Supplement of The Cryosphere, 20, 1543–1558, 2026
https://doi.org/10.5194/tc-20-1543-2026-supplement

The Cryosphere Open Access

Supplement of

Mapping Antarctic geothermal heat flow with deep neural networks optimized by particle swarm optimization algorithm

Shaoxia Liu et al.

Correspondence to: Xueyuan Tang (tangxueyuan@pric.org.cn)

Contents of this file

S1 Input feature maps

The following figures present the spatial distribution of all input features used in this study, displayed in both global Mollweide projection and Antarctic polar stereographic projection. Left panels show global distributions; right panels show Antarctic polar stereographic projections.

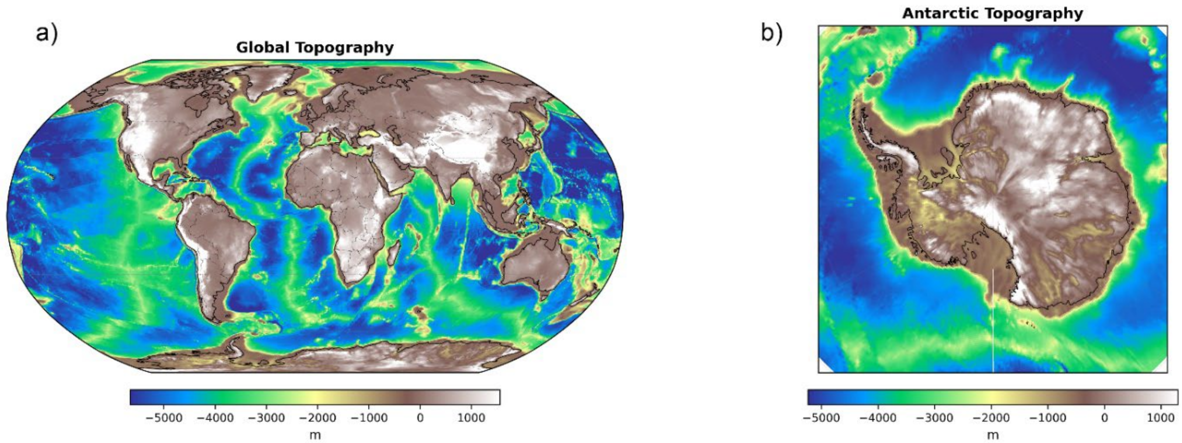

Figure S1. Bedrock topography. (a) Global bedrock topography. (b) Antarctic bedrock elevation. Data for both panels are derived from the ETOPO 2022 Global Relief Model (NOAA, 2022).

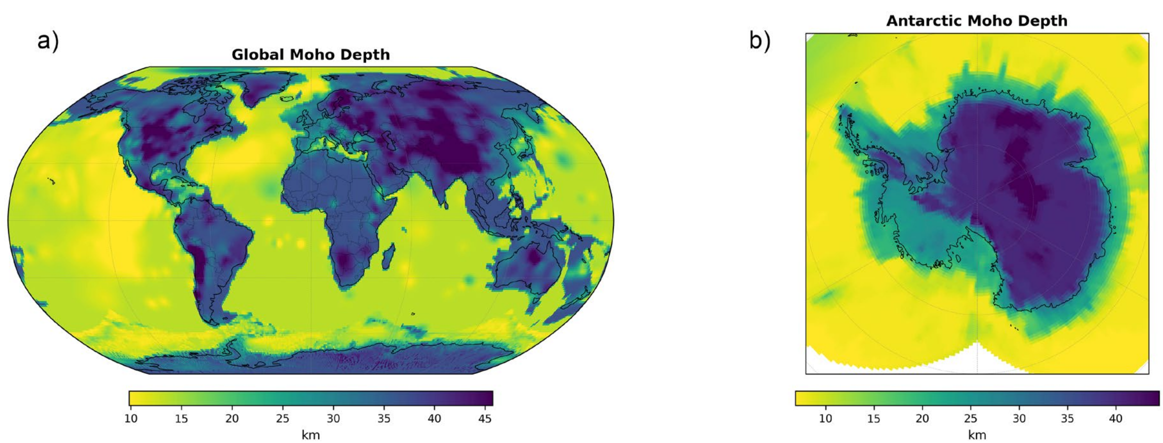

Figure S2. Moho depth. (a) Global Moho depth, sourced from Szwillus et al. (2019). (b) Antarctic Moho depth, derived from An et al. (2015).

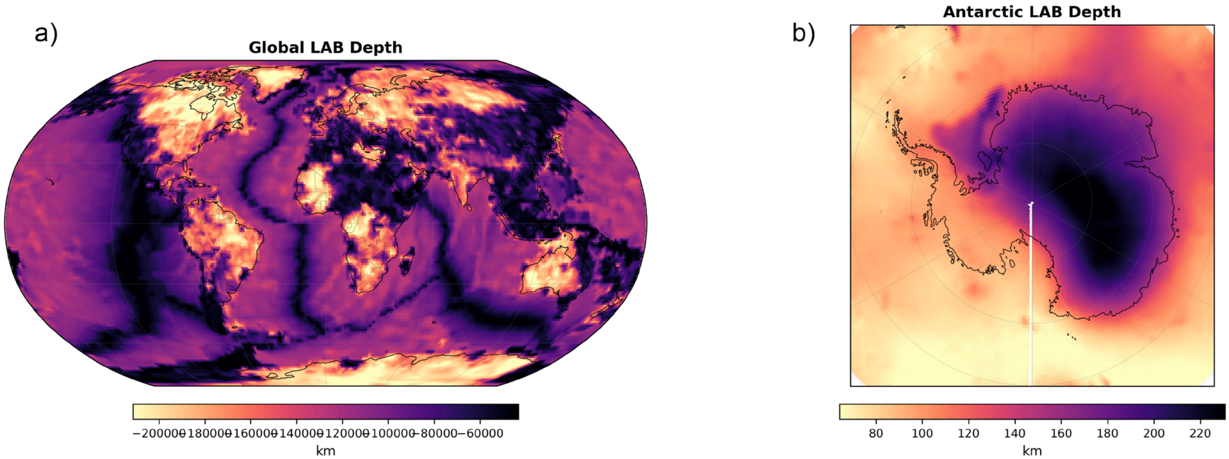

Figure S3. LAB depth. (a) Global Moho depth, sourced from Afonso et al. (2019). (b) Antarctic Moho depth, derived from Pappa et al. (2019).

