

SUPPLEMENTAL MATERIAL FOR

GLOBAL PERSISTENCE OF INTRAPLATE HEAT-FLOW ANOMALIES REVEALS LIMITS OF TECTONIC DISTANCE CONTROL

Cristiano Gaujac¹

¹ Independent Researcher, Villefranche-sur-Saône, France

Supplementary Methods

This supplement consolidates the technical support material for the study, with emphasis on implementation details, geometric and methodological audits, sensitivity checks, and reproducibility safeguards that underpin the main-text protocol without repeating its core narrative.

S1. Data ingestion and validation

Latitude, longitude, and heat-flow variables were identified in the 2024 IHFC database (Fuchs et al., 2023) through flexible column-name matching in order to accommodate variations in the database schema. After identification, the dataset was converted to geospatial format (GeoDataFrame) in WGS84 (EPSG:4326).

Validation included explicit numeric type coercion and removal of records with missing, non-numeric, or non-physical values ($q \leq 0$). No interpolation, smoothing, or model-based correction was applied, ensuring that the final dataset represented valid observational measurements only.

S2. Distance calculation and geometric audit

Distances between IHFC points and tectonic structures, defined from the GEM Global Active Faults Database (Styron and Pagani, 2020) and the PB2002 plate-boundary model (Bird, 2003), were calculated in an operational metric projection (EPSG:3857) using point-to-line operations with spatial indexing to ensure computational efficiency at the global scale.

The accuracy of this approximation was evaluated through an independent geometric audit, in which stratified subsets were reprocessed using local azimuthal equidistant (AEQD) projections centered on each observation. Differences between distances calculated in EPSG:3857 and AEQD were quantified and were typically on the order of tens of kilometers.

The distribution of these differences was examined for systematic bias and for potential effects on the intraplate classification based on the 110 km threshold, confirming that projection-induced distortions do not affect classification at the scale adopted in this protocol.

S2.1 Consistency between tectonic models (GEM vs. PB2002)

To assess consistency between independent tectonic datasets, distances derived from the GEM Global Active Faults Database and the PB2002 plate-boundary model were compared on a point-by-point basis.

Agreement between the two models was evaluated through correlation metrics and distributional comparisons of `dist_gem_km` and `dist_pb_km`. The aim was to test whether the conservative definition $\text{dist_min_km} = \min(\text{dist_gem_km}, \text{dist_pb_km})$ could introduce systematic bias because of differences in tectonic representation.

The results indicate substantial agreement between the two models, with no systematic divergence capable of materially affecting the classification of intraplate observations (Fig. SM1).

