Comments from Referee #2, followed by the authors' responses

General Comments

C: The study by Li et al. aims at presenting a method that allows the automated mapping of glacier extents in a challenging region using Google Earth Engine at high temporal resolution, up to annually from Sentinel-2. They used the results to consider area changes when calculating glacier mass balance. If the mapping had worked, this would have been an important study to improve related results also for other regions in the world. Unfortunately, the outcome of the mapping is not useful for any assessment. In this regard I want to acknowledge that the authors have provided the results of their mapping effort in the supplemental material. Without this, my evaluation would have been different as the paper is otherwise well written and the idea to just use all data available and combine it for the best possible result is fine. However, the largely arbitrary area changes from dataset to dataset (for individual glaciers partly larger than 50% from year to year) are obvious and glaciologically impossible. The authors mention that there are the usual problems with debris cover, clouds and shadow, but they have seemingly not recognized how large and arbitrary the variability is and that their method does not produce meaningful results.

R: We sincerely thank you for your professional and insightful comments. We acknowledge that the fully automated mapping approach employed in this study has certain limitations, particularly in challenging regions. Some descriptions inadvertently overstated the capabilities. Our study does not aim to reach the precision of the Randolph Glacier Inventory (RGI), which cannot realistically be achieved through fully automated mapping alone. Currently, automated glacier identification faces two major difficulties: first, the spectral characteristics of debris-covered glaciers are highly similar to those of bare ground, making accurate delineation extremely difficult; second, due to limited experience in interpreting such glaciers and a lack of high-quality samples, our manually drawn training data may not adequately represent debris-covered ice, further affecting the recognition accuracy. Thus, the current results should be interpreted more as changes in ice-covered area rather than precise glacier extent.

The considerable fluctuations in apparent glacier area primarily stem from extensive data gaps in the image series. In response, we have conducted additional quality control on a glacier-by-glacier basis, excluding unreasonable changes caused by severe data missing. This process significantly improves the reliability of the area change estimates. In the revised manuscript, we will aggregate the results into five-year intervals to provide a more robust assessment of glacier change, which also aligns better with the goal of supporting mass balance calculations. We fully appreciate your insights and have carefully addressed each comment, making corresponding improvements throughout the manuscript. we have revised the text throughout to more accurately reflect the scope and limitations of our work. Below, we provide a detailed, point-by-point response to your suggestions.

C: Neither the unrealistic area increase by about 500 km2 (estimated from Fig. 8, numbers for individual years are not provided) from 2019 to 2022 (naming it as a 'consistent decline' in L387), nor the sudden strong increase from 2016 to 2017 is discussed or considered as unrealistic. Instead, the authors correlate glacier elevation changes (wrongly labelled as 'Glacier Thickness' in Fig. 8) with glacier area changes as they assume there is a correlation (L407) and think that the correlation can be used as a validation (L413) of their (wrong) glacier areas. In fact, area changes are mostly driven by the ice thickness distribution along the glacier perimeter (thus depending on the shape of the glacier cross-profile) and are a longer-term response to changes in flow dynamics (glaciers have a response time). Hence, also the follow on analysis is a bit strange. In this regard, it is also unclear to me why the authors rely on results from Cryosat and ICESat (with their diverse range of issues) for such small glaciers instead of the Hugonnet et al. (2021) dataset that is widely used? As this dataset is not even mentioned in the comparison Table 1, I wonder why. Is the dataset too bad in quality?

R: We appreciate your suggestions. We acknowledge that the initial glacier extent results showed unrealistic year-to-year variability. To address this, we re-examined the classification outputs on a glacier-by-glacier basis and applied post-processing to flag and remove implausible area changes, including abrupt fluctuations or geometrically inconsistent outlines. Invalid outlines were replaced with the closest temporally consistent results. After these adjustments, the overall mapping is more stable, although uncertainties remain, particularly for complex glaciers. Glacier-by-glacier quality control was applied to detect and correct implausible changes, replacing them with temporally consistent outlines where possible, and marking unreliable years as no-data. Figures and analyses have been updated, and the Discussion now explicitly addresses these anomalies and their underlying causes. Regarding Fig. 8, we corrected the label from "Glacier Thickness" to "Glacier Elevation Change." We also removed any suggestion that the observed correlation between glacier elevation change and area change constitutes a form of validation, as glacier area and elevation changes are not necessarily synchronous or directly correlated per you suggest. For Hugonnet et al. (2021) dataset, a direct comparison with their study was not conducted initially. This is because we wanted to compare them with the results of time series of glacier mass balance, rather than trends. According to your suggestion, we have included the Hugonnet et al. (2021) dataset for comparison.