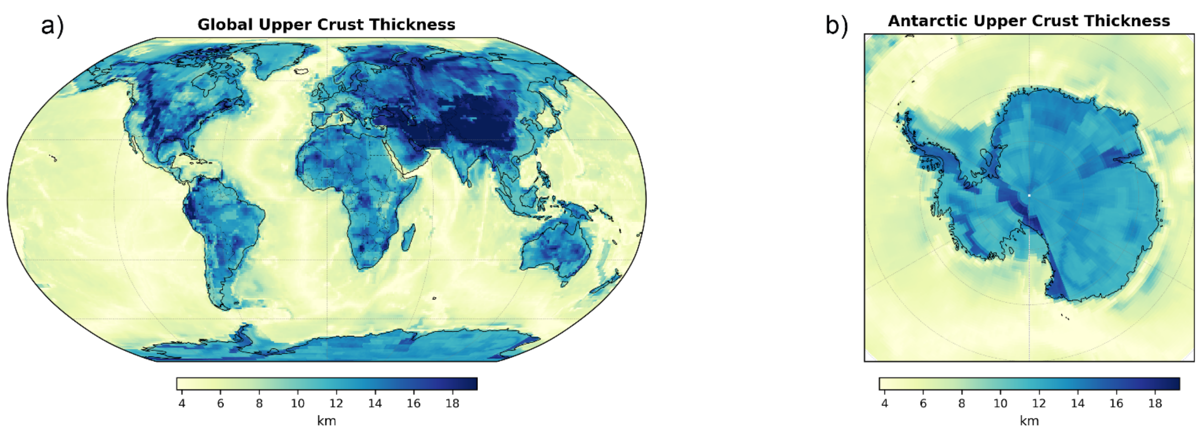

Figure S4. Upper crust thickness. (a) Global upper crust thickness. (b) Antarctic upper crust thickness. Data for both panels are derived from the CRUST1.0 Model (Laske et al. (2012)).

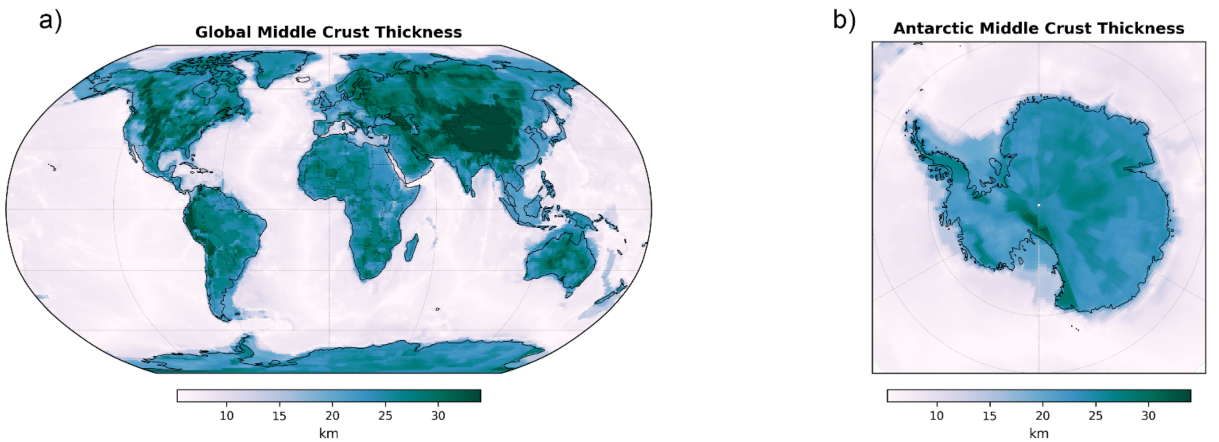

Figure S5. Middle crust thickness. (a) Global middle crust thickness. (b) Antarctic middle crust thickness. Data for both panels are derived from the CRUST1.0 Model (Laske et al. (2012)).

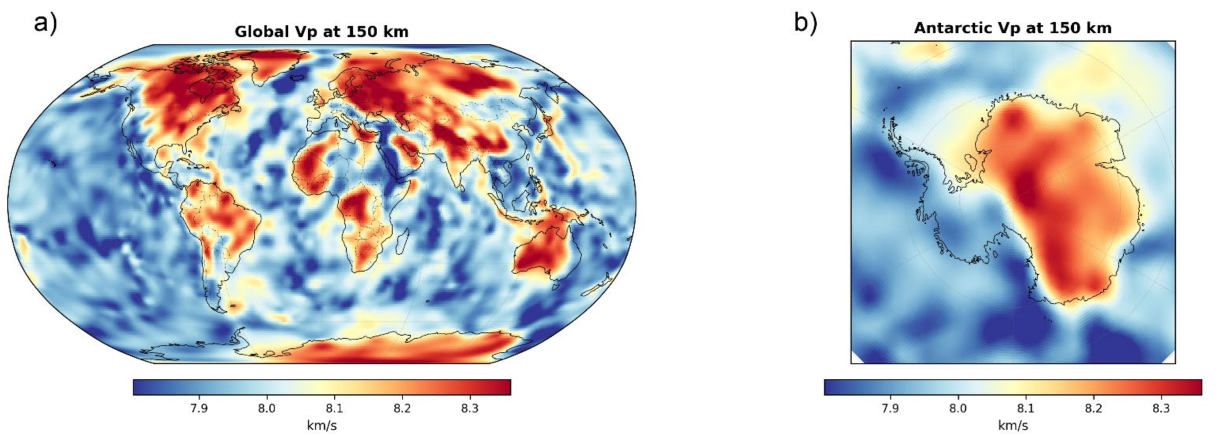

Figure S6. P-wave velocity at 150 km depth. a) Global P-wave velocity. (b) Antarctic P-wave velocity. Data for both panels are sourced from Schaeffer and Lebedev (2015).

