Reply on RC1.

Dear editor and reviewers:

We would like to express our sincere gratitude to you for your thoughtful comments and constructive suggestions of our manuscript, which clearly help us improve the manuscript. Please find our replies below. The reviewer's comments are shown in black, and our responses are in red.

Kind regards,

Authors

Reviewers Comments:
Reviewer: 1(Stål, Tobias)
Comments to the Authors:

The study demonstrates systematic optimization of neural network architecture using PSO for hyperparameter tuning, making it a significant methodological advancement over some previous work. The systematic optimization represents a robust alternative to ad-hoc tuning methods commonly used in Earth sciences applications, particularly for such challenging problems as geothermal heat flow prediction in data-sparse regions. The research validates the established understanding of Antarctic thermal structure by confirming the East-West pattern, with predominantly low heat flow values (30-60 mW m^{-2}) in East Antarctica and higher values (>60 mW m^{-2}) in West Antarctica. This consistency with previous studies strengthens confidence in Antarctic crustal thermal architecture. The combination of automated optimization and independent validation of the first-order approximation of the heat flow distribution makes this work valuable for advancing predictive modelling in polar geophysics.

As an output model of actual geothermal heat to expect and include, e.g., interdisciplinary models, I am more skeptical. I have several questions regarding the observables used (listed below). 1.The authors primarily include legacy data (e.g., as available when Aq1 was generated six years ago) along with a few additional datasets that I believe are not very robust. Some choices are not geologically meaningful, as outlined below, and the lack of qualitative assessment of the observables unfortunately invalidates the otherwise sensitive tests conducted. PSO is a valuable tool for DNN, and transparent enough to generate meaningful uncertainty metrics. However, the robustness that PSO is otherwise known for doesn't really help if the features are not meaningful, and we are treating interpolated grid values with the same weight as high-quality and representative observations (discussed by Al-Aghbary et al, 2025, link below). In general, gradient-based optimizers often outperform PSO in similar setups; however, there is certainly a value in testing and comparing various methods, and I believe there will be more development in this field over the coming years, including hybrid strategies (as introduced here with the Adam optimizer).

The ROI analysis offers a reasonable approach to address variations in in-situ data point density; however, a fundamental problem persists regarding how well a single heat flow measurement can represent an entire grid cell. Studies from West Antarctica demonstrate very large local variations in geothermal heat flow. While averaging measurements within global database cells could theoretically mitigate this issue, many cells contain only a single measurement. In these cases, we lack insight into the local conditions that were actually sampled, whether the measurement represents typical regional conditions or a localized anomaly. This spatial representativeness problem becomes particularly acute in Antarctica, where individual point measurements must characterize grid cells spanning millions of square kilometres, potentially introducing significant bias into the training dataset through disproportionate weighting of these sparse observations. However, those issues are not for this paper to resolve, and the methods and analysis are communicated very transparently and clearly. The paper contains many insightful comments regarding concerns and limitations, which are very welcome and still rare.

Some sections of the introduction are challenging to read and don't really make much sense, as if they were written by a language model rather than a scientist. The figures are very good; however, I suggest that Fig.7 be updated (as below).

I am supportive of the publication. However, I am not sure that this is the optimal journal, as the paper's main quality lies in the development of the DNN methods; however, I leave this for the editor to consider.

Thank you for your detailed comments. We greatly appreciate the recognition of our systematic PSO-based optimization approach and the transparent communication of methods and limitations. Your constructive suggestions regarding observable selection and spatial representativeness has helped us significantly improve the manuscript.

Following your suggestions regarding dataset selection, we have updated our datasets using newly available sources, including the Curie depth data for Antarctica from Martos et al., and removed variables that are not geologically meaningful in the Antarctic context (e.g., sediment thickness, rock type, and distance to hotspots). We have also re-conducted the sensitivity analysis based on the updated dataset. Furthermore, to address the black-box nature of the DNN model, we incorporated Bayesian probabilistic modeling to quantify its internal uncertainty and employed Particle Swarm Optimization (PSO) to optimize the model parameters.

We thank the reviewer for understanding the challenges associated with the inherent spatial representativeness limitations of the Antarctic heat flow dataset. To mitigate this issue, we applied a quality-weighted averaging method to the grid cells, assigning weights according to the measurement quality ratings. This approach effectively handles grid cells containing multiple measurements; however, as the reviewer rightly pointed out, it remains limited for cells with only a single measurement—this is an unavoidable constraint under the current data conditions and methodology.

We apologize for the introduction sections that were challenging to read. We acknowledge that large language models were used for language polishing in the original manuscript, which may have affected the clarity and scientific coherence of certain passages. This is mainly because our initial draft was written in Chinese, and we lacked confidence in expressing ourselves in the Introduction when translating it into English. Therefore, we used a large language model to assist us with polishing the Introduction. We have now rewritten the introduction section and have updated the figures accordingly.

