FONT STYLES: *Reviewer* | Response

Reviewer #1 (https://doi.org/10.5194/essd-2021-290-RC1)

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 _s_tudy"

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 _s_tudy", "This _s_tudy"

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.

Reviewer #2 (https://doi.org/10.5194/essd-2021-290-RC2)

This paper presents a database for heat flow measurements in Greenland. The database adds some more points and additional context to previous databases, so I think that the presentation of the database is a very useful contribution.

The paper then presents an analysis and discussion of the data, and presents a map of heatflow (Fig. 3) that is based on the data. Here, the authors use quite a bit of discretion and judgment in order to perform the interpolation, and this is necessary because the Greenland heat flow data are so unevenly distributed – almost no data in the interior of Greenland, and then sparse and variable data around the periphery. The authors do an excellent job of comparing their map to the heat flow maps from previous studies (e.g., Fig. 13), and they *provide an excellent and informative discussion (Section 4) of the issues surrounding the collection, interpretation, and interpolation of the Greenland heat flow data.*

Overall, I think this is an interesting and useful contribution that could be published after some revision. The paper is very well written and the figures are well-crafted and convincing. My one major concern is the authors' treatment of the NGRIP station, which I discuss below and I think it should be addressed upon revision. I also have several other specific points that I discuss below, and I think that addressing these will improve the impact of the work.

We thank the reviewer for their generally positive response to this work. We address the NGRIP anomaly in detail below.

Regarding NGRIP:

This station provides one of the only datapoints in central Greenland. It is thus quite valuable, but it is also problematic because it shows a large heat flow value (~130 mW/m^2) that is somewhat larger than other values. The authors here chose to disregard this point from their machine learning analysis (Fig. 3) although they did include this point in some tests with their jackknifing approach (Fig. 9). It seems to me that including or not including NGRIP station has dramatically large impacts on the resulting heat flow map (as can be seen by comparing the different maps in Fig. 13). This seems to indicate a larger uncertainty about the heat flow in central Greenland than the authors have expressed in their analysis – I think that they are underestimating the uncertainty about heat flow in central Greenland. For example, if they had included the possibility of high heat flow at NGRIP station (maybe NGRIP values with large uncertainty associated with them), then they would have a significantly higher max GHF values in Fig. 3c.

The reviewer is correct that the inclusion of NGRIP has a disproportionately large impact on the machine learning algorithm. Clearly, the algorithm identifies a large area of Central North Greenland that is geologically similar to NGRIP, based on the twelve input geophysical datasets. When NGRIP is excluded, this area is assigned lower heat flow characteristic of other subglacial sites. When NGRIP is included, however, this area is assigned the higher heat flow of NGRIP. This leads to our assumption that the measured heat flow at NGRIP likely results from a local phenomenon that is not captured in our twelve input geophysical datasets. We now describe this more explicitly in Section 3.2. We now also include both the "with NGRIP" and "without NGRIP" data products available online (https://doi.org/10.22008/FK2/F9P03L), and we also present the "with NGRIP" simulation in similar detail to the existing "without NGRIP" simulation (Figure 3). This allows the user to decide whether to include or exclude the NGRIP anomaly.

I suggest that the authors find a way to incorporate the additional uncertainty about heat flow in central Greenland that is expressed by their own uncertainty about what the NGRIP value actually represents. There might be several ways to do this. In principle, the machine learning algorithm could be trained to be smart enough to recognize an out-of-range station and disregard it to some extent if necessary. Presumably there are stations elsewhere in the world *that are similarly spurious but included in the analysis, and the machine learning algorithm can learn how to deal with them. But this might be too much for a revision of the current study. Instead, the authors could run the machine learning algorithm again but including NGRIP and take some sort of average of the estimates with and without NGRIP. Alternatively, they might present their jackknifing analysis with NGRIP in a bit more detail, to get an estimate of how large the uncertainty associated with central Greenland really is. (see my comment for Figure 9 below)*