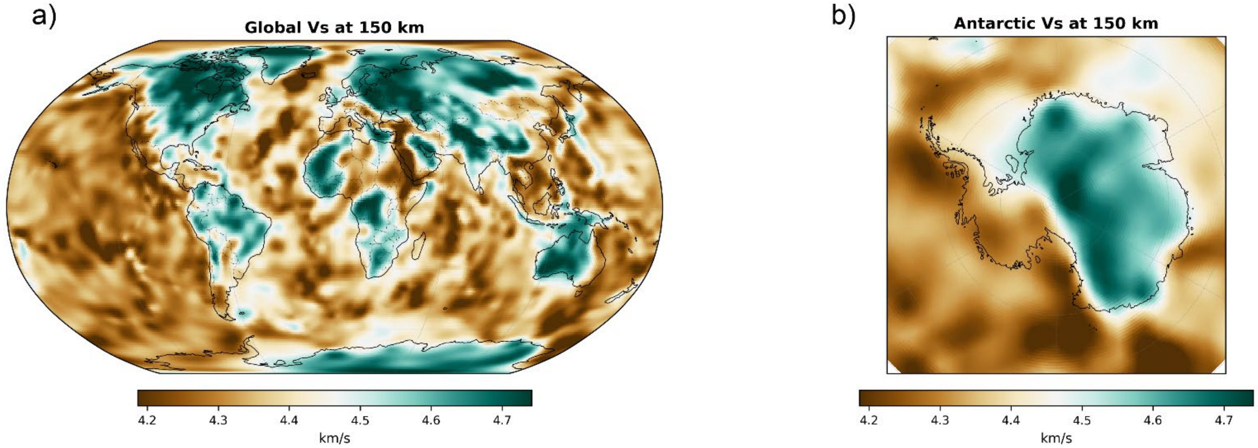

Figure S7. S-wave velocity at 150 km depth. (a) Global S-wave velocity. (b) Antarctic S-wave velocity. Data for both panels are sourced from Schaeffer and Lebedev (2015).

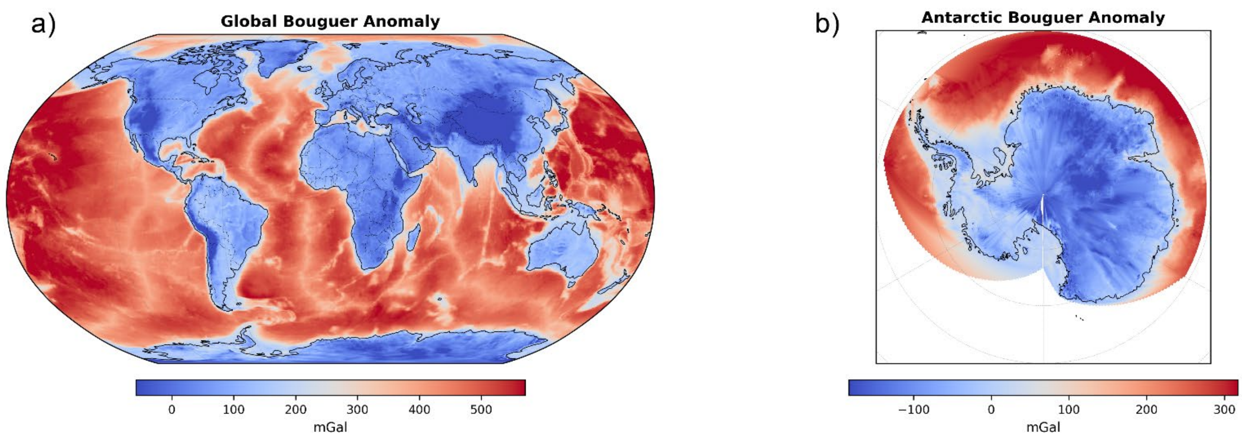

Figure S8. Bouguer anomaly. (a) Global Bouguer anomaly, sourced from Bonvalot et al. (2012). (b) Antarctic Bouguer anomaly, derived from Scheinert et al. (2016).

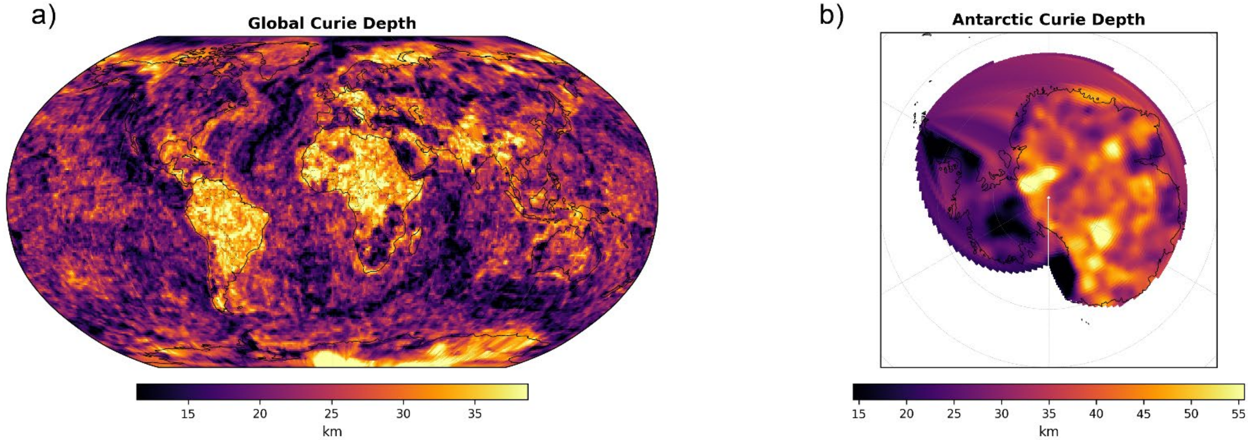

Figure S9. Curie depth. (a) Global Curie depth, sourced from Li et al. (2017). (b) Antarctic Curie depth, derived from Martos et al. (2017).