Main Items to Address Before Publication:

1. A review of the Introduction, correcting some references, and especially ensuring that the text has a clear and meaningful narrative and that statements are supported by discussion and/or references to studies.

   Following your suggestion, we have thoroughly revised the Introduction section and corrected the references. The specific revisions are as follows:

   "Geothermal heat flow (GHF) refers to the heat energy transferred from Earth's interior to the surface via conduction or convection (Pollack et al., 1993). As a critical heat source beneath the Antarctic ice sheet, GHF not only directly affects the subglacial hydrological system and promotes basal melting , but also serves as an important boundary condition for numerical models predicting the Antarctic Ice Sheet(AIS) mass balance and global sea-level change (Obase et al., 2023; Pollard et al., 2005; Seroussi et al., 2017; Wearing et al., 2024; Llubes et al., 2006). Furthermore,

characterizing the spatial distribution of GHF over Antarctica is crucial for comprehending the continent's past and present tectonic evolution (Reading et al., 2022).

Unfortunately, severe logistical challenges associated with collecting direct measurements in the Antarctic interior have resulted in a sparse and uneven distribution of in situ borehole GHF data (Fisher et al., 2015). Conventional approaches fall into two categories: on the one hand are those which use a single type of observation to infer GHF, most commonly seismic tomography (Shapiro & Ritzwoller, 2004; An et al., 2015a; Lucazeau, 2019; Shen et al., 2020; Haeger et al., 2022; Hazzard & Richard, 2024) or magnetic anomalies (Fox Maule et al., 2005; Purucker et al., 2012; Martos et al., 2017), although broad tectonic reconstructions have been used as well (Pollard et al., 2005). On the other hand, there are a newer set of statistical methods which integrate multiple types of observational constraints to infer GHF using multivariate similarity analysis (Stål et al., 2021), Bayesian inversion (Lösing et al., 2020) or machine learning (Lösing & Ebbing, 2021). While these approaches exhibit consistency at the continental scale, characterized by higher GHF beneath West Antarctica and lower values in East Antarctica, substantial discrepancies persist at regional scales. Methods relying on single observation types are typically constrained by limited data resolution and spatial coverage, as well as by underlying assumptions that may lack universal validity. For instance, seismic tomography-based approaches provide regional-average GHF estimates derived from data with limited sensitivity to upper crustal composition and a coarse lateral resolution of 600–1000 km across Antarctica (Shapiro & Ritzwoller, 2004). As demonstrated by Goutorbe et al. (2011) and Lucazeau (2019), integrating multiple observables yields more robust results than those derived from any single dataset. Specifically, Stål et al. (2021) showed that using 14–19 sets of observables produces a misfit of less than 10 mW m^{-2}, whereas additional datasets may introduce excessive noise without significantly improving estimates. Consequently, multi-observable approaches necessitate a careful selection of features with adequate Antarctic coverage and strict control over the number of inputs. Uncertainties in the original input data can propagate through the modeling process, and the resulting uncertainties in subglacial GHF estimates can substantially impact ice sheet mass balance simulations. Given that Antarctic ice sheet dynamics remain the largest source of uncertainty in future sea-level rise projections, with estimates for the year 2100 under the RCP8.5 scenario ranging from −7.8 to 30.0 cm in multi-model ensembles (Seroussi et al., 2020) to over 1 m when ice-cliff instability is considered (DeConto and Pollard, 2016), reducing GHF uncertainty is critical for improving the reliability of sea-level change predictions.

Recently, deep neural networks (DNNs) have emerged as powerful tools for synthesizing high-dimensional geoscience data, leveraging their formidable nonlinear mapping capabilities. Their efficacy has been proven in improving estimates of Antarctic ice sheet surface melt (Hu et al., 2021), estimating sea ice thickness from satellite radiometry (Herbert et al., 2021), and emulating basal melt rates beneath ice shelves (Burgard et al., 2022). However, current neural network models encounter two primary challenges. First, the performance of DNNs is highly sensitive to numerous hyperparameters; manual or suboptimal tuning often leads to poor generalization or overfitting. Second, as inherently opaque "black-box" models, DNNs seldom provide reliable probabilistic estimates or confidence *intervals. This lack of quantifiable uncertainty limits their applicability in downstream earth system modeling where error propagation is a concern.*

To address these issues, this study proposes a hybrid framework that couples DNNs with Particle Swarm Optimization (PSO) algorithms to refine parameter selection, underpinned by a Bayesian module for robust uncertainty quantification. This integrated approach introduces two key processes aimed at enhancing model generalization and reliability. First, the global search capability of PSO is leveraged to optimize DNN hyperparameters, thereby minimizing the objective function and improving predictive accuracy in data-sparse regions. Second, the integration of a Bayesian module facilitates the decomposition of uncertainty into aleatoric components (stemming from input data noise) and epistemic components (inherent in the model architecture and parameters). In the following sections, we detail the dataset construction and methodology, provide an analysis of discrepancies between the new GHF estimates and prior predictions, and discuss potential uncertainties along with their implications for future investigations."