We now highlight the influence of the NGRIP anomaly on heat flow uncertainty by comparing the jackknifing ensemble spreads of the “with NGRIP” and “without NGRIP” in a new Figure 9. This clearly shows that the inclusion of NGRIP introduces a region of elevated heat flow uncertainty around NGRIP. The manuscript also notes that in a pre-processing step we exclude all global heat flow measurements > 200 mW/m^2, as these are likely caused by local phenomenon. We cannot resolve the precise nature of the NGRIP anomaly, we can simply run simulations with and without the NGRIP data point and speculate that it is likely associated with local processes.

Overall, we really do not know much about the heat flow in central Greenland, and so it seems strange to throw away the one data point that we have from this region. Instead, it would be better to incorporate this uncertainty over the NGRIP point into larger uncertainty estimate for central Greenland.

We now provide the “with NGRIP” data product and associated uncertainty estimate.

Specific points related to NGRIP:

*Line 323 – Here the authors present an argument for excluding the NGRIP station (with its very high heat flow) from the machine learning training data. It is true that the heat flow estimates in central Greenland are very sensitive to the observation at NGRIP – this is because this point is much isolated from the others. I do not think that this makes a good argument for excluding the point – instead data points from sparsely-covered areas would be *more* valuable and important to include. I think the authors should develop some sort of general rule for excluding or including points in their analysis, for example based on proximity to other points that could indicate if a given observation is representative of its region. Otherwise, it seems like they are picking and choosing which points to include (and see my point about the next line).*

We now clarify that the spatial location of a measurement is not as important in machine learning as in a conventional spatial interpolation. From a machine learning perspective, it is the relations between data points and input fields that is of most importance. This means that a data point does not become valuable simply due to location. In this sense, NGRIP is not a spatial outlier, but rather a geophysical outlier. Simply put, heat flow at NGRIP is not consistent with other heat flow observations in similar geological settings. In terms of a general rule for exclusion, we only adopt the exclusion of heat flow measurements > 200 mW/m^2 from Lösing and Ebbing (2021). While we recommend the exclusion of NGRIP, we also provide a data product version that includes NGRIP.

Line 324 – The authors suggest that this point might not be statistically representative of the broader region – the authors have no way of knowing this, because there are no other measurements from this region. By this principle, other points that stand by themselves should be similarly dismissed from the database. This would remove pretty much all the points under the main ice sheet (Fig. 1), and there would be very little data left from interior Greenland. Instead, the main reason the authors are disregarding NGRIP is because of its high value – if it had been more "normal" then they would have included it. This is a bit dangerous territory, since excluding data points because they seem spurious can get subjective – and indeed this decision has a huge influence over the resulting heat flow map.

We now show that the influence of the NGRIP anomaly on heat flow uncertainty is disproportionately large by comparing the jackknifing ensemble spreads of the "with NGRIP" and "without NGRIP" in a new Figure 9. This clearly highlights that of all the on-shore measurements, the machine learning algorithm is most sensitive to NGRIP. While we continue to argue that these ensemble spreads justify regarding NGRIP as an outlier, we also now include full analysis and data product for the "with" NGRIP simulation.

Figure 4 – here the relative importance of the different input variables used for the machine learning algorithm are presented. But the most important one for continental Greenland is omitted – the decision to omit the NGRIP station. It seems like this decision should somehow be expressed in a figure like this.

We note that this figure reflects the internal structure of the trained machine learning and does not include pre-processing, such as decisions of which geophysical inputs or training data to include. We now note that the inclusion or exclusion of NGRIP does not fundamentally shift this importance ranking of input geophysical datasets.