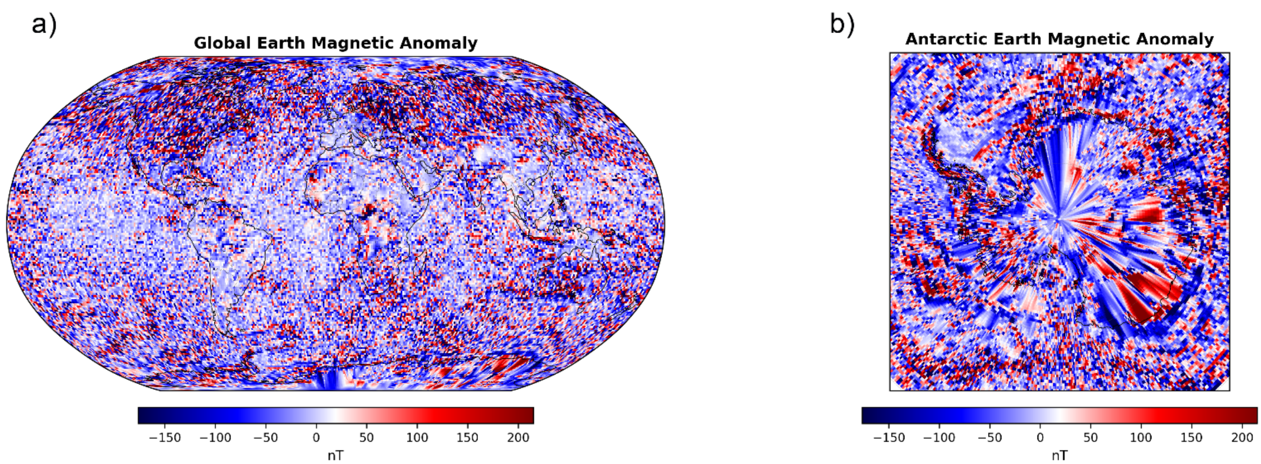

Figure S10. Earth magnetic anomaly. (a) Global Earth magnetic anomaly, sourced from Maus et al. (2009). (b) Antarctic Earth magnetic anomaly, derived from Golynsky et al. (2018).

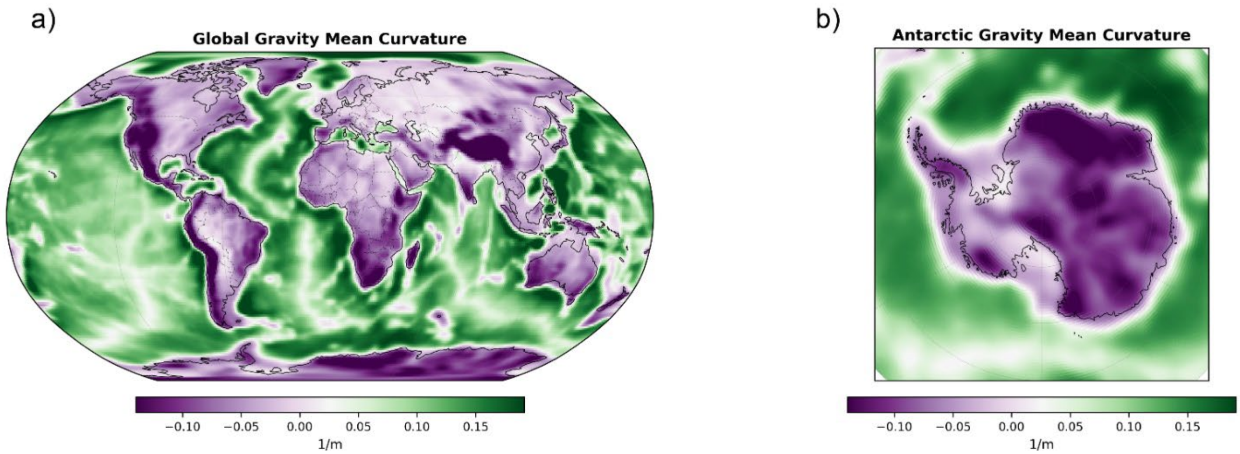

Figure S11. Gravity mean curvature. (a) Global gravity mean curvature. (b) Antarctic target gravity mean curvature. Data for both panels are sourced from Ebbing et al. (2018).

S2 Data Screening and Quality Assessment for GHF Measurements

The target variable for this study, geothermal heat flow(GHF), was compiled by merging two major databases: the International Heat Flow Commission (IHFC) database and the New Global Heat Flow (NGHF) database (Lucazeau, 2019). The combined dataset contains approximately 86,210 in-situ measurements, primarily acquired from bedrock drill holes and thermal probes. To ensure data quality and consistency, we applied the following screening criteria:

1. **Duplicate removal**: Records appearing in both databases were identified based on matching coordinates (within 0.01°) and GHF values, and duplicates were removed.
2. **Missing data exclusion**: Entries lacking geographic coordinates (latitude or longitude) or GHF values were excluded.
3. **Marine data exclusion**: Only continental measurements were retained; marine heat flow data were excluded as this study focuses on continental GHF prediction.
4. **Quality assessment filtering**: Only measurements accompanied by detailed quality grades were retained. The quality assessment in the NGHF database follows a hierarchical grading system based on measurement methodology, borehole depth, thermal conductivity determination, and correction for paleoclimatic or hydrological perturbations (Lucazeau, 2019).