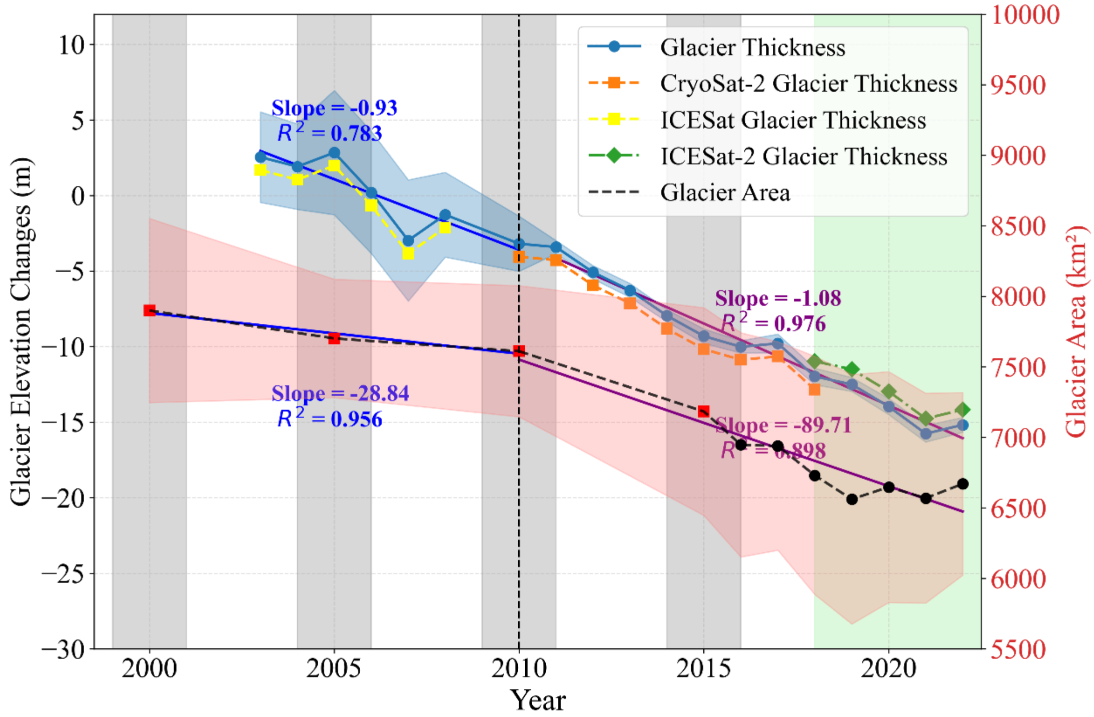

Fig1 Updated data on area changes

C: The very short results section (it has just 14 lines) mentioned the Kappa coefficient and overall accuracy along with three images showing outline overlays. I am aware that these statistical accuracy measures are frequently used in remote sensing studies to present the accuracy, but in my view they can be the result of anything and do not allow to obtain meaningful conclusions about the 'robustness and reliability of the classification approach'. At least for glaciers they do not work, as nicely confirmed here by the largely arbitrary results of the glacier mapping. The quality of mapped glacier extents can be shown by a) outline overlays and b) the sum of commission and omission errors (false positives and false negatives) divided by the common area. But as the former have been removed by the masking with RGI 7.0, I am unclear if the measures can be used here at all?

R: We appreciate your comment and agree that traditional statistical accuracy metrics can be misleading for glacier mapping due to complex glacier morphology and seasonal effects. To better assess classification reliability, we will supplement these metrics with comparisons to manually interpreted reference outlines and with spatial consistency checks across multiple time points. The approaches provide a more robust evaluation of the results. This clarification will be added to the revised manuscript.

