In this manuscript, the authors describe a newly compiled database for measurements of the geothermal heat flow (GHF) in and around Greenland. Based on these data and additional, geophysical information, they use a machine learning technique to assemble a gridded GHF map for the entire domain at a nominal resolution of 55 km. Several corrections are discussed, and a comparison to other, recently published GHF models is given.

As the authors correctly point out at the end of their summary remarks, increasing the level of understanding of Greenland’s GHF is of great scientific relevance. Therefore, the current study is highly welcome. I applaud the enormous effort that went into putting together all this information and the subsequent analyses. However, I would like to raise some issues that should be dealt with. 

We thank the reviewer for their generally positive response to this work. Below, we address the issues raised in detail.

As for the GHF data, I am a bit confused about which corrections have actually been applied, which ones are just mentioned as caveats, and what the rationale is behind including or excluding the corrections. The ones discussed in Sect. 2.2 are evidently applied. However, things become a bit obscure in Sects. 4.1 and 4.2, where some corrections are explicitly said to be used, whereas others are merely mentioned. This should be separated more clearly, perhaps by moving the explanation of all actually applied corrections to Sect. 2.2, while discussing in Sects. 4.1 and 4.2 only the conceivable corrections that are not applied in this study.

It is correct that topographic correction is the only systematic correction available in the database. We have now included a new Section 2.3 ("Topographic Section") that introduces this in the methods. We also clarify in the broader discussion of "Other Corrections" (Section 4.2), that topographic correction is the only systematic correction available in the database. We have also clarified in the "Summary Remarks" (Section 5), that topographic is the only presently available systemic correction.

Related to this topic, I am somewhat surprised that the authors do not attempt to impose a paleoclimatic correction for the effect of glacial-interglacial cycles on the basal
temperature gradient of the ice at subglacial sites. This is explicitly said in lines 468/469: "Indeed, the 61 +/- 2 mW/m2 present-day heat flow that we estimate at GRIP is ~20% greater than the 51 mW/m2 estimated for that site with paleoclimatic correction by Dahl-Jensen et al. (1998)." [BTW, the study by Greve (2019), which also accounts for the glacial-interglacial correction, gets exactly the same value as Dahl-Jensen et al. (1998).]

In a paper by Calov and Hutter (1997, 49(5), 919-962, https://am.ippt.pan.pl/index.php/am/article/view/v49p919), the authors demonstrate for Dye 3, Summit (GRIP) and Camp Century that the imbalance due to the time-dependent surface climate can be more than 10 mW/m2.

With Section 4.1, we clearly highlight the importance of paleoclimatic corrections when interpreting Greenland heat flow measurements. But, at this time, there is no systematic paleoclimatic correction that we can apply to subglacial, subaerial and submarine sites. Therefore, we cannot include any systematic corrections to the current version of the database. We hope to secure funding for dedicated talent to make these corrections on a site-by-site basis for version 2 of the database. We have now included the Greve2019 and Calov1997 references in the relevant discussions.

As for the constructed GHF map, I have some doubts whether the decision to omit the large GHF value at NGRIP for the main product was a good one. I understand the argument that a single, outlying value is spurious and may not be representative for a larger region. However, there is some additional evidence from the glaciological side, namely the existence of the North-East Greenland Ice Stream (NEGIS) that originates east of NGRIP and flows generally northeast towards the coast. The fast flow of this extended ice stream requires a continuously temperate base, which is hard to maintain with the 40ish mW/m2 GHF values in the area that I infer from Fig. 7. The situation is clearly better in Fig. 9, even though the main zone of elevated GHF values lies to the west of the NEGIS area. This issue deserves some further thinking/discussion.

Consistent with this NGRIP feedback, as well as similar NGRIP feedback from R2, we now also make the “with_NGRIP” data product available online (https://doi.org/10.22008/FK2/F9P03L). We also include new figures highlighting the methodological uncertainty with and without NGRIP included in the learning ensemble (Figure 9), as well as presenting the “with NGRIP” simulation in similar detail to the existing “without NGRIP” simulation (Figure 3). This allows the user to make their own decision about NGRIP inclusion. But, to be clear, we simply argue throughout the manuscript that NGRIP does not appear regionally representative, at least at 55 km resolution of our machine learning input data. We now clarify in Section 2.4 (“Greenland Heat Flow Map”) that the machine learning input fields do not support relatively high heat flow at NGRIP, and that the machine learning only simulates high heat flow at NGRIP if forced by inclusion of the NGRIP measurement.

Detailed comments:

Table 1: A bit more information would be helpful. What is "parent" vs. "child"? What is "TC pT"? Others at the authors' discretion. I fully understand that detailed explanations of all these entries are unnecessary, but they should at least be roughly understandable. We now define T and TC in the table caption and we have removed the redundant "parent" and "child" nomenclature unique to Fuchs et al. (2021). We now also direct readers to Fuchs et al. (2021) for a full description of the IHFC naming convention. We have also moved this table to the Appendix.

Line 92: I suggest adding the main information about EPSG:3413 (polar stereographic projection, parameters). We have now inserted additional details about the main parameters of EPSG:3413 projection.
Lines 130/131, "Heat flow uncertainties are also estimated for all 290 sites, based on the approach described in Section 2.1": I am not sure to what part of Section 2.1 this statement refers. This should be clarified.

