

# Review of Manuscript egusphere-2025-1772: Multi-Source Remote Sensing-Based Reconstruction of Glacier Mass Changes in Southeastern Tibet Since the 21st Century

William David Harcourt, University of Aberdeen

August 26, 2025

Firstly, I would like to apologise to the authors for the delay in commenting on this manuscript. I hope my detailed comments below will help shape the paper in a further revision.

## Paper Summary

Li et al. have developed a random forest classifier to produce a new set of glacier outlines over the southeastern Tibet region. One of the key issues with mass balance estimates is that they rely on a single set of glacier outlines, usually from the RGI, which does not account for glacier terminus changes. The paper produces a new set of outlines for 2000, 2005, 2010, 2015 and then 2016-2022 at annual resolution. The produced outlines were determined to be of high accuracy as determined by the kappa score in a confusion matrix. Furthermore, the authors compare mass balance estimates using a fixed outline and the changing outlines derived in this study and find that there is a 10% difference, although it was not clear from the paper if this was an under- or over-estimate. Finally, the authors find that the loss of glacier area in the region has been accelerating, although no discussion was made on potential drivers (although this was not the aim of this study).

## General Comments

The paper produces some useful results, particularly around the use of dynamic glacier outlines for quantifying glacier mass balance. The methods are thorough and mostly well thought out with some caveats, and the findings appear to be of good quality, although some further information is required to improve understanding of these. There are several areas that require major improvement in a revised manuscript:

- The introduction requires a more detailed discussion of recent machine learning methodologies used to track glacier area and margin changes. More details are provided below in my technical comments, but the authors have missed a growing body of literature on this topic. This will help the authors better justify their choice of a random forest classifier used in this study.
- A more critical review of the features used to train the random forest classifier is needed. In particular, a sensitivity analysis will assist in understanding which features in the model are dominating the training. Furthermore, the authors used a broad range of features in the random forest classifier, hence it would be interesting to see if the model is overfitting in some way due to the diversity of input data. Quantifying this would help improve reliability in the final results.

- The methods section is verbose and could be shortened significantly. This will allow for more space to discuss model performance later in the paper.
- It would be useful to understand the performance of the random forest classifier in different contexts. In particular, how does it perform for different satellite images e.g. Sentinel-2, Landsat 7, Landsat 8 etc. Currently, the uncertainty is taken as 1 derivative of the pixel size, but it should really reflect the accuracies of the glacier outlines which will vary with different data sets.
- A key outcome of the study is the impact of dynamic glacier outlines on mass balance calculations, but this is not explored sufficiently in the study. I would urge the authors to present these results more fully and discuss the implications of this for mass balance studies in Tibet and the wider globe.

There are several typos and gramatical mistakes throughout the paper, some of which I have highlighted in my technical comments, but I would encourage the authors to thoroughly review the manuscript upon revision.

## Technical Corrections (References to line (L) numbers in preprint)

L10: Better to say ‘glacier area’? Also, the latter part of the sentence only applies to optical data.

L12: ‘the Landsat satellite series’

L14: ‘for this region’

L15: ‘integrating a three-year dataset’ isn’t clear to me - do you mean delineating glacier area for 3 years and then the median year is taken to be the time satmp?

L19: ‘we calculated glacier mass balance’

L20: ‘glacier areas calculated in this study, resulting in an annual mass loss of 6.20’

L33-35: Does this sentence refer to the Tibetan Plateau specifically? If so, can the authors state this.

L36: ‘hence the region is dominated by maritime glaciers’

L37: What glacier changes? The natural cycle of accumulation/ablation or a longer term trend? This is not clear.

L39-52: The description of NDSI could be improved e.g. the use of a manual threshold is only mentioned at the end. The authors state a weakness is lack of automation, which is true, but there is a wider point that the application of NDSI varies in different geographic regions, which makes it hard to automate the process. This should be acknowledged.

L39: ‘Glacier area mapping from satellite imagery’

L40: ‘substantial time for human interpretation’

L43-44: Extracting what component of glacier and snow cover? Area changes? Differentiating between the two surfaces? Probably both.

L53-63: This is quite a vague paragraph that misses a lot of important studies mapping glaciers with ML e.g. for terminus mapping, glacier area estimates and surface features (e.g. ???). There is a growing body of literature in this field and this should be acknowledged with a more detailed literature review in this section.

L53: ‘Recent developments in machine learning algorithms have enabled large volumes of satellite imagery to be used as training data for automated classification of glaciers’ - or something like this. It’s important to be clear what ML does and how it improves over the traditional techniques.

L64: ‘from optical satellite imagery’

L70: Define ‘high temporal resolution’ - either weeks, months, seasonal or years.

L72-77: What are the details of this inventory? What is their estimate of the number of glaciers, area etc.?

L72: ‘Qinghua,2020), who’

L91-98: I would like to see a bit more discussion of the importance of glacier inventories (e.g. areas) for quantifying mass changes e.g. how do the GLAMBIE/IMBIE community estimates tackle this problem and what is the consensus approach when multi-temporal data sets aren’t available? What is the impact on uncertainty estimates? This will naturally then lead onto the objectives in the paragraph.

L109: For those unfamiliar with this region, it might be worth zooming out a bit and placing an inset map to show the position of this region in the wider regional context.