2. A discussion on why the datasets are chosen (e.g., "for backward compatibility to be able to compare the results with studies using simplistic statistical approaches", or what you prefer).

   Thanks, this suggestion is very helpful. We have substantially revised Section 2.2 to provide a comprehensive rationale for dataset selection. We have also clarified that legacy datasets were included to enable direct comparison with previous studies. Please refer to Section 2.2 of the revised manuscript for details.

3. Reduced dependency on Antarctic in-situ measurements for validation.

   We agree. It is important to treat in-situ inferences carefully, since they are representative of localized temperature structure and are potentially susceptible to contamination by thermal signals caused by frictional heating at the base of the ice sheet, hydrological circulation, and local topography. Given the sparsity of Antarctic GHF estimates derived from in-situ temperature probe observations in boreholes and unconsolidated sediment, we have now treated these data as an independent reference rather than as a validation dataset.
4. As Fig. 7, and in text, confirm what is the most similar model and maybe provide some rudimentary numerical test, e.g., average difference.

   We thank the reviewer for this suggestion. We have revised Figure 7 to identify the most similar model and have included numerical comparisons to quantify the model similarities.

   Quantitative comparison of our model against existing Antarctic GHF estimates reveals that the model of Martos et al. (2017) shows the highest overall agreement with our predictions, with a mean difference of −2.5 mW/m^2.
5. Check and confirm the correct references, and organize the bibliography.

   We have rechecked all references and reorganized the bibliography accordingly. The reference list is now organized alphabetically by first author surname.

Detailed Comments:

L37: Mareschal and Jaupart (2013) is a good overview; however, it's not relevant to reconstruct Antarctic tectonic history. Reading et al (2022, NREE) is probably the most suitable example here.

L43: The text here is not clear; I understand, but it needs some editing.

L47: Citing Lösing, Ebbing et al (2020[should be 2021?]) here also appears a bit out of place. Rather, acknowledge how this study helped us contextualise previous temperature-gradient-based studies. Or is it Lösing and Ebbing (2021)?

L50: The statement that "process-based modelling [depends on] complex mathematical formulation" requires some explanation of what this means, how this is a problem, and why the study at hand addresses this.

L50-55: This section is very hard to follow and doesn't really make any sense. It appears to contradict the previous sentence somewhat.

We thank the reviewer for the detailed comments on the Introduction section. We have carefully considered all suggestions and have substantially rewritten this section to improve clarity.

L37: We have revised the citation and replaced Mareschal and Jaupart (2013) with Reading et al. (2022, NREE).

L47: You are correct. The correct citation is Lösing and Ebbing (2021), which we have now corrected.

L43、L50、L50-55: We have removed this confusing section in the revised manuscript and rewritten this section to clarify the meaning. Please refer to the revised manuscript for details.

L61: I respectfully disagree that deep learning has been particularly successful in polar regions. Whilst there have been a few very useful studies recently (Notably by Prof. Tang), and a lot of method development, the general applications of studies have largely been limited by data availability and lack of consistency and structure. Compared to other regions of the world, DL/ML methods in polar regions have often failed to generate outputs that have been widely accepted to advance our understanding. The statement requires some analysis of why Polar regions have been more successful than elsewhere. In general, I believe that extra caution is required, and uncertainty must be communicated well when dealing with the unknown subglacial geology and in interdisciplinary studies. In the past, we've seen many examples of how research outputs have found interdisciplinary applications that they are not suited for. This is due to change, of course.

We agree with this assessment. While deep learning is a powerful tool, its application in polar regions faces certain limitations due to data scarcity and the "black-box" nature of these models. We have revised the text to provide a more balanced view. Improving model interpretability and addressing data constraints remain key directions for future development in this field.

Section 2.1: One major problem for empirical heat flow models, and related models, is that the training set, or reference set, is not an unbiased representation of Earth's surface. Some settings are highly overrepresented (Stål et al, 2022, Frontiers). How does this impact your results?