C: I see missing debris-covered glacier parts and that large regions in shadow are sometimes missing. Hence, intensive manual editing would be required before resulting outlines could be used for change assessment. The statement that glaciers 'are well identified' (L377) seems misleading in this regard. Although the year 2000 dataset is likely the most complete regarding shadow and debris mapping, showing a region where the method does not work and discussing it would have been more helpful. One can see the problems of the classification a little bit for the 2022 outlines of the right glacier in the upper left panel [please name them properly a), b) and c)] of Fig. 8, but the image is very dark (what about some contrast stretching?), the lines are hard to see (also the red outlines on a reddish background in the insets of Fig. 7 are barely visible) and the wrong mapping results are not really discussed.

R: We are grateful for your comment. We acknowledge the limitations of our approach, as accurately identifying debris-covered glaciers remains a major challenging. As noted in our previous response, the spectral characteristics of debris-covered glaciers complicate their discrimination, and our sample selection strategy significantly influenced the results. For the year 2000, we relied on the RGI 7.0 for training sample delineation, which helped achieve relatively high mapping accuracy for debris-covered glaciers. In subsequent years, however, we intentionally avoided using the RGI to account for glacier retreat, but our limited experience in manual sample selection led to insufficient representation of debris-covered ice, resulting in notable omissions and inconsistent delineation over time. This explains the substantial variability observed in debris-covered glacier extents between different periods. Good train dataset is vital in our approach.

We also recognize that the original figure suffered from low contrast, making some lines difficult to discern, particularly the red outlines on reddish backgrounds in the insets of Fig. 7. In the revised manuscript, we applied contrast stretching and adjusted the color schemes to improve visibility. Additionally, the discussion on mapping inaccuracies has been expanded to clarify the sources of misclassification and their potential impacts on the results.

