Review of 'A gradient-boosted tree framework to model the ice thickness of the world's glaciers (IceBoost v1)' by Maffezzoli et al.

December 2, 2024

In 'a gradient-boosted tree framework to model the ice thickness of the world's glaciers (IceBoost v1),' the authors present a method based on random forests for estimating ice thickness globally. The method differs substantially from more classical methods for ice thickness estimation in that it does not utilize principles of glacier-wide mass conservation. Instead it is trained on a large number of geometric and other variables calculated locally. The results are encouraging, especially when the model is trained on data from the same glacier on which it is making predictions.

I think this is an interesting paper that may be useful to some practitioners who are interested in estimating global glacier volume. The methods make sense and it is clear what is being done. I would like to see a somewhat more comprehensive description of some of the method's shortcomings. For example, I think a better job could done of trying to understand in what circumstances the method performs well versus not – the Malaspina example could do this, but instead we just get 'works okay here, but look at this other example where it works better!'. I also would like to see at least a discussion of the suitability of the resulting product for different tasks, e.g. as input for modelling, which this will invariably be used for, whether its suitable or not. It would be nice to hear a bit about predictive uncertainty - even if it's just to say that you don't know because the model doesn't produce that. As a small concern, it would be helpful to add a section describing why – not just what – features were included in the analysis. It's not immediately clear why having slope smoothed over 8 different length scales (some of which aren't so different) is a good idea. Finally, after all the fuss made over this being globally applied, why not provide an estimate of global ice volume and compare it to previous works?

Detailed comments are below.

L3 Here and elsewhere, I think the word 'distribution' needs to be used with more precision. In particular, I think it would be better to say that the method evaluates the thickness at arbitrary coordinates and can produce a map.

L28 A comment on this section - it's not necessarily the case that other methods for the 'global' problem don't exist, it's just that many people working in this field don't believe that it's defensible to produce such estimates because input observations - e.g. surface mass balance - are not reliable.

L24–39 It is worth noting potential limitations of these methods (of which there are many). What does the present work bring to advance beyond these previous efforts?

L46–48 This is an awkward sentence, and it won't be clear to the lay reader what the distinction is between image and tabular data, particularly since we almost always look at thickness as maps (i.e. images).

L57 The absolute number of training data 'points' isn't all that relevant because they are adjacent to one another and many (perhaps even most) of these data points are strongly correlated with one another.

L60 'The DEM choice ...' I don't really know what this line means.

L61 I suggest being more explicit about what these features are. For example, instead of calling RACMO 'mass balance information', it would be more helpful to describe it as 'modelled estimates of spatially-distributed surface mass balance'.

L77 I think that this line about BedMachine will only make sense to someone who is already very familiar with how BedMachine is constructed.

L110 Symbols are arbitrary of course, but glaciological literature usually uses H (or sometimes (h) to represent ice thickness.

L113 This decision makes sense because you include features that already encode this regional variability (SMB, elevation, etc.). It would be weird to include these things and then also need regionally stratified models.

L121-L136 I don't think this section does a good job of justifying not holding out a validation set - after doing 200 trials, the hyperparameter choices ought to be robust to randomness in the splits because there are so many of them. If these hyperparameters don't perform well on another independent data set, then the model isn't very good.

L137 I don't think 'accuracy' and 'precision' should be used here. Just state the metrics that are used (median residual and RMSE). Also, why these two?

L146 I think I understand what this paragraph is saying, but I would like to see this clarified. Are regions exlcuded from analysis because there's no data or because the performance isn't good? What does it mean for performance to be 'indicative'?

L150 This is an awkward jump to start talking about the combined models when there has yet to be any description of the performance of the individual models.

L153 The more cynical take regarding the better performance of the 'supervised' model is that the model is overfit and only performs well when it can look up memorized observation points that are close in space to the query.

L161 Shouldn't this appear in some kind of 'data availability' statement?

L209 Malaspina is currently land terminating, so its terminus is mostly grounded at or above sea level. When using 'terminus' do you mean the piedmont lobe?

L227 I am not sure how this argument is justified. Is it possible to be more quantitative?

L234 'Conversely, the features not based on satellite products are not discrete'. I don't understand this sentence.

L236 There are fundamental limits to the resolution of bed features that can be determined solely from surface observations due to the diffusive nature of ice flow.

L267 These conclusions are fine, but please write them in narrative form, rather than as bullet points.