We now describe that where site-specific measurements of both temperature gradient and thermal conductivity are available, we assume an uncertainty of ±5%. Where only site-specific temperature gradient is measured and thermal conductivity is assumed, we assume an uncertainty of ±10%. Where only heat flow is reported, without a specific temperature gradient or thermal conductivity, we assume an uncertainty of ±15%. We also clarify that this uncertainty treatment is applied to both Type 0 and Type 1 sites.

Table 3: "This study"
We have made this correction.

Line 168: Reference for the 86°C?
We now clarify that "86°C" is the average top-to-bottom temperature difference for the 14 deep exploration wells presented here. As described in the data availability, the primary temperature data remain proprietary for these exploration wells. We have only negotiated permission to publish the secondary heat flow data.

Line 169: Reference for the 2.00 W m-1 K-1?
We now clarify that this bulk thermal conductivity applied to all 14 deep wells approximates the bulk thermal conductivity applied to 5 of these wells by Rolle (1985). Here, we note that Rolle (1985) examined stratigraphy of each well, but it is unclear to us whether this determined their choice of thermal conductivity. Ultimately, we are clear that this is an assumption, upon which future researchers may improve.

Table 4: "Previous study", "This study"
We have made this correction.

Equation (1): Units are missing.
We have now added units.

Lines 231-237: Basal melting is not the only problem at subglacial sites with a temperate base (T = T_pmp). Another one is frictional heating due to basal sliding, which works in the opposite direction: Basal melting consumes sensible heat, while frictional heating produces it. This makes it very difficult to estimate the geothermal heat flux in the underlying rock from the heat flux into the ice sheet ("basal temperature gradient approach").

We have now noted this point -- that both frictional heating and basal melting can influence basal ice temperatures in temperate settings -- in Section 2.2.4 ("Type 3 - No Heat Flow").

Lines 296-298: I find it a bit inconsistent to keep Table 5 in the main text, while outsourcing Figure A1 to the Appendix. Since it is not explained in much detail, what about moving everything to the Appendix, and perhaps giving a bit more detail there to make the paper more self-contained?
We have now removed previous Table 5 to the appendix and renamed it Table A2. We have also inserted more description to transfer the key points of this table into the main text.

Lines 390-392: Why does the inclusion of NGRIP also produce an island of large GHF values (~70 mW/m2) around ~69°N, 43°W (Fig. 9 vs. Fig. 7)?
We now clarify that this “island” of elevated heat flow is caused by the machine
learning algorithm classifying the “island” as geologically similar to NGRIP, and thereby assigning similar heat flow as NGRIP. Fundamentally, the inclusion/exclusion of NGRIP influences the decision trees of the algorithm, which affects a broader region with similar features as NGRIP.

Line 459, "snowfall rates are comparatively high": This statement is quite vague. Compared to what?
We now rephrase this to state that vertical velocity is effectively equivalent to snowfall rate at the ice-sheet surface. This contrasts with deeper in the ice column, where vertical advection rates become small.

Lines 461/462, "convolution of complementary advection and conduction": In general (exception: ice domes/ridges), strain heating (viscous dissipation) also plays a significant role in the deeper parts of an ice sheet.
We have now added a caveat to this sentence to clarify that this mismatch occurs even in the absence of heat source/sink terms, such as the strain heating noted here.

Line 473: "This _study_ provides"
We have now removed this sentence.

Line 495: "heat production from radioactive sources can also influence the apparent geothermal gradient": My understanding is that a significant part of the GHF we see at the Earth’s surface is due to radiogenic heat production in the crust. So, it’s a normal process rather than merely a perturbation. It should be clarified how to differentiate this normal background from a correction-worthy anomaly.
We now clarify that approximately half of Earth’s contemporary heat flow is ultimately derived from radioactive decay, primarily within the mantle, but near-surface radioactive sources can influence the apparent magnitude and distribution of this background geothermal heat flow.

Lines 525/526: I would find it more logical to swap the order of Tables 6 and 7 (methods first, results later).
We have now made this change in table order.

Table 6: "This _s_tudy" (2 x)
We have made this correction.

Line 566: "This _s_tudy" (also in the first column of the table body)
We have made this correction.

Line 979: "of _t_his _s_tudy"
We have made this correction.

Figure 5: Are the two different panels really needed (subaerial + subglacial = on-shore, submarine = off-shore)? They could be combined into a single one by either adding the green curve (on-shore) to the left panel, or dropping the green curve altogether.
These two subplots are admittedly very similar. But there seems to be a strong tradition of thinking in terms of onshore/offshore. We cannot simply put the “on-shore” curve into the left subplot, as the component distributions sum to the total. We therefore defer to an editorial decision on whether to include or exclude the onshore/offshore right subplot.

Figure 8a: A nonlinear scale would be better, such that O(10 mW/m2) differences can also be discerned (especially for the subglacial sites).
We note that the marker sizes already scale nonlinearly in this figure, which makes small
differences appear disproportionately larger. Perhaps the fundamental issue is that residuals span an order of magnitude. This makes it difficult for small residuals to visually compete with large residuals, but we feel this is appropriate.