We acknowledge that this sampling bias represents a limitation that cannot be fully resolved through methodological improvements alone—it requires fundamental expansion of the observational database in underrepresented regions. Neural networks learn weights from large datasets, and when certain geological settings or regions are overrepresented in the training data, this inevitably influences the model's predictions. This bias is also reflected in our results, where very few predicted values exceed 120 mW/m², likely due to the underrepresentation of high heat flow settings in the global training dataset. Addressing this limitation will require continued efforts to acquire new measurements in undersampled regions and geological settings.

L87: "marine measurements excluded," but Figure 1 map shows many marine measurements in, e.g., the North Atlantic, Mediterranean, and East China Sea. Are those included or not?

We have re-excluded data with the "marine" domain attribute from the IHFC database. For the NGHF database, we only retained data with geography codes A, B, C, D, E, F, G, and H (representing continental regions: Africa, North America, South America, Australia, Europe/Greenland, miscellaneous lands, Antarctica, and Asia/Arabia/India, respectively), excluding all oceanic measurements. Figure 1 has been updated to reflect this correction.

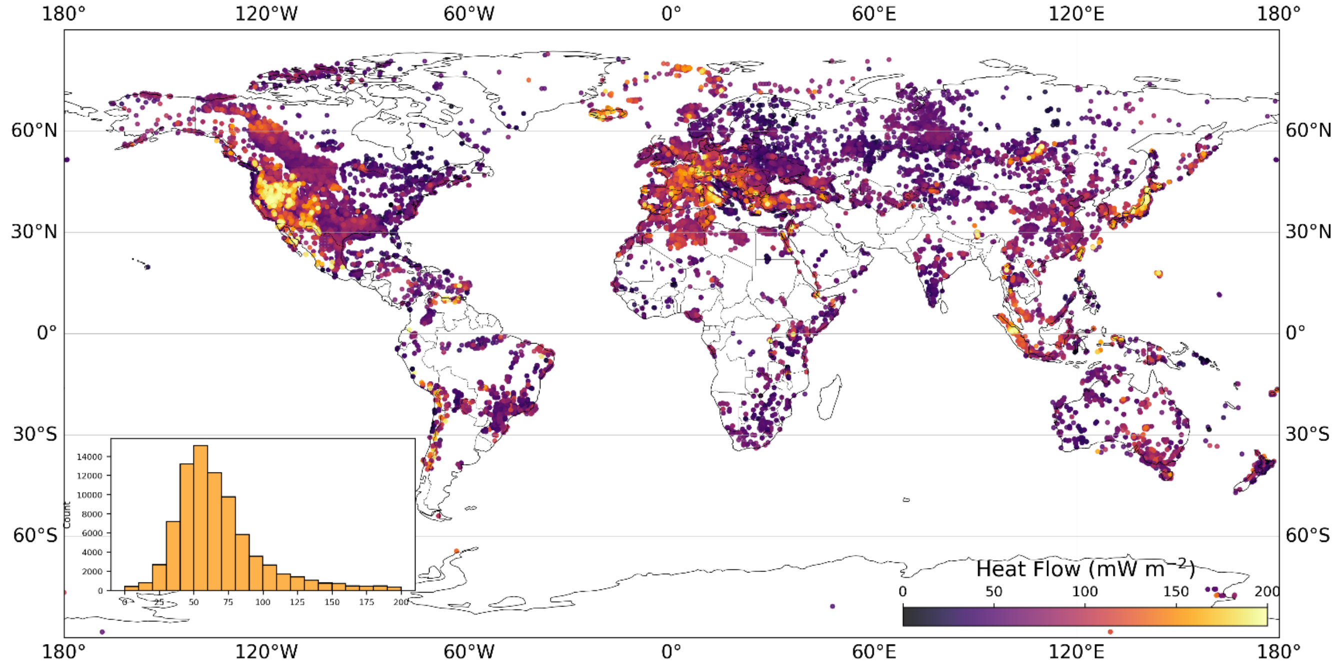

Figure 1. Spatial distribution of global GHF measurements used for model training.

Fig. 1: “Dataset obtained from IHFC and NGHF”; however, the text only mentions IHFC. Were the two databases merged? Wouldn’t that duplicate most records in NGHF?

Yes, the two databases were merged. After excluding marine data, the IHFC database contains 41,511 measurements and the NGHF database contains 44,698 measurements. There is indeed substantial overlap between these two databases. After merging and removing duplicate records, we obtained a final dataset of 57,654 unique heat flow measurements.