L113: What does ‘glacier distribution area’ mean?

L124: This section is not consistent - sometimes the sampling is described, in other sections it is not. Either describe the sampling within each section or create a new section where it is fully described.

L126: What does ‘analysis-ready’ mean? What processing has been applied before these images are provided on GEE?

L139-145: Given the introduction focuses on the limitations of optical data, the authors should discuss somewhere the pros and cons of using SAR data as an alternative.

L146-153: What is the time stamp of the NASADEM? Or is it a dynamic data set?

L155-161: Time stamp of 2000 for RGI7.0.

L167: Vague - define exactly in which period the data were acquired. If  $T$  is the sampling year, did you obtain all suitable summer images in years  $T \pm 2$  years?

L201: Do the image data cubes represent the ‘image composites’ described above? It would be useful to have consistent language throughout the manuscript to avoid confusion.

L203: ‘Spectral reflectance alone is insufficient’

L204: ‘this study extracts spectral, terrain, texture, and radar interferometric features to train a Random Forest classifier for delineating glaciers in satellite imagery.’

L210: I’m confused here, how do Figures 4a-f represent cloud-free image composites?

L226: Which images are used to generate the NDVI image for each? Did you merge the NDVI values for a single year?

L233: Same point as for NDVI, not clear to me which images are being used to calculate this.

L242: Same as for L226 and L233.

L248: I am confused by this figure - I assume each of the horizontal squares represents an image, so what do the colours represent? And what do the vertical boxes represent

L266-265: Image textures are better defined as the spatial arrangement of pixels in an image

L277: Is this the mean texture from GLCM? It's not clear why this was chosen - the authors state that a previous study found it is 'consistent with other textures' - why would this mean it is the best feature to use? If it is consistent with other features, then any other texture feature could be used e.g. autocorrelation, entropy etc.?

L287: What is a 'mean synthesis'? Also the 'salt-and-pepper noise' I assume is referring to 'speckle' - calling it noise is incorrect as speckle is a repeatable feature in SAR data.

L291-304: Without a suitable figure (ie. Figure 5), it is difficult to interpret the feature layers described in this section. The inclusion of RGI outlines would help, but also subtitles and a larger legend will help readability.

L292-293: This is not clear in Figure 5, see comment below.

L305: Figure 5: It's not clear what the values represent, the legend is way too small. One legend for all composites is sufficient, unless the values are significantly different between each panel. It would also be useful to overlay the RGI outlines here so the reader can visually assess how well each image feature matches the glacier area. Also, if this is referenced before Figure 3, it should also be first in the order of figures.

L314-319: RF has been widely used, although arguably it has been superseded by CNNs and foundation models. Can the authors comment on why they did not apply these other methods?

L322: Are the labels used for all images or a subset? For the images labelled, are the labels shown in Figure 6 suitable for all images given the potential for changes in surface characteristics at different times of the year?

L327-336: Are you discussing here the training data, validation data, or both? Subtitle is misleading, 'Selection of Classification Samples' doesn't really say anything here. How many images where the training data taken from?

L338-342: F1 score might be more suitable here if there is class imbalance - I suspect there is imbalance in the training data, but it is not stated.

L358: What is meant by a 'decision-level fusion strategy'?

L371: '4 Results'

L372-380: This is a surprisingly short section that only gives the headline figures. I would like to see a sensitivity analysis of the random forest classifier, particularly an understanding of which texture features were more important for classification than others. One possibility of using a diverse range of features is that the model could be overfitting, potentially leading to errors in the resultant classification maps. Furthermore, how do the accuracies compare for different data sets? I would expect there to be differences in Sentinel-2 vs Landsat, whilst Landsat 7 would likely yield different accuracies to Landsat 8. This information must be included to better understand the performance of the technique.

L373: Referring to the 'annual' classification results, I assume you mean for the results after 2016 with the Sentinel / Landsat results? Furthermore, the authors should show here the confusion matrix to better highlight true positives, true negatives, false positives and false negatives. A single accuracy score may be misleading.

L386: Why is the mapping error based on half the image element? Surely the graph should be representing uncertainty calculated from the random forest model outputs?

L395-405: This section is a bit confusing. It might be helpful to construct a table with the key results from previous studies to make it clear how the results in this paper compare?

L397: 'It is noted that'

L406-419: This reads like a results section - also, the inclusion of ICESat, ICESat-2, and CryoSat-2 should really be discussed in the methodology section. Are the data sets extracted simply just the time series as presented? Or did the authors process the data sets in some way? I also don't think thickness and area should be presented on the same graph, it might cause confusion - I would use 2 panels instead.

L423-424: The variables in the equations need to be stated.

L430: Missing word, glaciers are losing mass at a rate of 6.20 Gt/y?

L432-434: This is an important result, but it is not shown graphically. Can the authors make a figure showing this key result? Although, what does the 10% refer to - an under- or an over-estimate compared to the fixed outlines?

L420-447: These are results, and the methods described here should be presented in a methodology section. The small discussion towards the end should be expanded particularly focusing on the importance of updated glacier outlines for mass balance estimates, as this is key moving forward in future studies.

L473: Define the number of images used and over what time period

L472-483: I would expect the conclusions to mention the performance of the random forest model as well.