

## ***Interactive comment on “Glacio-hydrological melt and runoff modelling: a limits of acceptability framework for model selection” by Jonathan D. Mackay et al.***

### **Anonymous Referee #1**

Received and published: 19 February 2018

Review of Glacio-hydrological melt and runoff modelling: a limits of acceptability framework for model selection by J.D. Mackay et al.

This paper presents a truly comprehensive investigation into the performance of a set of glacio-hydrological models of varying complexity. Such models basically take climate and topographic data and calculate river discharge via model calculations of distributed glacier melt, linked to overland and river flow models. The models investigated here are all at what might be judged as the simpler end of model configurations used within glaciology, as the melt models are variations on the well-established temperature index approach, and the runoff models are all reservoir models, again of varying complexity.

C1

The key advance in the paper is the use of a comprehensive range of output measures used to judge model performance, and within that, the adoption of a sophisticated estimate of the limits of acceptability for each performance test based on uncertainties and errors within these datasets due to intrinsic errors or errors induced by sampling, spatial scale and the like. This is a powerful advance over most previous studies, which have used a limited set of performance measures, normally judged with rather simple statistics for model agreement, such as correlation or the various mean error statistics available. I think the argument made that such simplistic model/data comparison approaches have limited glaciological studies in comparison with other model disciplines, such as hydrology, is powerful, although it has been addressed occasionally within glaciology before. Rye et al. 2012, for instance, show how there is a need for multi-objective optimisation within melt modelling, as no single measure of model fit is adequate to fully capture the performance of any given model, and different outcomes, in terms of longer-term mass balance predictions are possible with equally acceptable models. The methodology used by Rye et al., and the overall scope of their investigation is different to the current study, but the over-arching aims seem very similar.

The paper sets out a clear methodology to develop the limits of acceptability for each ‘test’ dataset. As examples, the LOA as defined for ice melt include the need to allow that model-based estimates of melt when compared with in-situ stake ablation measurements can only be expected to be as close as the actual spread in melt over the equivalent node area within the model. This spread was calculated using high resolution terrestrial LIDAR scans made during the stake measurement campaign, and used to define 95% confidence bounds around the stake measurements. MODIS-based (and therefore 500m spatial resolution) snow cover estimates were transformed to use within the 50m resolution DEM of the study site with a monte-carlo approach in which for each MODIS snow pixel the relevant number of 50m snow pixels were distributed randomly 1000 times to generate confidence bounds around the snow cover-elevation measure used as a model test. Such processes were adopted for 33 data ‘signatures’.

C2

The comparison exercise was based on the use of three variations in temperature index model (of differing complexity), run with each of three runoff models, giving 9 model configurations. Each was then run with a randomised range of parameter values, yielding 45,000 calibration runs to be judged against each of the 33 'signatures'. This set of results is evaluated both in terms of measuring the various models' performance against the range of signatures, and also in terms of the signatures' discriminatory power. Ultimately, no single model nor set of parameters met all of the LOA across all the signatures, but the approach was capable of showing the 'trade-offs' within model configuration/parameterisation in terms of the different patterns of acceptability identified, and it also showed that additional complexity did not always lead to improved acceptability.

This might be seen as a limitation of the current study. It may be that such a comprehensive analysis of a range of models and data signatures would lead to a different 'specific' conclusion in terms of the best-fit model, or the most powerful discriminatory data, when applied to a different glacier system. I do feel, however, that whilst such a complex study as this one is not going to be possible for every model application, a more sophisticated methodology to evaluate model fit is important within glaciology. I think a key specific outcome of this study (which to my mind merits publication by itself) are the descriptions and methodologies used to generate the limits of acceptability for the range of data available. These will form a useful resource for future studies which may adopt this, or similar, methodologies. Additionally, as well as potentially improving the discriminatory power of models and our ability to discriminate between different model performance, work such as the current study, and a few earlier papers, will also allow a better understanding of the uncertainty inherent in model predictions. I think this is a careful and extremely thorough study, and I strongly recommend it for publication. I hope it gains impact and this type of more sophisticated evaluation of model performance gains traction within the glaciological community.

The paper is also commendably well written. I have very few specific corrections which

C3

I feel should be made.

P2 L10 '...adhere to, ...' (insert comma)

P3 L22 'therefore we' (delete comma)

P3 L24 'definition imperfect' (delete comma)

P5 L15 change 'on' to 'of'

P9 L 25 delete 'an' before 'can'

P15 L1 Suggest rewording to 'As in many glaciated catchments, topography controls spatial temperature gradients to a large extent'.

P15 L7 Suggest change 'shallow' to 'reduce'

P18 L 6 'curves provide' (delete comma)

P24 Figure 8 caption or key needs to include explanation of the dots on Fig 8a

P40 L16 'all but two of' (delete both commas)

Reference Rye, C.J., Willis, I.C., Arnold, N.S. and Kohler, J., 2012. On the need for automated multiobjective optimization and uncertainty estimation of glacier mass balance models. *Journal of Geophysical Research: Earth Surface*, v. 117, doi:10.1029/2011JF002184

---

Interactive comment on The Cryosphere Discuss., <https://doi.org/10.5194/tc-2017-268>, 2018.

C4