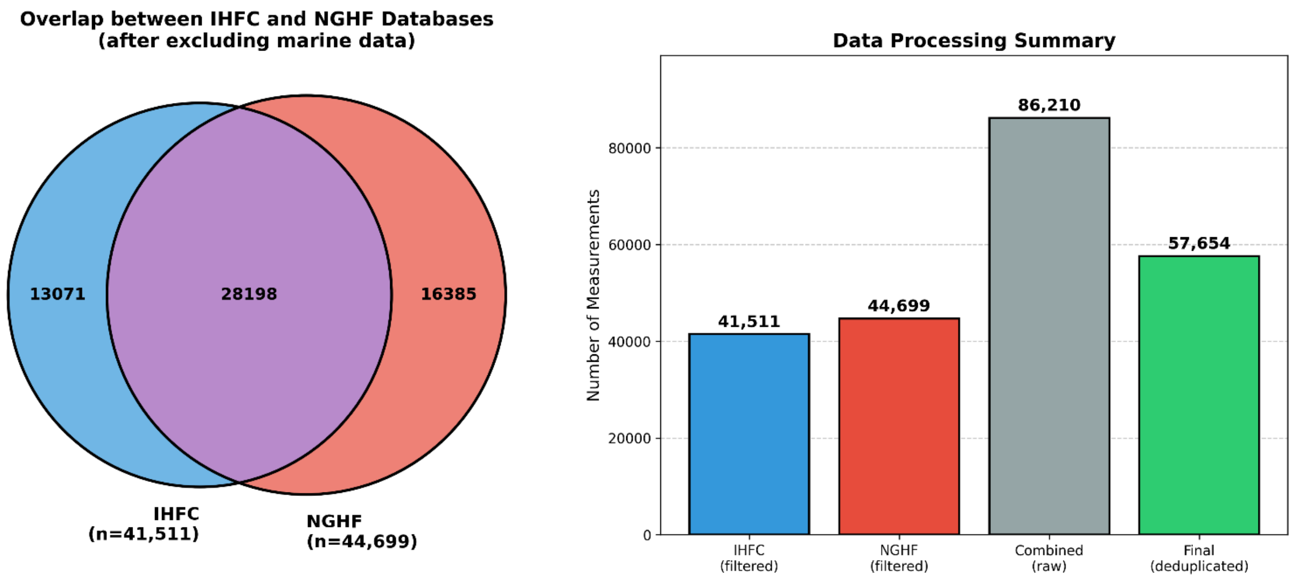

Figure 2. Overlap between IHFC and NGHF.

Table 1 (and general comments on features used):

Why are you using such relatively old datasets? With all respect to the legacy, the results from CRUST1, Shaeffer and Lebedev (2015), and An et al. (2015) are all good studies; however, they are over ten years old, and a lot of data have been collected since then. I notice a significant similarity with the observables used to produce Aq1 back in 2019-20; however, I would have used different datasets today.

We acknowledge that some of the datasets used in this study are relatively dated. As noted earlier, to enable direct comparison with previous studies using statistical and machine learning approaches (e.g., An et al., 2015; Martos et al., 2017), we included several legacy datasets that have been standard inputs in Antarctic GHF modeling. However, following the reviewers' suggestion, we have updated several datasets in the revised manuscript. Specifically, we have incorporated the Curie depth data for Antarctica from Martos et al. (2017) to supplement the GCDM model, and removed features that lack physical significance in the Antarctic context (e.g., sediment thickness, rock type, and distance to hotspots). We acknowledge that further improvements could be achieved by incorporating more recent datasets as they become available, and we consider this an important direction for future work.

Rock type is likely not a very useful observable, as most of Antarctica is classified as ice, and we know that the crustal geology is important but challenging to model (Stål et al., 2024, GRL).

We agree. Rock type is not a meaningful observable in Antarctica since most of the continent is classified as ice (see figure below). We have removed this feature from our model in the revised manuscript.