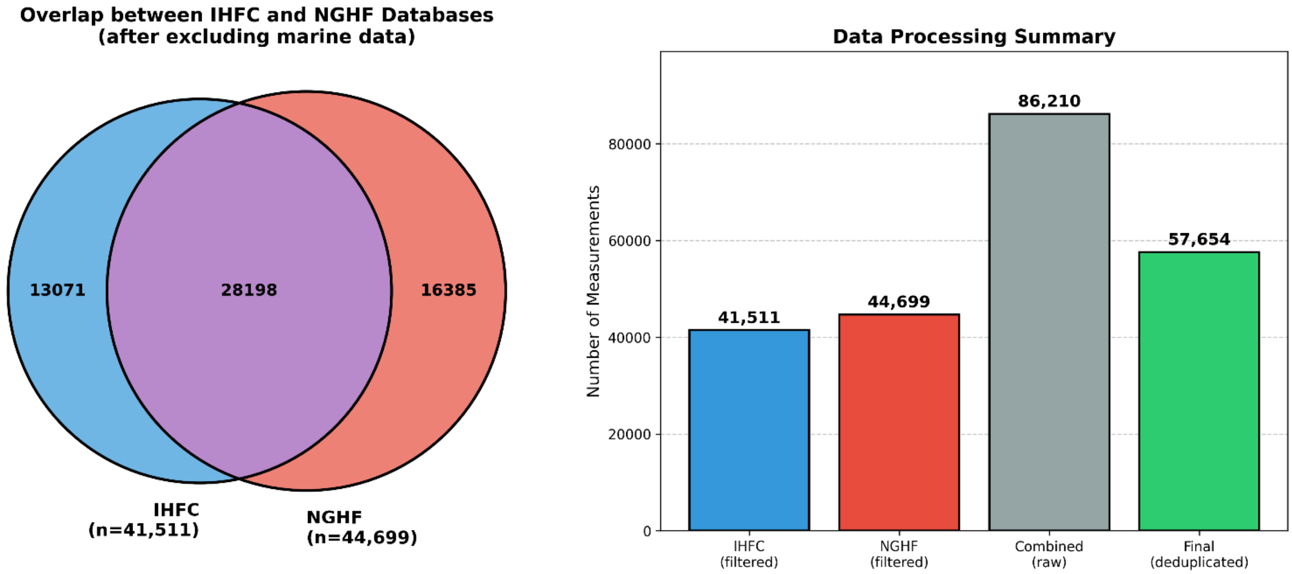

Figure S12. Overlap between IHFC and NGHF.

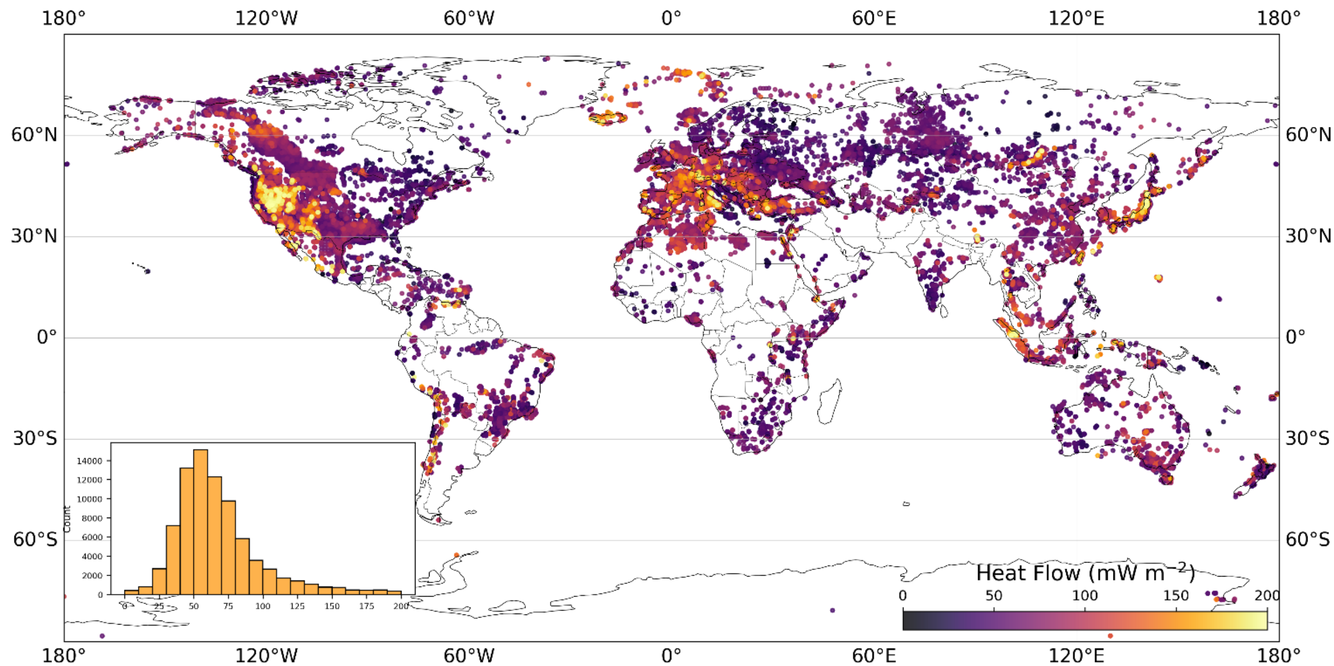

Figure S13. Spatial distribution of global GHF measurements used for model training.

S2.1 Quality Assessment and Classification

To account for varying reliability in the source data, we assigned quality scores to all observed Geothermal Heat Flow (GHF) data points. Following the assessment standards of the International Heat Flow Commission (IHFC), observations were categorized into five quality grades (A, B, C, D, and Z) based on measurement precision and methodological reliability. We assigned a specific Quality Weight (w_k) and an Uncertainty Coefficient (δ_k) to each grade. High-quality measurements (Grades A and B) contribute significantly to the grid estimation with low uncertainty, while low-quality or unknown data (Grade Z) are assigned minimal weights to reduce their impact on the final model. The detailed classification scheme is presented in Table S1.

Table S1. Quality classification, weights, and uncertainty coefficients for GHF data.

Quality Grade	**IHFC / NGHF Code**	**Quality Weight (w_k)**	**Uncertainty Coefficient (δ_k)**	**Description**
A	A / U1 M1	1.0	0.05	High precision; minimal error
B	B / U2 M2	0.8	0.10	Good reliability; standard error
C	C / U3 M3	0.6	0.15	Moderate reliability
D	D / U4 M4	0.4	0.25	Low reliability; high uncertainty
Z	Z / Ux Mx	0.01	1.00	Unknown quality or estimated

S2.2 Grid Design and Coordinate System

To spatially normalize the observed GHF data, this study adopts a standard global grid system with a spatial resolution of $0.5° \times 0.5°$. The coordinates of the grid centers are defined as follows: longitude ranges from $-179.5\ to 179.5$ and latitude ranges from -89.5 *to* 89.5 both with a step size of $0.5°$. This configuration results in a matrix of 720×360 standard grid cells, covering a total of 259,200 global grid points.