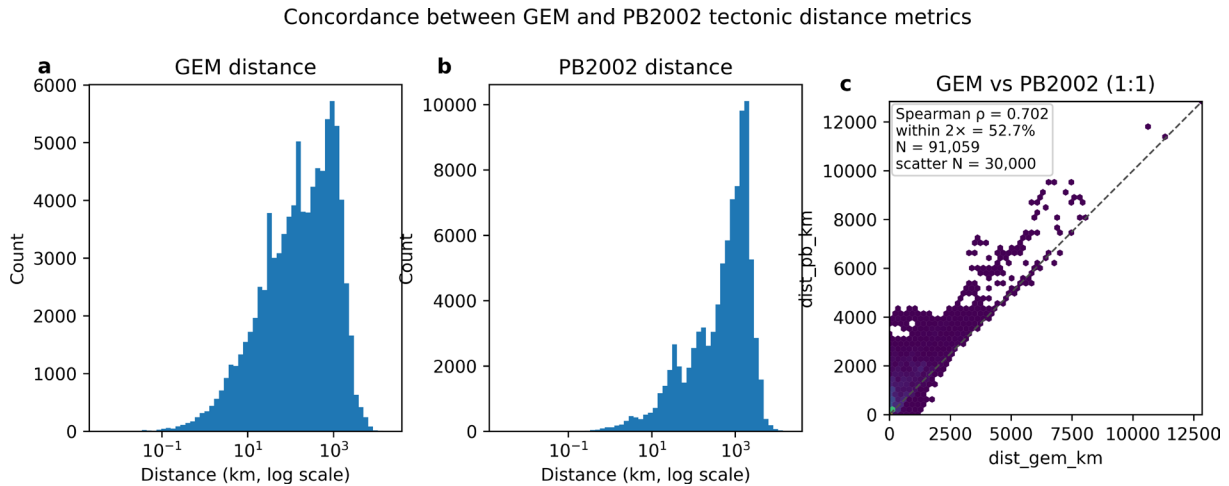


Figure SM1. Agreement between tectonic distance metrics derived from GEM and PB2002. Panels (a) and (b) show the global distributions of site-to-structure distances obtained from the GEM Global Active Faults Database and the PB2002 plate-boundary model, respectively. Panel (c) compares `dist_gem_km` and `dist_pb_km` for a deterministic subsample ($N = 30,000$), with the dashed line indicating 1:1 agreement. The overall concordance indicates that intraplate classification is not strongly dependent on a single tectonic dataset and supports the conservative use of $\text{dist_min_km} = \min(\text{dist_gem_km}, \text{dist_pb_km})$.

S3. Diagnostic analysis of the intraplate threshold

The validated dataset was stratified into distance intervals (0-50, 50-100, 100-150, 150-300, and 300-1000 km), and for each interval the median and 90th percentile of q were calculated.

The evolution of these statistics across distance classes was examined to identify where thermal behavior shifts from a regime strongly influenced by tectonic margins to one that is comparatively more stable. This diagnostic therefore defines the transition range used to motivate the operational intraplate cutoff.

Sensitivity tests using alternative thresholds were then used to assess the stability of the thermal statistics across an approximate range of 80 to 150 km, placing the adopted value (110 km) within a statistically robust interval (Fig. SM2).