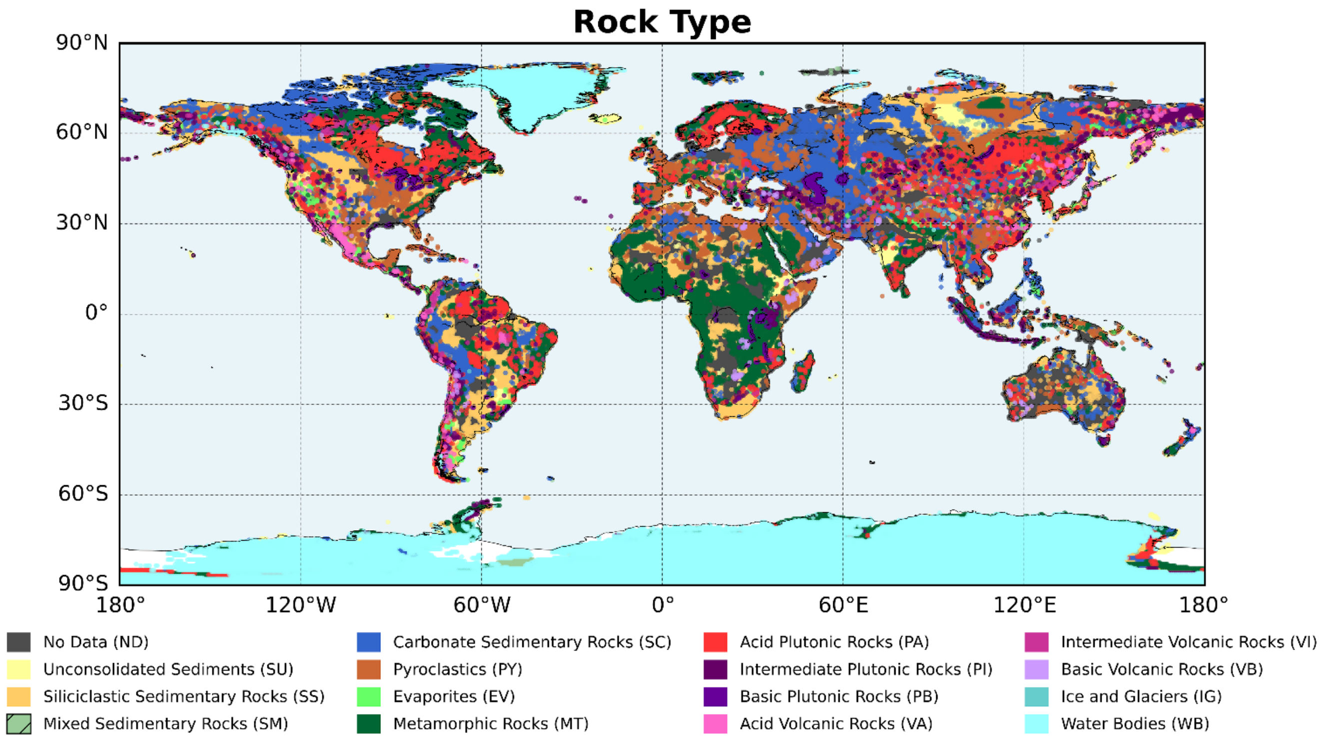

Figure 3. Global Lithological Map

Some observables, e.g., CTD from Li et al (2017), have very little coverage in Antarctica.

Thanks for the comment. We have now augment the dataset with data from Martos et al. (2017).

What depths are the P wave speed and S wave speed taken from? Can the tomographic model suggest values for the crust, as suggested on L117?

The P-wave and S-wave velocities were extracted at 150 km depth from the Schaeffer and Lebedev (2015) model. This depth is within the upper mantle rather than the crust. We have revised the text to clarify that these seismic velocities serve as proxies for the thermal state of the upper mantle, rather than the crust, and reflect the lithospheric thermal structure that influences surface heat flow.

"We selected shear-wave velocity (Vs) and compressional-wave velocity (Vp) data at 150 km depth from the Schaeffer and Lebedev (2015) model, which is constructed from cluster analysis of global surface wave tomography without requiring a priori assumptions about Earth's structure, thus objectively reflecting upper mantle velocity anomalies beneath Antarctica."

Distance to hotspot cites Anderson (2016); however, this study is not in the reference list, only Anderson (1998), which, to my understanding, doesn't provide the spatial data referred to here. Is it the Complete Hot Spot Table? This list, as far as I know, has not been peer-reviewed, and I am rather sceptical of it. As above, it should probably be regarded as legacy work, as there was very little to constrain some of those suggestions 25-30 years ago.

You are correct, the data source was the Complete Hot Spot Table. Given the concerns regarding its reliability and lack of peer review, as well as its limited applicability in the Antarctic context, we have removed this feature from our analysis in the revised manuscript.

Distance to Volcanoes and hotspots is, as I understand, not distance-weighted in any way. Hence, heat flow values and target locations are equally linked, e.g., if the distance to the nearest volcano is 2000 km or 20 km. This is not a useful predictor of geothermal heat. The Adam optimizer compounds this issue by learning to exploit statistical correlations between raw distance and heat flow in the training data, regardless of physical plausibility. Since Adam operates purely on numerical gradients, I suppose it will adjust network weights to minimize prediction error even when the learned relationships violate fundamental geology. This creates a model that may perform well statistically but also might generate physically meaningless patterns.

Thanks for your suggestion. The reviewer is correct that unweighted distance values are not physically meaningful predictors of geothermal heat, as they treat locations 20 km and 2000 km from a volcano equally in terms of their relationship to heat flow.

