Referee's comments are in blue, our reply in black, quotes in the revised manuscript in red.

Review of Wang et al. (2025) 'Quantifying Temperature-sliding Inconsistency in Thermomechanical Coupling: A Comparative Analysis of Geothermal Heat Flux Datasets at Totten Glacier'

Summary

This paper presents a new set of metrics to assess the inconsistencies between modelled basal temperatures and observed surface ice velocities in ice-sheet simulations, and applies them to evaluate the likely shortcomings of eight different geothermal-heat-flux datasets in the Totten basin in East Antarctica. The paper finds that its results in this regard agree with previous work that assessed the eight datasets using radar specularity observations to determine the presence of basal water and thus whether the ice was warm or cold, which suggests it is indeed performing well in identifying problem areas.

I think the new method is a useful way to quantify the mismatches between modelled ice temperatures and observed velocities, and the paper is well-executed. I do think it would benefit from a small amount of restructuring, as presenting the application of the method and the comparison of the geothermal datasets at the same time makes it harder for the reader to understand what's going on. I also find it dissatisfying that the method cannot say anything about the source of the inconsistencies, however. Pointing out where the inconsistencies are is, of course, a useful first step, but without knowing where they're coming from, there's not necessarily much one can do with the information. That would obviously be too much for this paper though, and the authors suggest this might be the focus of future work – I would strongly encourage them to pursue it, as that would be really very useful!

Overall, I think the paper is suitable for publication subject to minor revisions to improve the clarity of the presentation.

Reply: Thank you for your encouraging comments.

Page and line numbers refer to those in the clean version of the submitted manuscript.

Major Comments

• Focus of the paper: This is a a bit of a minor gripe, but I feel the focus of the paper is not always very clear: is it presenting the new set of metrics or is it assessing the differences between the GTF datasets? I think it would be clearer to start with one simulation where the authors show how the method works in practice, and then move on to the eight simulations comparing the GTF data. As it stands, the authors are trying to do both things at the same time and the consequent seesawing makes it harder for the reader to understand what the metrics are showing in concrete terms. This need not be a new simulation or anything – just pick one of the eight and add a

short section to the results showing the fields for the metrics and explaining the interpretation, before launching in to the full bore comparison.

Reply: Thank you for your comments. We add a section in the revision. We show the spatial fields of the inconsistencies metrics (Section 2.1) for the modelled result in Huang et al. (2024) with Martos et al. (2017) GHF as an example, and provide an interpretation, before conducting a comprehensive comparative analysis for the result with 8 GHF datasets.

Here is the section we add:

3.1 Spatial Distribution of Inconsistencies with one GHF dataset

In this section, we show the spatial fields of the inconsistencies metrics (Section 2.1) for the modelled result in Huang et al. (2024) with Martos et al. (2017) GHF as an example, and provide an interpretation before conducting a comprehensive comparative analysis for the result with 8 GHF datasets.

