Comment on essd-2021-290

Anonymous Referee #2

Referee comment on "Greenland Geothermal Heat Flow Database and Map (Version 1)" by William Colgan et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-290-RC2, 2021

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 judgement 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.

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.

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)

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.

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).

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 here 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 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.

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.

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 hotsprings in general (they usually are located in high heat flow areas). (see also my comment below for line 493)

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.

Table 2 – I think that all of the fields specified in the database 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.

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?

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.

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