metadata, which says little about the glacier. Was a glacier-clipped cloud fraction considered?
 - **I. 190–192.** State how overlaps between adjacent orbits with different viewing geometries are resolved in the daily mosaics (median, first valid pixel, simple mosaic?).

- **I. 196–199.** NDSI ≥ 0.7 is used simultaneously to exclude clouds *and* rock/debris, two objectives with different spectral regimes. Separate them conceptually in the text.
- **I. 222–224.** Quantify "a small number of late-summer images".
- **I. 234–235.** With a 30 m DEM and several metres of vertical error, 10 m elevation bins are below the effective resolution; justify, or use 25–50 m bins and show that nothing changes.
- **I. 251–253.** The 10 % minimum-coverage cut-off is permissive given the 162.6 m RMSE in that regime (Fig. 3d); please justify it or flag those observations prominently.
- **I. 286–287.** Justify the 100 m buffer around the manual lines and show sensitivity (50/100/200 m).
- **I. 292–294.** Report the number of glacier-year pairs available from WGMS; that is the effective n of Fig. A2.
- **I. 520–521.** The **"first region-wide assessment of spatio-temporal SLA variations for glaciers across the Alps" claim does not stand as written.** Davaze et al. (2020), cited by the authors at I. 26, produced annual, glacier-specific, end-of-summer SLAs for 239 Alpine glaciers over 2000–2016 from multi-mission optical imagery, using a NIR/SWIR band ratio and the annual maximum of the transient SLA. That is a region-wide, glacier-specific SLA assessment for the Alps. The present work has real and substantial novelty over it: a longer record (to 2025), more glaciers (409 against 239), *transient* rather than annual-only SLA, an explicit dh/dt correction, a per-scene quality flag, and, above all, **the trend analysis, which Davaze et al. do not perform.** I would encourage the authors to state exactly that instead. A precise claim of what is new is stronger than a broad one that a reader can falsify with a paper already in the reference list. Please also qualify with respect to Naegeli et al. (2019) and, in particular, Hu, Z., Dietz, A., Zhao, A., Uereyen, S., Zhang, H., Wang, M., Mederer, P., and Kuenzer, C.: Snow Moving to Higher Elevations: Analyzing Three Decades of Snowline Dynamics in the Alps, *Geophys. Res. Lett.*, 47, e2019GL085742, <https://doi.org/10.1029/2019GL085742>, 2020. That paper reports significant snowline rises of about 4–11 m yr^{-1} in four to five of six Alpine catchments over 1984–2018 and is absent from the reference list. It measures regional seasonal snowline rather than glacier-specific SLA, so the claim probably stands if narrowed to *glacier-specific*; citing Hu et al. and stating the difference in metric would make the claim stronger rather than weaker. It is also a useful external check on magnitude: their range brackets the density-controlled value obtained in G1.
- **Fig. 2.** Central to G1. Please add the number of scenes *per glacier* (median and percentiles), not only the regional total, since the density argument operates per glacier.
- **Fig. 3.** Add the 1:1 line and the number of pairs in each panel; separating (e) and (f) by month would locate the problematic early- and late-season regimes.
- **Fig. 5b.** Show *all* 398 trends on a diverging scale rather than only the 363 positive ones; displaying only the positive trends is a presentational bias. Mark significance with a symbol (e.g. stippling) instead of excluding glaciers.
- **Fig. 6.** See G7; report n and the test statistic in each panel.
- **Fig. 8.** Add the median number of scenes per glacier and year as a second axis; it is the variable G1 calls into question, and this figure is the natural place to settle it.
- **Fig. A1.** Coercing the error to zero above 95 % coverage understates the uncertainty of the best observations, which dominate the recent years. Justify, or use the fitted value.
- **Table A2.** Add debris cover and elevation range to allow the per-glacier bias analysis of G3(b); sorting by RMSE would help the reader.

- **Table A3.** Add the *signed* mean and its CI (G4), and explain the absence of the LT05–S2SR pair (no temporal overlap).
- **Sect. 5.3.** A well-written and honest section. Please add a sentence on which QA cut-off a user should apply for calibration purposes, referring to the flags requested in G8.
- **Code.** The NASADEM swb (surface water) band is used in `main_gee_rgi-id_tsla_proc.py` but never mentioned in the paper. The code also includes the Landsat-4 collection (LANDSAT/LT04/C02/T1_L2), which the paper does not declare and Table A1 does not list; I verified that no LT04 observations exist in the product, so the branch is inactive; please remove or document it.