S2.3 Grid Assignment Algorithm

We employed a spatial discretization approach to map the irregular observation points onto a regular grid. For each raw observation point with coordinates (ϕ_k, λ_k) (latitude, longitude), we assigned it to the nearest grid center $\boldsymbol{C_{ij}}(\phi_i, \lambda_j)$ using the Nearest Neighbor algorithm based on Euclidean distance:

$$d_{ij} = \sqrt{(\phi_k - \phi_i)^2 + \left(\lambda_k - \lambda_j\right)^2} \tag{S1}$$

The observation point is assigned to the grid cell (i, j) that minimizes d_{ij}.

S2.4 Quality-Weighted Data Aggregation

When multiple observation points (n) are assigned to a single grid cell, we apply a quality-weighted averaging method to determine the representative GHF value. The weighted mean heat flow q_{grid} is calculated as:

$$q_{\text{grid}} = \frac{\sum_{k=1}^{n} w_k \cdot q_k}{\sum_{k=1}^{n} w_k} \tag{S2}$$

where q_k is the observed heat flow value of the k-th point, and w_k is its corresponding quality weight.
The uncertainty of the grid cell is estimated using a weighted standard deviation approach, incorporating the uncertainty coefficients. The grid uncertainty σ_{grid} is defined as:

$$\sigma_{\text{grid}} = \sqrt{\frac{\sum_{k=1}^{n} w_k \cdot (q_k \cdot \delta_k)^2}{\sum_{k=1}^{n} w_k}} \tag{S3}$$

where δ_k is the uncertainty coefficient corresponding to the quality grade of the k-th point. This formulation ensures that grid cells dominated by high-quality data (Grade A/B) exhibit lower calculated uncertainty, while those relying on lower-quality data reflect higher uncertainty.

S3 Bayesian Neural Network and Uncertainty Quantification

S3.1 Bayesian Neural Network Framework

In standard neural networks, weights are treated as deterministic parameters optimized to minimize a loss function. In contrast, Bayesian Neural Networks (BNNs) treat weights **W** as random variables governed by probability distributions. Given a dataset $\mathbf{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, the goal is to infer the posterior distribution over weights:

$$p(\boldsymbol{W}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{W})p(\boldsymbol{W})}{p(\mathcal{D})} \qquad (S4)$$

where p(**W**) is the prior distribution, p(**D**|**W**) is the likelihood, and p(**D**) is the marginal likelihood (evidence). Since the exact posterior is generally intractable, we employ variational inference to approximate it.

S3.2 Variational inference and Evidence Lower Bound (ELBO)

Variational inference approximates the true posterior p(W|D) with a tractable variational distribution $q_\theta(\boldsymbol{W})$, parameterized by θ. The optimal approximation is found by minimizing the Kullback-Leibler (KL) divergence between $q_\theta(W)$and p(W|D):

$$D_{\mathrm{KL}}[q_\theta(\boldsymbol{W})|p(\boldsymbol{W}|\mathcal{D})] = \int q_\theta(\boldsymbol{W}) \log \frac{q_\theta(\boldsymbol{W})}{p(\boldsymbol{W}|\mathcal{D})} d\boldsymbol{W} \qquad (\boldsymbol{S5})$$

Since direct minimization requires knowledge of the intractable posterior, we instead maximize the Evidence Lower Bound (ELBO):

$$\mathcal{L}_{\mathrm{EB}}(\theta) = E_{q_\theta(\boldsymbol{W})}[\log p\,(\mathcal{D}|\boldsymbol{W})] - D_{\mathrm{KL}}[q_\theta(\boldsymbol{W})|p(\boldsymbol{W})] \qquad (S6)$$

The first term is the expected log-likelihood, encouraging the model to fit the data. The second term is the KL divergence between the variational posterior and the prior, acting as a regularize that prevents overfitting.

S3.3 KL divergence for Gaussian distributions

We assume both the prior and variational posterior are factorized Gaussian distributions:

$$p(\boldsymbol{W}) = \prod_i \mathcal{N}\left(w_i \middle| 0, \sigma_p^2\right) \tag{S7}$$

$$q_\theta(\boldsymbol{W}) = \prod_i \mathcal{N}(w_i | \mu_i, \sigma_i^2) \tag{S8}$$

where σ_p^2 is the prior variance (optimized via PSO), and μ_i and σ_i^2 are the variational parameters for each weight. The KL divergence between two univariate Gaussians has a closed-form solution:

$$D_{\mathrm{KL}}\left[\mathcal{N}(\mu, \sigma^2) | \mathcal{N}\left(0, \sigma_p^2\right)\right] = \log \frac{\sigma_p}{\sigma} + \frac{\sigma^2 + \mu^2}{2\sigma_p^2} - \frac{1}{2} \tag{S9}$$

The total KL divergence is the sum over all weights:

$$D_{\mathrm{KL}}[q_\theta(\boldsymbol{W}) | p(\boldsymbol{W})] = \sum_i \left(\log \frac{\sigma_p}{\sigma_i} + \frac{\sigma_i^2 + \mu_i^2}{2\sigma_p^2} - \frac{1}{2}\right) \tag{S10}$$

S3.4 Reparameterization Trick

Direct sampling from $q_\theta(\boldsymbol{W})$ is not differentiable with respect to θ. The reparameterization trick (Blundell et al., 2015) addresses this by expressing a sample as a deterministic transformation of a noise variable:

$$w_i = \mu_i + \sigma_i \cdot \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0,1) \tag{S11}$$

This transformation shifts the stochasticity from the weights to the auxiliary variable epsilon, enabling backpropagation.