Line 397 – Here the authors suggest that some of the extra heat flow at NGRIP may be due to hydrological processes. But wouldn't the associated hydrological processes likely indicate high overall heat flow from this region? That seems to be the case with the other hot springs discussed in the paper (Fig. 6a), and my understanding of hot springs in general (they usually are located in high heat flow areas). (see also my comment below for line 493)

We now clarify that local hydrological processes such as subglacial water flow or hot springs are sub-grid cell processes relative to the global scale of our machine learning algorithm. But here, we also caution that the available heat flow measurements generally do not support high regional heat flow where hot springs are found. For example, the hot springs along 70N in both East and West Greenland are not associated with high local heat flow measurements.

Other specific points within the paper:

Line 40 – The authors present a list of reasons that good heat flow information is necessary. I agree with this list, but I would also add that heat flow data provides useful constraints on the thermal structure of the lithosphere: its elastic thickness, density of heat producing elements in *the crust, etc. I think it would be useful to add this to the list, so as to also make the paper relevant for tectonophysicists.*

We have now included that improved geothermal knowledge help constrain the thermal structure and properties of the lithosphere.

Table 2 – I think that all of the fields specified in the database are useful. There are uncertainties specified for all the components, except for the parameters that go into computing heat flow – namely the temperature gradient and the conductivity. I would think this information could be useful to those using the dataset. For example, if a user feels that the uncertainty in conductivity should be higher (e.g., if they have measurements that suggest this) then they could develop their own uncertainty measure.

Unfortunately, uncertainties in gradient and conductivity are very seldom reported. Indeed, there are many sites for which gradient and conductivity themselves are not reported, it is simply heat flow that is reported. We do, however, report when conductivity is assumed, rather than measured, in the comment section of a site, which provides a qualitative flag. It does not presently appear possible to systematically assess empirical uncertainties in conductivity and gradient at site level.

Table 2 - I also do not fully understand what is meant by "where only gradient or conductivity is reported" (in the statement about heat flow uncertainty). How would heat flow be computed if only one of these is reported? Do the authors mean "if only uncertainty in the temperature gradient or the conductivity is reported"? But in that case, why assign the uncertainty to a set value (e.g., 10%) and not simply use the reported values?

We have now clarified that this ±10% assumed uncertainty relates to sites where temperature gradient is reported, but thermal conductivity is assumed. This contrasts with ±5% assumed uncertainty at sites where both the gradient and conductivity are reported (i.e. no assumptions). While we use reported uncertainties where available, the vast majority of sites have no reported uncertainty, which compels our fractional uncertainty system. This uncertainty system is clearly imperfect, but it is sufficiently transparent and systematic to allow subsequent users to modify uncertainty assumptions.

*Line 127 – I do not understand what is meant by "diminishing extreme values from surveys conducted in the late 1970s and early 1980s". Which are the extreme values – heat flow, conductivity, or temperature gradient? It seems that the authors are identifying these problematic points as having abs((dT/dz)*k – q) > 2, so why do they need to provide another explanation as "diminishing extreme values" – which doesn't seem to have a real meaning (in my understanding). Incidentally, in these instances it seems that dT/dz, k, or q could be reported wrongly, and the authors here are assuming that it is q that is the bad value. Isn't it equally likely that it is k or dT/dz? It seems that the authors should at least consider this possibility if they are replacing a value of q that was reported from a previous study. In any case, I think a bit more explanation here would help.*

We now clarify that the majority of these reassessments are “down revising” extreme values from the 1970s and 1980s. This sentence is not meant to explain a selection criteria, but rather summarize the cohort of reassessed values. We feel it is important to explain that these reassessments generally pertain to older measurements that yielded high heat flow values along the mid-Atlantic ridge. Simply put, there is a clear spatial and temporal coherence to the lower quality IHFC data that we reassess. We also clarify that k and dT/dz are the primary measurements and that q is a secondary derived product.

Line 133 – I would say “lower resolution” instead of “relatively low resolution” since resolution is a relative quantity.

We have made this correction and now provide the respective horizontal resolutions of the BedMachine and ETOPO1 DEMs for quantitative comparison.