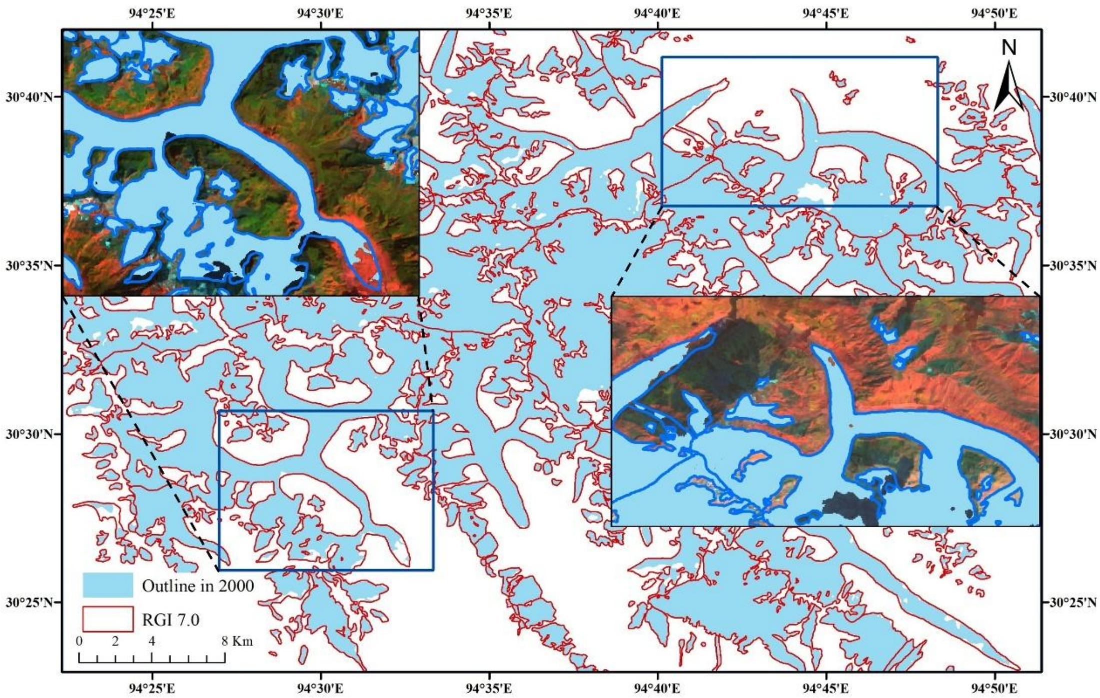

Fig2 Updates to figure 7

C: A final major point of concern is the general set-up of the study. First, the elevation change datasets are introduced in the Discussion Section 5.2 rather than in Sections 2.2 and 3. Their description is thus very short and the processing method unclear (e.g. how has the radar penetration into snow been corrected?). This can likely easily be adjusted. The motivation to determine annual area changes is mentioned, but not critically discussed. Even when the resulting glacier outlines would have been correct, a one-pixel (two pixels for debris-covered regions) uncertainty at 30 m resolution relates to a 30 and 60 m location uncertainty of the outline. With an assumed annual terminus retreat of 5 to 10 m / year (much less around the perimeter), one has to wait several years before new outlines make sense compared to uncertainties. But here the mapped termini could be wrong by several km, so change assessment is not an option. To make my major objection of the high variability in the mapping results clearer, I have added all glacier maps (setting the no data value to 0 before) and received a very colourful picture. On the last two pages of this review, I show a few examples for illustration. If the mapping had been correct, colours should only appear near the terminus and around the perimeter. As a note, this is just the result of a simple adding without a timeline, not revealing the partly strong year-to-year jumps in mapped glacier areas.

R: We appreciate your suggestions. For clarity, we specify that the datasets were derived from previous studies, and no additional processing was performed in our work. We also recognize that the original approach was somewhat aggressive in interpreting short-term glacier changes. Considering the positional uncertainties at 30 m resolution, we have revised our analysis to use 5-year intervals between mapped glacier outlines. This adjustment reduces the relative impact of positional errors and provides a more robust basis for assessing glacier changes over time.

C: In conclusion, this brute-force mapping using sophisticated image processing without a sufficient understanding of the mapped subject (glaciers) and how it should change over time is not recommended. When being harsh, I would ask the authors to please first learn the basics about glaciers and how they work, then proceed with the user needs (are annual updates really required?) and then do the mapping. A bit less harsh I would ask the authors to first get the mapping right for one year before applying it to several years. When glacier area changes are mostly due to changes in the mapping results rather than real changes, there is no need to perform change assessment. In my view, it is possible to publish a study revealing that a method has not worked. However, in this case an honest discussion and illustration of the problems is required to be helpful for future studies. Concluding that this study provides 'effective support for future glacier inventories' (L482) is in my view highly misleading.

R: We appreciate your suggestion and fully acknowledge that our original approach cannot perfect well for mapping individual glacier outlines. In response to your comment, we will remove any overstated claims related to annual glacier inventorying and clarified that our results reflect changes in ice-covered area. A more cautious error assessment has also been provided to better quantify the uncertainties involved in regional scale.

Specific comments

I do not comment here on all details of the study, but include some general remarks:

C: Providing area changes in km2 and mass changes in Gt (sure to use 900 rather than 850 kg/m^3 in this region?) are not useful as they are incomparable across regions. In future studies, please give relative area changes in % and the related change rates per year for area and specific mass changes per unit area and year for mass balance (and please do not use the latter to 'validate' the former, this makes glaciologically little sense).

R: We appreciate your suggestion. We agree that absolute glacier area and mass changes are difficult to compare across regions. In the revised manuscript, we now report glacier area changes as relative percentages and annual rates, and mass changes as specific mass balance per unit area per year. Regarding the density value used for converting geodetic volume change to mass change, the choice of 900 kg/m^3 follows one previous work (Zhao et al., 2022) to maintain consistency within regional research. *"Glaciers in the SETP are mainly maritime glaciers, which have a slightly larger density than continental glaciers. Therefore, a density of 900 kg m^{-3} was used"* (Zhao et al., 2022). We also acknowledge that a density of 850 kg/m^3 is more commonly used in the literature, and this point has been duly noted in the revised discussion.

Zhao F, Long D, Li X, et al. Rapid glacier mass loss in the Southeastern Tibetan Plateau since the year 2000 from satellite observations[J]. Remote Sensing of Environment, 2022, 270: 112853.

C: Please carefully check text errors. Often spaces are missing or units are wrong (area in km instead of km2, volume in m instead of m3). Also the citation style is strange. For example, in L398 it is written 'Ye et al. (Quinghua, 2019 ...)reported ...' So is it now Ye or Quinghua and why is it first et al. and then without et al.? In the reference section it is actually Quinhua, Y.E., again different. Correct would have been to write: 'The datasets by Qunighua (2019 and 2020) reported ...' or in L409: 'The results of Jakob et al. (2021)' … . As a small note, the References Section becomes more readable when indenting the text from the second line a bit, making it 'hanging'.