S3.5 Heteroscedastic aleatoric uncertainty

To capture data-dependent (aleatoric) uncertainty, our network outputs both the predictive mean $\mu(\boldsymbol{x})$ and the predictive variance $\sigma^2(\boldsymbol{x})$. The negative log-likelihood (NLL) for a single observation is:

$$-\log p\,(y|\boldsymbol{x},\boldsymbol{W}) = \frac{\big(y-\mu(\boldsymbol{x})\big)^2}{2\sigma^2(\boldsymbol{x})} + \frac{1}{2}\log\sigma^2\,(\boldsymbol{x}) + \frac{1}{2}\log(2\pi) \qquad (S12)$$

In practice, the network outputs $\log\sigma^2\,(\boldsymbol{x})$ for numerical stability. The data loss term becomes:

$$L\text{NLL} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\big(y_i-\mu(\boldsymbol{x_i})\big)^2}{2\sigma^2(\boldsymbol{x_i})} + \frac{1}{2}\log\sigma^2\,(\boldsymbol{x_i})\right] \qquad (S13)$$

S3.6 Total loss function

The complete loss function combines the heteroscedastic negative log-likelihood with the KL divergence term:

$$Ltotal = LNLL + \beta\cdot D_{\mathrm{KL}}[q_\theta(\boldsymbol{W})|p(\boldsymbol{W})] \qquad (S14)$$

where β is a weighting coefficient. Both beta and the prior variance σ_p^2 are optimized via Particle Swarm Optimization (PSO).

S3.7 Predictive uncertainty decomposition

For a new input x*, the predictive distribution is obtained by marginalizing over the posterior:

$$p(y^*|\boldsymbol{x}^*,\mathcal{D}) = \int p(y^*|\boldsymbol{x}^*,\boldsymbol{W})q_\theta(\boldsymbol{W})d\boldsymbol{W} \qquad (S15)$$

Approximated via Monte Carlo sampling (T=50):

$$E[y^*] \approx \frac{1}{T}\sum_{t=1}^{T} \mu_t(\boldsymbol{x}^*) \tag{$S16$}$$

The total predictive uncertainty for a single BNN model is decomposed as:

$$\underbrace{\mathrm{Var}[y^*]}_{\text{Total}} = \underbrace{\frac{1}{T}\sum_{t=1}^{T} \sigma_t^2(\boldsymbol{x}^*)}_{\text{Aleatoric}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T} (\mu_t(\boldsymbol{x}^*) - E[y^*])^2}_{\text{Epistemic}} \tag{$s17$}$$

where the aleatoric component represents irreducible data noise, and the epistemic component represents model uncertainty.

S3.8 Ensemble Uncertainty Aggregation (Cross-Validation)

To further improve robustness, we employ a K-fold cross-validation ensemble (K=10). The final uncertainty incorporates both intra-model (BNN) uncertainty and inter-model (ensemble) variance.

1. Aleatoric Uncertainty (Ensemble Average):

$$\sigma_{\text{alea}}^2(\boldsymbol{x}^*) = \frac{1}{K}\sum_{k=1}^{K} \sigma_{\text{alea},k}^2(\boldsymbol{x}^*) \tag{$S18$}$$

2. Epistemic Uncertainty (Ensemble Average):

$$\sigma_{\text{epis}}^2(\boldsymbol{x}^*) = \frac{1}{K}\sum_{k=1}^{K} \sigma_{\text{epis},k}^2(\boldsymbol{x}^*) \tag{$S19$}$$

3. Inter-model Variance (Structural Uncertainty):

$$\sigma_{\text{inter}}^2(\boldsymbol{x}^*) = \frac{1}{K}\sum_{k=1}^{K} \left(\mu_k(\boldsymbol{x}^*) - \hat{\mu}(\boldsymbol{x}^*)\right)^2 \tag{$S20$}$$

The total ensemble uncertainty is computed as the quadrature sum:

$$\sigma_{\text{total}}(\boldsymbol{x}^*) = \sqrt{\sigma_{\text{alea}}^2(\boldsymbol{x}^*) + \sigma_{\text{epis}}^2(\boldsymbol{x}^*) + \sigma_{\text{inter}}^2(\boldsymbol{x}^*)} \tag{S21}$$

S4 Model Performance

S4.1 Coefficient of variation

To assess the reliability of the PSODNN predictions, we calculated the Coefficient of Variation (CV) for each grid point based on the ensemble spread of the five cross-validation folds. Here, the CV is defined as the ratio of the ensemble standard deviation ($\sigma_{\text{ensemble}}(\boldsymbol{x})$) to the mean predicted GHF ($\hat{\mu}(\boldsymbol{x})$):

$$\text{CV}(x) = \frac{\sigma_{\text{ensemble}}(x)}{\hat{\mu}(x)} \times 100\% \tag{S22}$$

This metric specifically quantifies the epistemic uncertainty arising from model variability. A lower CV indicates high agreement among the five independent training folds, suggesting robust prediction performance despite the scarcity of training data.

We categorized the prediction reliability into four levels and the statistical results (Table S3) reveal that the majority of the Antarctic continent is predicted with high confidence by our ensemble model:

Table S2. Distribution of prediction confidence levels across Antarctica

Confidence Category	Count of Grid Points	Percentage
High confidence (< 15%)	28,961	**67.13%**
Moderate confidence (15–30%)	14,086	**32.65%**
Low confidence (30–50%)	93	**0.22%**
Very low confidence (> 50%)	0	**0.00%**