Following this suggestion, we have implemented distance weighting using an exponential decay function to better reflect the physical relationship between distance and geothermal influence. Specifically, the thermal influence weight is calculated as:

$$w = \exp(-d/\lambda),$$

where d is the distance to the nearest volcano (in km) and λ is the decay parameter. We tested multiple decay parameters (500 km, 1000 km, and 2000 km) to capture different scales of thermal influence. This approach ensures that nearby volcanic sources have a stronger influence on predicted heat flow, which is more physically plausible than using raw distance values.

L207: What procedure to avoid overfitting is applied? Here, I would urge the authors to consider alternative and informative metrics of uncertainty. The recent paper by Al-Aghbary et al (preprint link below) would be a good starting point. What uncertainty and error could/should we actually optimize to reduce?

Thankyou for recommending the work of Al-Aghbary et al. As we stated in the manuscript:

"To control model complexity and mitigate overfitting, we implemented a multi-faceted regularization strategy: (1) L2 regularization (weight decay) was applied to penalize large weight magnitudes and encourage simpler model representations; (2) batch normalization was implemented after each hidden layer to stabilize training dynamics and accelerate convergence; (3) dropout layers were incorporated between hidden layers to randomly deactivate neurons during training, reducing co-adaptation and improving generalization; and (4) an early stopping mechanism was established, terminating training if validation loss failed to decrease for 10 consecutive epochs, with model weights corresponding to the lowest validation loss retained.

Our analysis reveals that aleatoric uncertainty—arising from inherent observational variability and unresolved geological heterogeneity—constitutes the dominant fraction of total uncertainty across Antarctica. This finding suggests that while our model architecture has sufficient capacity to capture underlying patterns, irreducible noise in heat flow observations and small-scale geological complexity impose fundamental limits on prediction accuracy.

We have reviewed the recent work by Al-Aghbary et al. (2025), which demonstrates a promising approach to address this limitation. By applying unsupervised clustering to partition geophysical observables into homogeneous subsets and training dedicated local expert models within a Mixture-of-Experts (MoE) framework, they achieved substantial reductions in aleatoric uncertainty (up to 29% in synthetic datasets and 8% in real-world GHF data) while maintaining stable epistemic uncertainty. This cluster-specific modeling strategy offers a compelling direction for future improvement of Antarctic GHF predictions, particularly given the pronounced geological heterogeneity between tectonically active West Antarctica and stable East Antarctic cratons. We have added discussion of this approach in the revised manuscript as a promising avenue for future work."

L281: Including the few in situ measurements in Antarctica is very problematic. 1. Most of them don't reach the bed and represent the paleoclimate and hydrology rather than geothermal heat. 2. They are very sparse, and there are no measurements to average in each grid. 3. Some measurements should be treated with some particular care, as they are either old or associated with large technical challenges when made. 4, and most importantly, they get a disproportional weight as they will be very similar to the surrounding region.

Correct! The Antarctic in-situ measurements indeed present the issues you outlined. We have therefore excluded Antarctic heat flow data from our dataset and use these measurements only as reference points for evaluating model performance.

Figure 6: There appear to be gridding artifacts from the projection of observables, such as lineations pointing toward the South Pole.

Thanks for your comment. We have revised Figure 6 to address these artifacts.

L340 and L438: This statement does not seem to agree with Fig. 7 (?). Instead, it appears that your distribution resembles Lösing et al. (2020) most, which I believe should be Lösing and Ebbing (2021).

We have retrained the model and re-performed the comparisons. The result shows that the model of Martos et al. (2017) generally best matches our predictions, with an average difference of -2.5 mW/m². We have updated the text at L340 and L438 accordingly to ensure consistency with Figure 7.

L353 Are the values you get at Lake Vostok simply the in-situ measurement extrapolated? This measurement is likely to get a very high weight.

We thank the reviewer for raising this concern. The elevated GHF values predicted around Subglacial Lake Vostok are not extrapolations from the in-situ borehole measurement, as Antarctic heat flow data were entirely excluded from the training dataset. Our model was trained exclusively on global continental heat flow data outside Antarctica, and the Antarctic predictions are therefore purely extrapolated based on the relationships learned from other continents.

Figure 7. Show the difference, with the sign, rather than the absolute difference.

OK. We have revised Figure 7 as requested, now presenting the difference with its respective sign instead of the absolute difference.

L361 As explained above, I agree with Fisher, as you cite, and I don't think this is a valid test. The measurements are too sparse and represent the very local conditions. The measurements in the interior don't reach the bedrock. We need some further discussion and evidence to claim that the measurements "nonetheless" provide good support. Two papers to consider here are Talalay et al. (2020, Cryosphere) and Mony et al. (2020, Glaciology).