C: Figure captions: I suggest inserting a . or : after the figure number, e.g. ‘Figure 1: Study area’

C: Table 1: I think the brackets around the authors of the cited studies are not required.

R: We appreciate your comments. We have carefully checked the manuscript for text errors, correcting missing spaces, unit inconsistencies, and citation issues. All references have been verified and standardized. Figure captions now include a colon after the figure number, and Table 1 has been reformatted to remove unnecessary brackets around author names. A comprehensive review of all text, figures, tables, and references has been completed to ensure consistency and clarity throughout the manuscript.

C: L261: Figure 4: The blue and red lines and squares are difficult to see against the dark background. Also the annotations and legends of the insets are partly hard to see. It needs also to be explained what is what. Just writing in the text that types can be clearly distinguished is a bit thin. The same applies to all panels in Fig. 5. The panels are too small, the legends are unreadable and it is unclear what is what.

R: We appreciate your suggestion. We have enhanced the visibility of Figures 4 and 5 by adjusting line and marker colors to higher-contrast shades, enlarging the panels, and ensuring that all legends and annotations are clearly readable. The captions and main text have also been revised to explicitly explain the meaning of each line, marker, and panel, making the figures fully interpretable independently of the text.

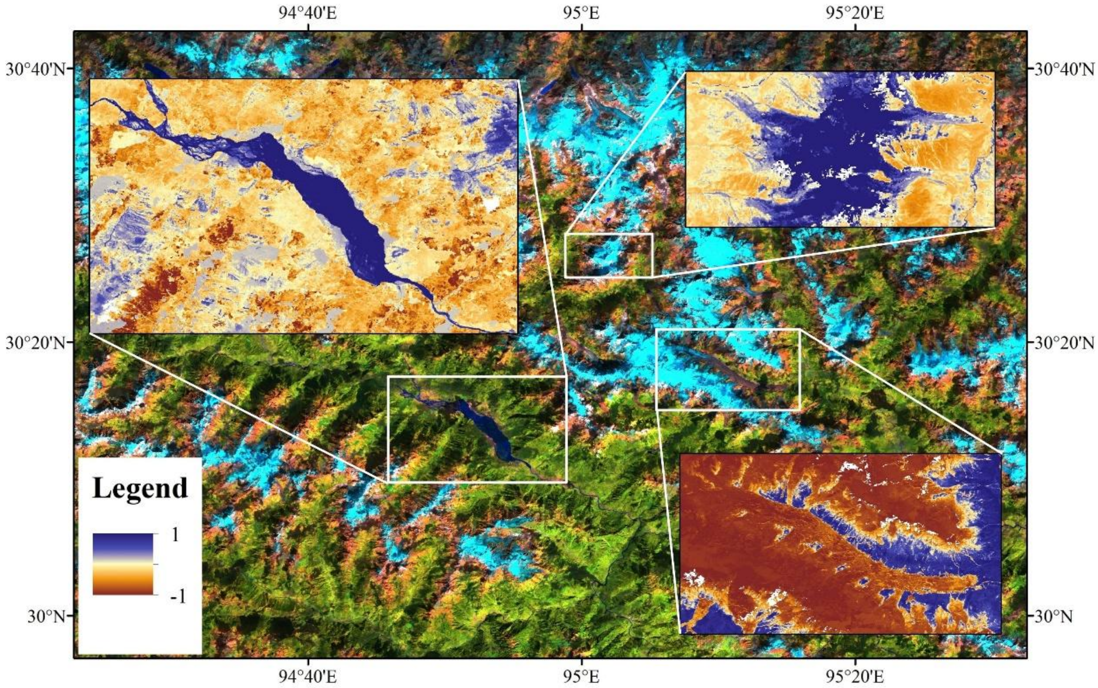

Fig3 Updates to figure 4

C: L306: I am also a bit unclear what Fig. 5 should tell me? That many datasets have been used and none of them shows glaciers clearly?

R: We appreciate your comment. Figure 5 is intended to show all the features used in our classification, highlighting the variety of datasets incorporated and their individual contributions to glacier mapping. In a future revision, we will overlay RGI glacier boundaries on each feature to make it clearer how each dataset aids in glacier extraction.

C: L390: Incomplete caption. Please note that sudden area gains as shown in Figure 5 (and 6) are glaciologically not possible. This is not how glaciers work.