S4.2 Uncertainty decomposition

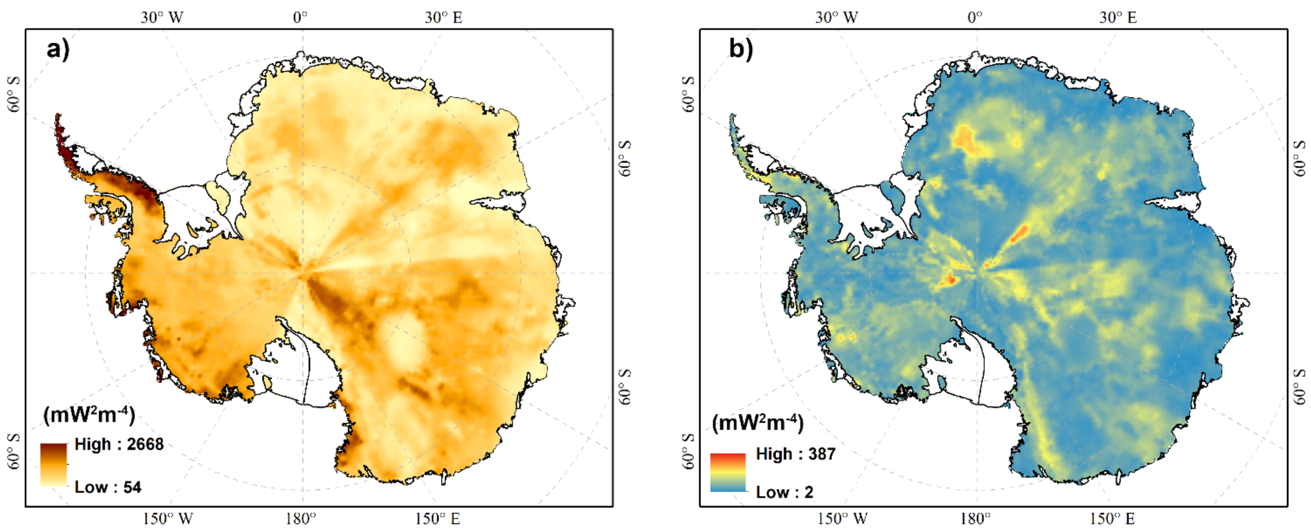

Figure S14. Spatial distribution of the decomposed predictive uncertainty variances across Antarctica. (a) Aleatoric uncertainty variance (mW^2m^{-4}). (b) Epistemic uncertainty variance (mW^2m^{-4}).

References

- Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D.: Weight Uncertainty in Neural Networks, in: Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), Lille, France, 6–11 July 2015, PMLR, 37, 1613–1622, 2015.
- Gal, Y.: Uncertainty in Deep Learning, PhD thesis, University of Cambridge, Cambridge, UK, 2016.
- Kendall, A. and Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, in: Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017, 5574–5584, 2017.
- NOAA National Centers for Environmental Information: ETOPO 2022 15 Arc-Second Global Relief Model, NOAA [dataset], https://doi.org/10.25921/fd45-gt74, 2022.
- Szwillus, W., Afonso, J. C., Ebbing, J., and Mooney, W. D.: Global crustal thickness and velocity structure from geostatistical analysis of seismic data, J. Geophys. Res.-Solid Earth, 124, 1626–1652, https://doi.org/10.1029/2018JB016593, 2019.
- An, M., Wiens, D. A., Zhao, Y., Feng, M., Nyblade, A. A., Kanao, M., and Lévêque, J. J.: S-velocity model and inferred Moho topography beneath the Antarctic Plate from Rayleigh waves, J. Geophys. Res.-Solid Earth, 120, 359–383, https://doi.org/10.1002/2014JB011332, 2015.
- Afonso, J. C., Salajegheh, F., Szwillus, W., Ebbing, J., and Gaina, C.: A global reference model of the lithosphere and upper mantle from joint inversion and analysis of multiple data sets, Geophys. J. Int., 217, 1602–1628, https://doi.org/10.1093/gji/ggz094, 2019.

- Pappa, F., Ebbing, J., Ferraccioli, F., and van der Wal, W.: Modeling satellite gravity gradient data to derive density, temperature, and viscosity structure of the Antarctic lithosphere, J. Geophys. Res.-Solid Earth, 124, 12053–12076, https://doi.org/10.1029/2019JB018037, 2019.
- Laske, G., Masters, G., Ma, Z., and Pasyanos, M. E.: CRUST1.0: An updated global model of Earth's crust, EGU General Assembly 2012, Vienna, Austria, 22–27 April 2012, EGU2012-3743, 2012.
- Schaeffer, A. J., and Lebedev, S.: Global heterogeneity of the lithosphere and underlying mantle: A seismological appraisal based on multimode surface-wave dispersion analysis, shear-velocity tomography, and tectonic regionalization, in: The Earth'sHeterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective, edited by Khan, A., and Deschamps, F., Springer, Cham, 3–46, https://doi.org/10.1007/978-3-319-15627-9_1, 2015.
- Scheinert, M., Ferraccioli, F., Schwabe, J., Bell, R., Studinger, M., Damaske, D., and Richter, T. D.: New Antarctic gravity anomaly grid for enhanced geodetic and geophysical studies in Antarctica, Geophys. Res. Lett., 43, 600–610, https://doi.org/10.1002/2015GL067439, 2016.
- Bonvalot, S., Briais, A., Kuhn, M., Peyrefitte, A., Vales, N., Biancale, R., and Sarrailh, M.: World Gravity Map (WGM2012),Bureau Gravimétrique International, 2012.
- Li, C. F., Lu, Y., and Wang, J.: A global reference model of Curie-point depths based on EMAG2, Sci. Rep., 7, 45129, https://doi.org/10.1038/srep45129, 2017.
- Martos, Y. M., Catalán, M., Jordan, T. A., Golynsky, A., Golynsky, D., Eagles, G., and Vaughan, D. G.: Heat flux distribution of Antarctica unveiled, Geophys. Res. Lett., 44, 11417–11426, https://doi.org/10.1002/2017GL075609, 2017.
- Maus, S., Sazonova, T., Hemant, K., Fairhead, J. D., and Ravat, D.: National Geophysical Data Center candidate for the World Digital Magnetic Anomaly Map, Geochem. Geophys. Geosyst., 8, Q06017, https://doi.org/10.1029/2007GC001643, 2007.
- Golynsky, A. V., Ferraccioli, F., Hong, J. K., Golynsky, D. A., von Frese, R. R. B., Young, D. A., Blankenship, D. D., Holt, J. W., Ivanov, S. V., Kiselev, A. V., Masolov, V. N., Eagles, G., Gohl, K., Jokat, W., Damaske, D., Finn, C., Aitken, A.,Bell, R. E., Armadillo, E., Jordan, T. A., Greenbaum, J. S., Bozzo, E., Caneva, G., Forsberg, R., Ghidella, M., Galindo-Zaldivar, J., Bohoyo, F., Martos, Y. M., Nogi, Y., Quartini, E., Kim, H. R., and Roberts, J. L.: New magnetic anomaly map of the Antarctic, Geophys. Res. Lett., 45, 6437–6449, https://doi.org/10.1029/2018GL078153, 2018.
- Ebbing, J., Haas, P., Ferraccioli, F., Pappa, F., Szwillus, W., and Bouman, J.: Earth tectonics as seen by GOCE-Enhanced satellite gravity gradient imaging, Sci. Rep., 8, 16356, https://doi.org/10.1038/s41598-018-34733-9, 2018.