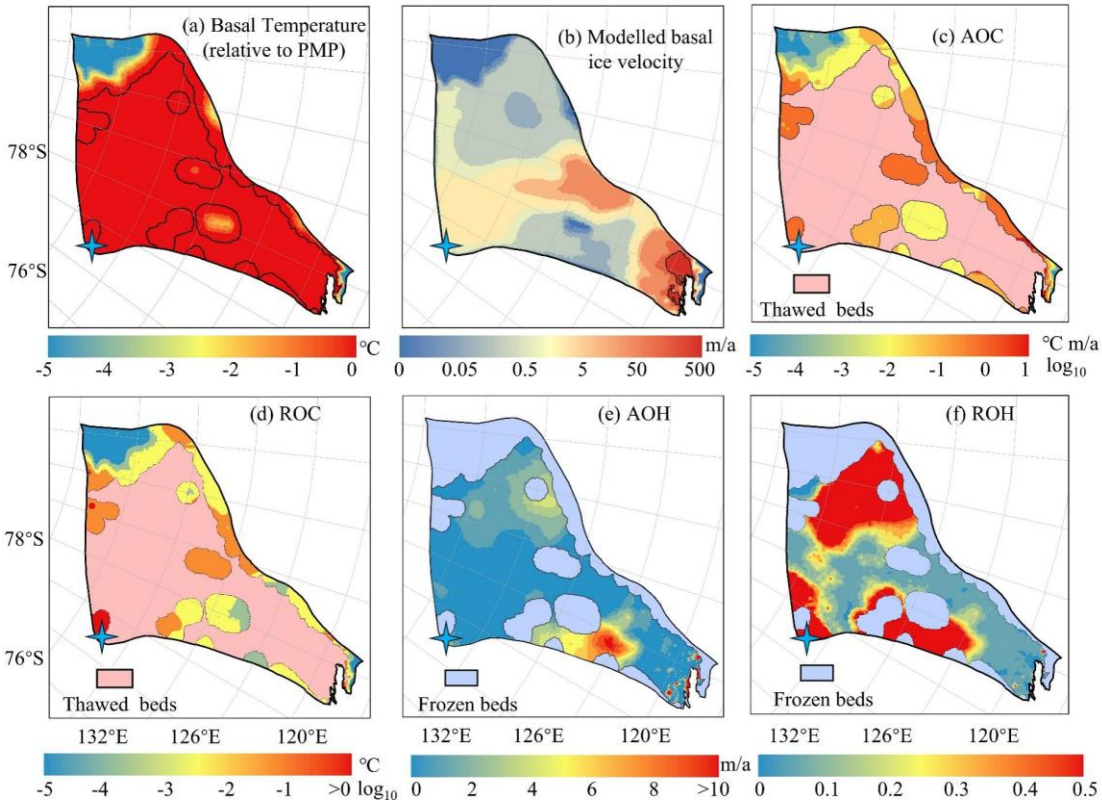

Figure 3. Spatial distribution of modelled basal ice temperature using the Martos et al. (2017) GHF. (a), modelled basal ice speed (b), *AOC* (c), *ROC* (d) inconsistencies in modelled cold-bedded regions, and *AOH* (e) and *ROH* (f) inconsistencies in modelled warm-bedded regions. The colormap in (c) and (d) is on logarithmic scale. The pink region in (c) and (d) represents modelled thawed bed, while the blue region in (e) and (f) indicates frozen-bedded areas.

The modelled result based on the Martos et al. (2017) GHF reveals extensive

regions of thawed bed with limited areas of frozen bed. The frozen bed is predominantly located in the southern corner of the study domain, where the modelled basal ice speed approaches zero, consistent with cold basal ice temperature. Consequently, the AOC inconsistency at this marginal zone is negligible (Fig. 3). Along the western margin of Totten Glacier, basal ice temperature remains below the pressure melting point, albeit approaching it. However, localized regions exhibit high basal velocities of several tens of meters per year, contradicting the presence of a frozen bed and resulting in large AOC inconsistencies.

On the other hand, large AOH values are observed between 69°S and 71°S in the eastern Totten Glacier region, where the simulated surface ice speed exceeds observational data by >5 m a^{-1} (Fig. 3e). In this area, the modelled basal ice temperature reaches the pressure melting point, with the modelled basal ice speed at approximately 0.05 m a^{-1}. Basal friction inversion failed to reproduce observed surface ice speed due to the model's overestimation of ice temperature and softness. This pronounced velocity mismatch highlights a fundamental inconsistency in the eastern glacier region, likely originating from discrepancies in the input datasets. Regions of high ROH and ROC values coincide with areas of relatively high AOH and ROC, particularly where the observed surface velocities are slow, as per their formulations.

• Future steps: This paper is very much a first step – yes, knowing where the inconsistencies are is good, but knowing why they're there is a lot more useful if one wants to improve one's ice-sheet model. I know the authors hint at this being a possibility for future work, but maybe they could suggest in the discussion some ways in which modellers might use the method to investigate the causes of inconsistencies? This would make the paper a lot more directly useful to the ice-sheet modelling community, one feels.

Reply: Evaluating the inconsistencies is like doing basal friction inversion. It tells us where is the misfit, but it alone cannot tell us why there is a misfit. Anyhow, there is a procedure that could be worth checking. So we make a list in the discussion that one could check in the revision.

Evaluating the inconsistencies reveals where mismatches occur but not why they arise. The primary factors to investigate are surface temperature, GHF, accumulation rate, and ice thickness, representing the most critical boundary conditions. Furthermore, integrating multiple sources of prior knowledge can help constrain model parameters:

1. High-resolution radar measurements: Check availability of ice thickness data along flight lines to validate geometric boundary conditions.
2. Paleoclimate context: Historical climate reconstructions indicate significantly colder surface temperatures during glacial periods compared to present-day conditions, with correspondingly lower accumulation rates. These paleo-temperature conditions likely induced a long-term thermal memory within the ice column, potentially contributing to observed discrepancies between modeled and measured basal properties.

Therefore, we recommend systematically evaluating:
(1) Spatial distribution of radar-derived ice thickness measurements;
(2) Temporal consistency of surface temperature boundary conditions;
(3) Sensitivity of model results to GHF;
(4) Accumulation rate reconstructions during key climatic periods.
This multi-faceted approach helps isolate the causes of inconsistencies in ice sheet simulations.

Minor Comments

• p.1, l.30-35: I can't quite work out from the phrasing in the abstract for the first inconsistency if this means the GHF datasets are too cold or whether the model's too cold. Similarly, with the second one, are you saying the model is overheating compared to the GHF datasets, or that the datasets themselves are too hot? Or both? Obviously, it's explained later, but best to not confuse people with the first thing they'll read! Some rephrasing to make it clearer what's going on here would be beneficial.

Reply: We rephrase these sentences

"Examples of the method utility are 1. an inconsistency characterizing overcooling with all GHFs near the western boundary of Totten Glacier between 70°S-72°S, where there is a bedrock canyon and fast surface ice velocities, which suggests that GHF is low in all published datasets; 2. an overheating inconsistency in the eastern Totten Glacier with all GHFs that leads to overestimation of ice temperature due, in this case, to an unrealistically warm surface temperature."

to

"Examples of the method utility are 1. an overcooling inconsistency with all GHFs near the western boundary of Totten Glacier between 70°S-72°S, where there is a bedrock canyon and fast surface ice velocities, suggesting that all GHFs are low; 2. an overheating inconsistency in the eastern Totten Glacier with all GHFs suggesting overestimation of ice temperature, in this case, due to an unrealistically warm surface temperature."