We thank the reviewer for this clarification and for recommending the papers by Talalay et al. (2020) and Mony et al. (2020). We agree that the in-situ measurements are too sparse to serve as a robust validation dataset and represent only very localized conditions. As Talalay et al. (2020) demonstrate, the basal temperature at the Antarctic Ice Sheet and the temperature gradient in subglacial rocks have been directly measured only a few times. Furthermore, Mony et al. (2020) emphasize that thermal gradients within the ice cannot be used to estimate the solid Earth contribution with any certainty unless the exact basal conditions are known and the borehole reaches sufficient depth.

Accordingly, we have revised our approach: these in-situ data are no longer used for model training or validation, but are instead provided solely as an independent reference for qualitative comparison.

The Conclusion is too long and mainly repetitive. I suggest shortening it and merging some items with the Discussion if required.

Thanks for your advice. We have now shortened and streamlined it to focus clearly on the principal findings:

"In this study, we present an integrated framework combining PSO-optimized deep neural networks with Bayesian uncertainty quantification for predicting Antarctic GHF. Through regional density experiments, we found that our model is significantly outperforms linear regression in terms of prediction accuracy and nonlinear mapping capacity, particularly in data-constrained environments. The resulting GHF distribution reveals a pronounced East-West dichotomy. Elevated heat flow anomalies are concentrated along the coastal margins of West Antarctica, primarily driven by active lithospheric extension and tectonic activity. Notably, our model predicts elevated GHF values in East Antarctica compared to previous studies, suggesting that the East Antarctic Shield may not be as uniformly cratonic or thermally stable as formerly assumed. Furthermore, uncertainty decomposition reveals that aleatoric components dominate the total predictive variance, highlighting the fundamental limits on predictability imposed by inherent observational noise and unresolved small-scale geological heterogeneity. Future research should prioritize the acquisition of high-quality borehole measurements in data-sparse regions and the integration of physics-informed constraints to enhance model interpretability and geophysical fidelity."

Bibliography:

The list is not organised and rather chaotic. A few key studies appear to be missing, and even some citations in the text are missing. Please check Lösing et al. (2020) and Lösing and Ebbing (2021); I suspect that the papers have been mixed up a few times in the manuscript. Anderson (2016) is also missing or might have been confused by another study.

We apologize for the oversight and the resulting confusion in the reference list. We have reorganized the reference list into alphabetical order by first author's surname, following standard formatting conventions.

I would recommend that the authors have a look at the following suggestions:

Al-Aghbary et al.

(In review, https://www.authorea.com/doi/full/10.22541/au.175373261.14525669)

Mony et al. (2020, Glaciology).

Reading et al (2022, NREE)

Stål et al. (2022, Frontiers).
Stål et al. (2024, GRL).
Talalay et al. (2020, Cryosphere)

We thank the reviewer for these valuable references. We have carefully reviewed all suggested studies and incorporated them into the revised manuscript. These references have significantly improved the quality of our discussion.

Reference:

Reading, A. M., Stål, T., Halpin, J. A., Lösing, M., Ebbing, J., Shen, W., McCormack, F. S., Siddoway, C. S., and Hasterok, D.: Antarctic geothermal heat flow and its implications for tectonics and ice sheets, Nat. Rev. Earth Environ., 3, 814–831, https://doi.org/10.1038/s43017-022-00348-y, 2022.

Stål, T., Reading, A. M., Halpin, J. A., and Whittaker, J. M.: Properties and biases of the global heat flow compilation, Front. Earth Sci., 10, 963525, https://doi.org/10.3389/feart.2022.963525, 2022.

Stål, T., Halpin, J. A., Goodge, J. W., and Reading, A. M.: Geology matters for Antarctic geothermal heat, Geophys. Res. Lett., 51, e2024GL110098, https://doi.org/10.1029/2024GL110098, 2024.

Mony, L., Roberts, J. L., and Halpin, J. A.: Inferring geothermal heat flux from an ice-borehole temperature profile at Law Dome, East Antarctica, J. Glaciol., 66, 509–519, https://doi.org/10.1017/jog.2020.27, 2020.

Talalay, P., Li, Y., Augustin, L., Clow, G. D., Hong, J., Lefebvre, E., Markov, A., Motoyama, H., and Ritz, C.: Geothermal heat flux from measured temperature profiles in deep ice boreholes in Antarctica, The Cryosphere, 14, 4021–4037, https://doi.org/10.5194/tc-14-4021-2020, 2020.

Al-Aghbary, M., Awaleh, M. O., Jalludin, M., et al.: Improving Geothermal Heat Flow Predictions and Uncertainty Quantification using Clustering-based Quantile Regression Forests, Authorea [preprint], https://doi.org/10.22541/au.175373261.14525669/v1, 28 July 2025.