R: We acknowledge your point. The sudden area increases shown in Figures 5 and 6 are indeed unrealistic. In the revised manuscript, we plan to reconstruct the glacier dataset at five-year intervals. Given the extent of these revisions, we kindly request additional time to carefully implement these changes.

C: L414: Figure 9b: This comparison makes glaciologically no sense.

R: We appreciate your suggestion. This part has been removed.

C: L470: This is correct, but the NDSI has been shown to be very sensitive to path radiance in the green band, creating problems with ice in shadow. Additional classification problems are introduced when using the analysis ready reflectance datasets instead of the raw data, which allow for a better separation of details when the SNR is low.

R: We appreciate your suggestion. We used decision-level fusion to address this issue, which removes cases where either Landsat or Sentinel-2 images are affected by shadows on the ice.

C: L471: As far as I can see it, most gaps are due to not mapping debris-covered glacier parts.

R: We appreciate your suggestion. For the frequency maps generated from our preliminary updated data, including the 2000 dataset greatly exaggerates apparent glacier retreat. This occurs because the 2000 classification samples were manually delineated based on the RGI dataset, enabling relatively accurate mapping of debris-covered glaciers. For subsequent years, limited experience in distinguishing debris-covered glaciers resulted in samples that are less reliable, causing the classification to miss a substantial portion of these glaciers.

Following the demo area you provided, we show the results of the initial modifications to the area:

