Summary: The paper investigates inconsistencies in ice sheet models when the basal thermal state is inferred from ice surface velocities. The authors derive six metrics with which they quantify these inconsistencies, apply their method to Totten Glacier, Antarctica, using eight geothermal heat flux models, and compare their results with radar specularity. The authors further use the discovered inconsistencies to rank the eight geothermal heat flux models in their reliability. I commence the authors for finding this novel approach and I see the added value in having an evaluation method that can be applied rapidly and with relatively little input data. However, there are inconsistencies with terminology, methodology, and clarity of the manuscript so that revisions ae required.

I have two major comments and a number of minor comments that are outlined below:

Major:

1. The manuscript would benefit from a clearer structure. As the paper combines a multitude of modelled and observed data, as well as several different metrics, it is hard to follow. The results section in particular is very hard to follow as it switches between GHF products, absolute and relative metrics as well as overcooling and overheating. I suggest the following:
   a. Clearly define your input datasets and metrics in the methodology section. Use different subheadings for a) Definition of Metrics B) Normalization and ranking. The section could also benefit from a table that shows all input datasets.
   b. While the maps are useful to assess the spatial distribution of different metrics, I would also add a table that shows the key differences for each metric and GHF.
   c. Structure the results section by metric with individual subheadings e.g. absolute inconsistencies, relative inconsistencies, comparison.
   d. Section 4.3. should be more on the caveats as it mostly discusses the influence of near-surface air temperatures on your results. I would change the subheading to e.g. "Impact of Input datasets", and rather than calling it an additional experiment, state that a caveat of your study is that it is influenced by the input data.
2. At various times in the manuscript, the authors label datasets and values as unrealistic, gold standard or otherwise. These labels are unscientific and should be removed (see more detailed comments below for occurrences that I have spotted).

Minor:

Throughout the manuscript the authors switch between surface velocity and surface speed. For consistency, it should be one or the other.

Line 85: Insert "the" before basal friction coefficient

Line 125 – 132: This paragraph could go into the introduction as you define inconsistencies there. I would, however, remove the sentence in Lines 92-94 and rephrase the sentence in Lines 125 – 126 to the following: "*For this study we define inconsistencies as differences between modelled frozen bed and modelled basal sliding (which is tuned to match the observed fast surface velocity during the inversion), and between modelled warm bed and observed slow surface velocity. The inconsistencies originate from multiple causes, including uncertainties in GHF, surface ice temperature, ice sheet geometry, bed topography, surface velocity, ice density and incomplete ice flow mechanics.*"

Lines 155 – 156: This sentence is unnecessary. Consider removing.

Lines 197 – 199: This sentence can be incorporated into the next one e.g. *"We obtain three absolute inconsistencies (AOH, AOC, ACI) and three relative inconsistencies (ROH, ROC, RCI), with which we can comprehensively analyze ..."*

Line 230: It is not quite clear from the map where 71°S is as the map only shows 76°S and 78°S. The map should either show the coordinates referred to in the text or the area should be highlighted somehow.

Line 234: Being colder than what? I assume than the other GHF products. Consider adding *"... than the other four GHF products."*.

Line 240: The canyon is not really apparent in Figure 1b - Consider adding an outline.

Lines 244 – 245: This should go further up in the methods section where you define AOH and AOC.

Lines 246 – 247: Again, it is not clear where the area referred to in the text is located as the coordinates in the map do not correspond.

Line 254 – 256: This sentence refers to ice flow and references Figure 3, which does not show ice flow. I would suggest referring to Figure S2 for ice velocity and Figure 3 at the end of the sentence.

Line 261 – 263: I would expect the spatial distribution of the two metrics to be different as they are derived differently. Consider removing the sentence as I think it is not providing any additional information. If it stays in the manuscript, it should be absolute "overheating" inconsistencies in Line 262.

Line 263: Mention how to find Dome C in the figure (e.g. Dome C (Blue Star, Figure 4).

Figures 4 – 6: Keep the wording consistent. Where the color scale is logarithmic, say that instead of "non-linear". Consider changing the color for the Dome C marker – It is very hard to see. Also, if Dome C is marked in all figures, it should also be referred to in the figure caption.

Lines 282 – 284: Repetition from Lines 216 – 218. Consider removing.

Line 287: Again, the coordinates are meaningless if the map doesn't reflect them.

Figure 7 a,b: Not the best colors for colorblindness. Consider changing (check here: https://colorbrewer2.org/)

Lines 345 – 347: Repetition from Lines 125 – 132. Consider removing.

Line 349: "check" sounds a bit informal. Maybe use "assess".

Lines 361 – 362: This is a very strong statement, and I don't think you can say that unless you have a product that captures GHF correctly or a citation to back it up. Consider removing or at least toning it down.

Lines 367 –368: In contrast to what - Not quite clear why this sentence follows the section on uncertainty metrics. Maybe it was supposed to go after Line 365. Also, it should be "in contrast" not "by".

Line 376: Considered "gold standard" by who? Either add a citation or refrain from using ratings.

Line 395: Insert "is" between "impact" and "beyond".

Line 405: Remove "we know for sure".

Figure 8 – caption: Change "thick black curves" to "black lines".

Line 480: Remove "the" before "dynamic ice loss".

Lines 481 – 482: Rephrase. Currently, the sentence doesn't make any sense.

Lines 482 – 483: This sentence needs a citation.

Line 494: Change "under climate change" to "under future climate change scenarios".

Line 501: Change "checking" to "assessing".

Line 507: This is the first time you talk about englacial temperature. I assume you mean basal temperature.

Line 509: Which simulation results? Add a citation or refer to a specific simulation.

Line 511: Again, this is a strong assumption unless you know what is realistic. Maybe use "due to differences in".

Lines 518 – 519: See comments above on coordinates.

Line 519: Find a different word unless you can prove that is not realistic

Line 520: Remove "there".

Supplementary material

Figure S1 – caption: Add abbreviation for pressure melting point. It is otherwise not clear what PMP in the legend stands for.

Figure S4: The labels should include the year of the publication as is present in all other figures (e.g. Purucker (2012) – Lösing and Ebbing (2021))

As you are citing in the figure captions, you should add a reference list.