5. Technical corrections

- **Eq. 1 and l. 170–171.** The code computes `band_nir.multiply(band_nir.divide(band_swr))`, i.e. $NSIR = NIR^2/SWIR$. The equation is correct but ambiguously typeset; the *prose* ("the ratio (NSIR) of near-infrared (NIR) and short-wave infrared (SWIR)") is what is wrong. Please correct the text, typeset the equation unambiguously, and note that the index is quadratic in NIR and therefore not scale-invariant: a relative error in BOA reflectance propagates twofold, where a normalised ratio would cancel it to first order. A useful check, cheap because the data are already published: show that the Otsu threshold (TH_{NSIR}) does not correlate with solar elevation angle or with sensor.
- l. 222: "seamingly" → "seemingly".
- l. 231: "misdetected" → "misclassified".
- l. 304: "Error the DEM-differencing elevation change correction" → "Error of the ...".
- l. 382: "manual and automated SLE measurements" → "SLA".
- l. 442: please rewrite "can be often easily detected posterior based on".
- l. 487: "relying on Otsu thresholding Landsat imagery" → "... Otsu thresholding of Landsat imagery".
- l. 491, 503: "(Rastner et al., 2019) demonstrated", "(Loibl et al., 2025) provided" → "Rastner et al. (2019) demonstrated", "Loibl et al. (2025) provided".
- l. 584: "exposition" → "aspect", used consistently throughout.
- l. 609: "retrival" → "retrieval".
- Sect. 3.5 lists $\delta^{\wedge}AREA_covered$ while Eq. 5 uses $\delta^{\wedge}AREA_coverage$; please unify.
- "snow line" and "snowline" are both used throughout; please choose one.
- In `functions.py`, the NDSI condition is `ndsi.gte(0.7).And(ndsi.lte(2))`. An upper bound of 2 is unreachable for a normalised difference bounded in $[-1, 1]$; the clause is dead code and suggests the index is scaled, which it is not.

Numbers that disagree with the published product (all verified against the delivered tables):

Manuscript	Published product
408 glaciers (abstract, Sects. 2.2, 4.2, conclusions)	409 tables and 409 columns in the EoS table
"~200,000" observations; 120,000 Landsat + 103,000 S2 (Sect. 4.1)	235,707 at QA ≥ 0.75 (130,337 Landsat, 105,370 S2); 355,752 unfiltered

Manuscript	Published product
To 20 October 2025 (Sect. 2.1); "November 2025" (Sect. 3.1); 2000–2025 (title); 2025-12-31 (Zenodo)	First observation 2000-04-08 , last 2025-11-23
Sect. 3.6 (405 of 408) vs Sect. 4.2 (n = 398)	405 have a mean, 398 have a trend; please make this explicit

6. Questions I would like answered in the response

14. Does the main result use $QA = 1$ or $QA \geq 0.75$?
 15. Which coverage threshold is correct, the 30 % in the code or the 50 % in the paper?
 16. Why does the product contain 409 glaciers and 235,707 observations where the paper reports 408 and ~200,000?
 17. How is a per-observation uncertainty field justified whose median lies 1.4 m above its theoretical minimum?
 18. By how much does the regional trend change when sampling density is controlled?
-

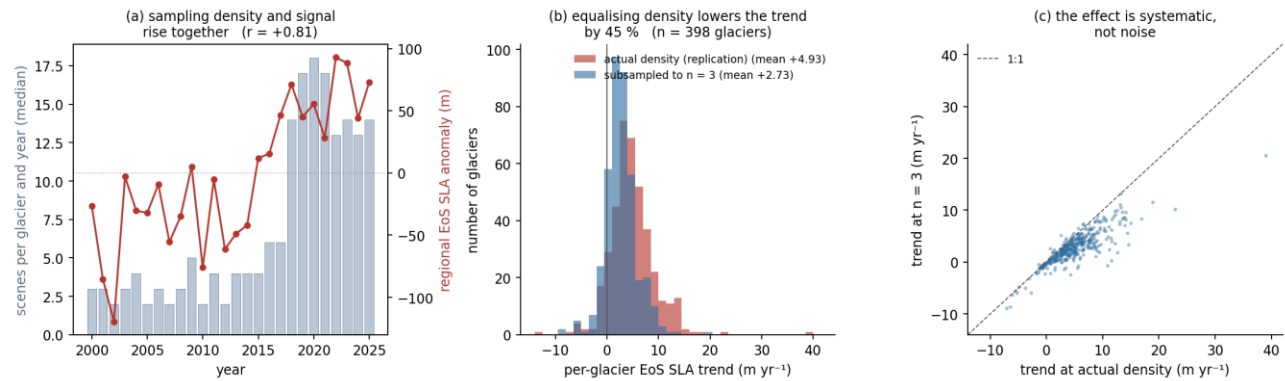
7. Closing

To summarise what I think is needed. First, demonstrate the trend against the fivefold increase in acquisition density, or better, report the density-controlled value, since the measurement above indicates that roughly half the published magnitude is an artefact of the growing constellation. Second, measure the cross-sensor offsets *with sign* and recompute the trend on a sensor-homogeneous subset. Third, rebuild the uncertainty budget so that it is consistent with the validation. Fourth, bring the data description into the paper, correct the QA-FLAG rule, flag the physically meaningless retrievals, and archive and license the code.

None of this requires reprocessing the archive, and none of it threatens the paper's central claim: the Alpine snowline is rising, and this dataset is the best regional evidence for it. What the revision would add is that the number attached to that statement becomes defensible. I would be glad to review a revised version.

The figure supporting G1 is attached below.

Attached figure



Sampling density and the end-of-summer trend, measured on the published tables. (a) Median number of scenes per glacier and year inside the EoS window (bars) and the regional EoS SLA anomaly (line). (b) Distribution of the 398 per-glacier trends at the actual sampling density and after subsampling every glacier-year to $n = 3$. (c) The same comparison glacier by glacier against the 1:1 line.