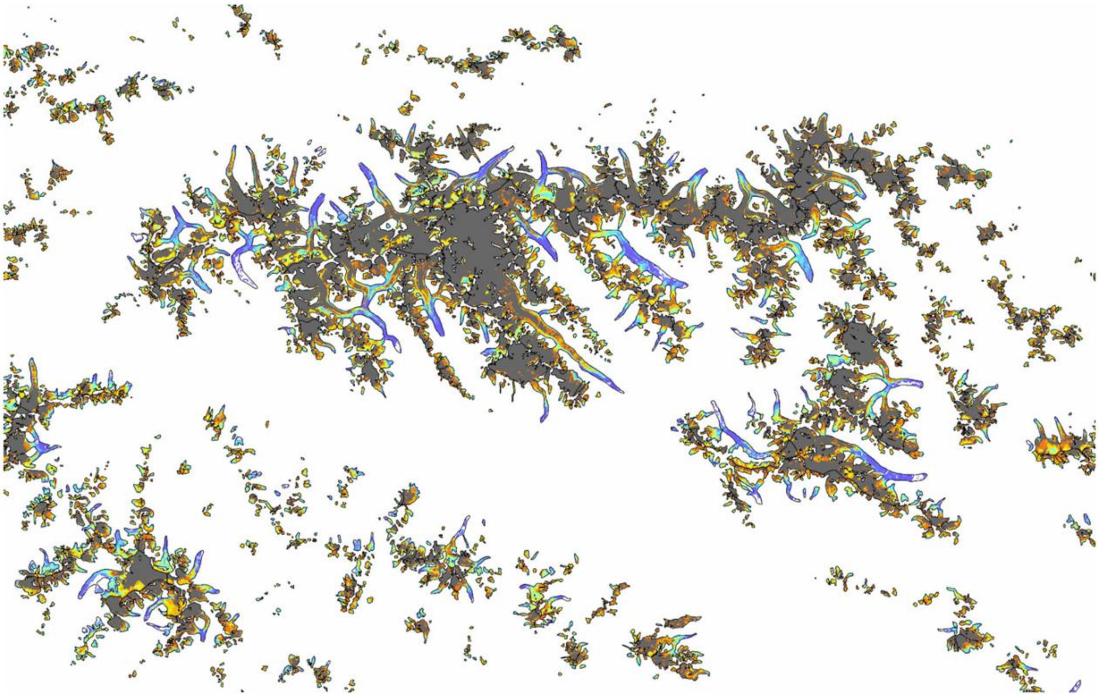

Fig 4 example3

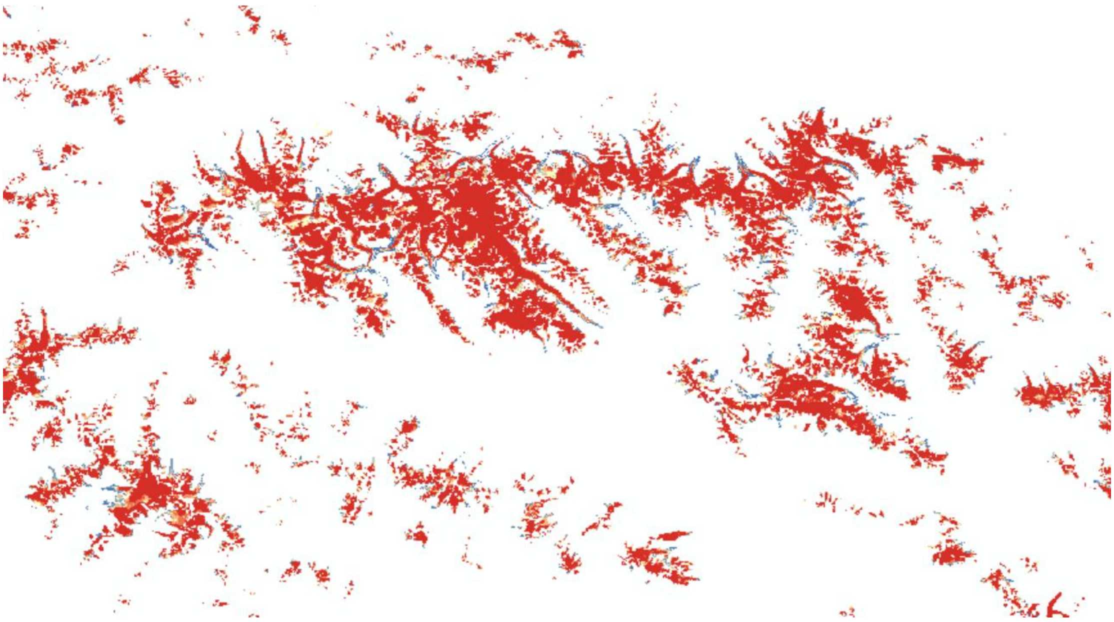

Fig 5：The same area as your example3, showing our modified results: a graph of glacier frequency for 2015-2022 year-by-year data.

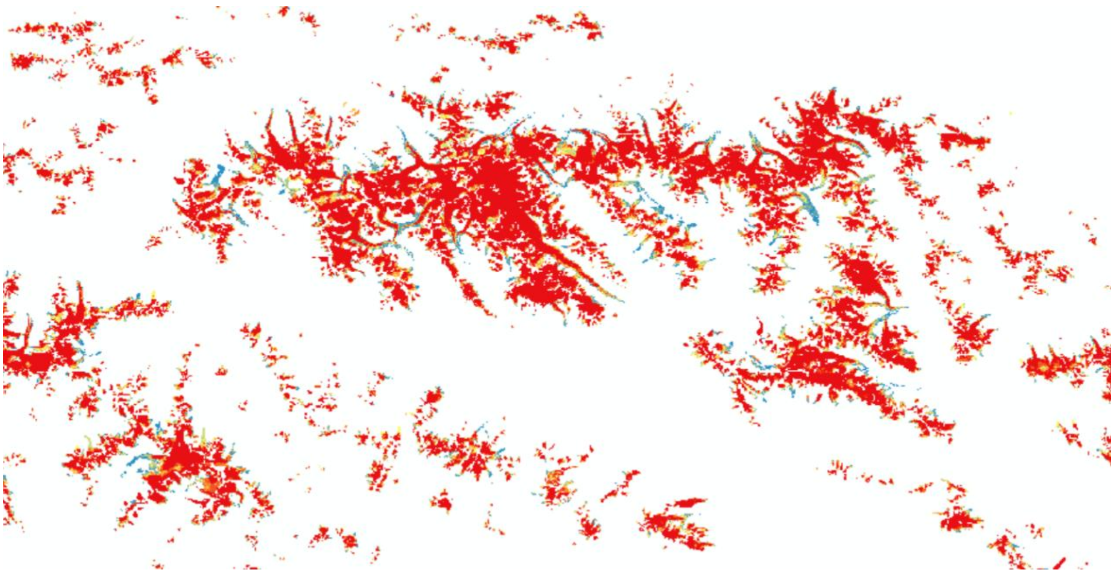

Fig 6： The same area as your example3, showing our modified results: Glacier Frequency Plots for 2005, 2010, 2015-2022 data.