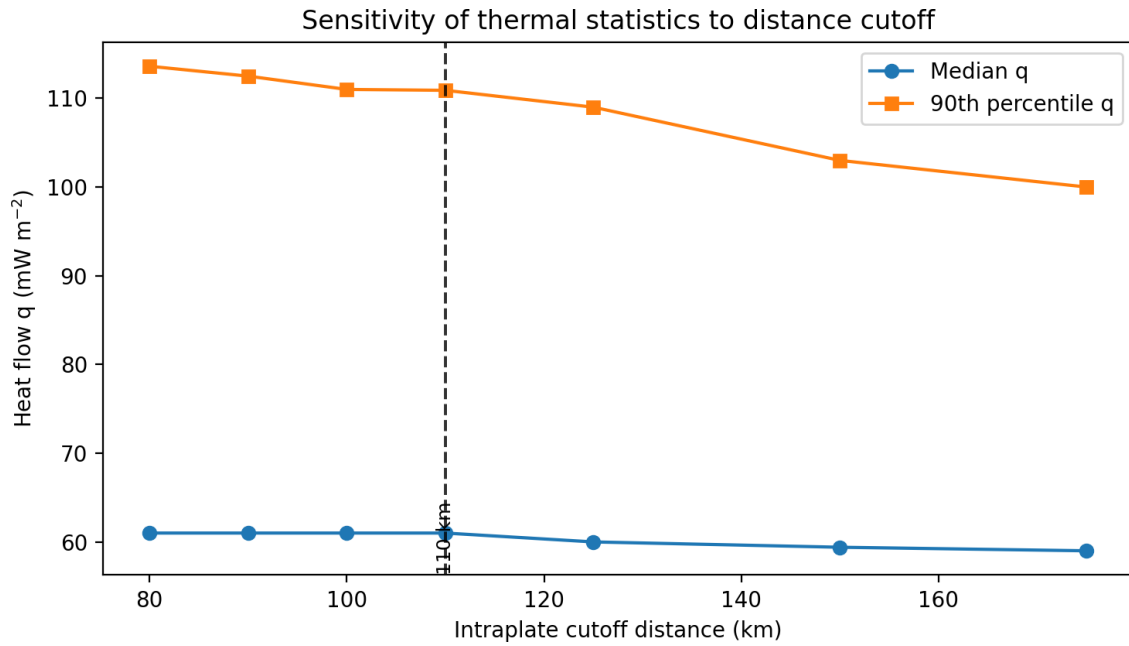


Figure SM2. Sensitivity of thermal statistics to the intraplate cutoff. Median and 90th percentile heat-flow values (q) for intraplate subsets defined by alternative threshold values of `dist_min_km`. The dashed vertical line indicates the adopted cutoff (110 km). The relatively limited variation of both summary statistics across the tested interval shows that the selected threshold falls within a statistically stable transition range, supporting its use for defining the canonical intraplate domain.

S3.1 Distributional properties of the heat-flow data

The global distribution of heat-flow values was evaluated in both linear and logarithmic space in order to characterize its skewness and its suitability for statistical modeling.

The distribution in linear space shows strong right skewness, with a long tail of elevated values, whereas the logarithmic transformation yields a more symmetric distribution and stabilizes the variance.

This behavior supports the use of $\log_{10}(q)$ in the regression analyses and justifies adoption of a log-log analytical framework (Fig. SM3).

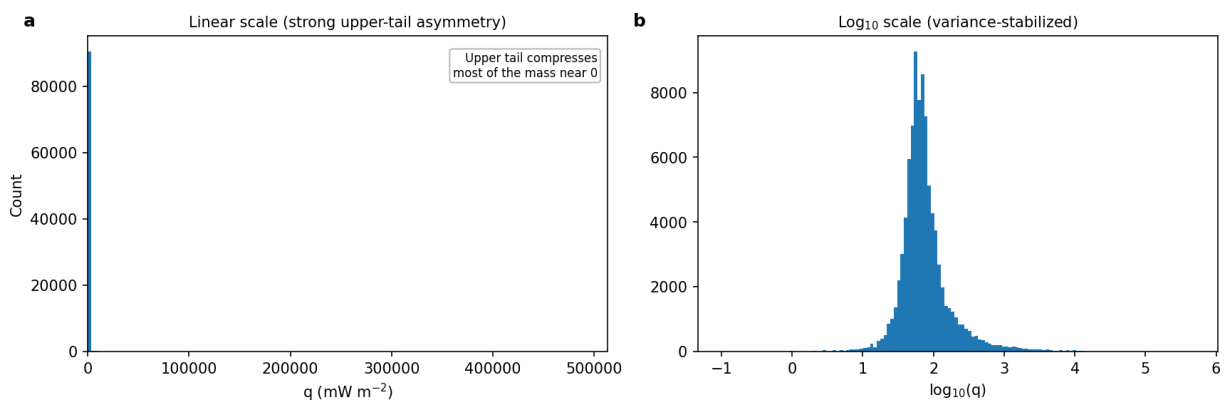


Figure SM3. Global heat-flow distribution after physical validation. Histograms of q for all physically valid stations ($q > 0$, $N = 91,059$) are shown on (a) a linear scale and (b) a logarithmic scale. The linear representation emphasizes the pronounced right skewness and extended upper tail of the global heat-flow distribution, whereas the logarithmic representation reduces asymmetry and stabilizes variance. Together, these diagnostics support the use of $\log_{10}(q)$ in the analytical framework adopted in the manuscript.

S4. Analytical structure and canonical dataset

The primary inference on the relationship between heat flow and tectonic distance was conducted on the canonical point-by-point dataset using ordinary least-squares (OLS) regression in log-log space with HC3 robust standard errors. The principal functional form was $\log_{10}(q)$ versus $\log_{10}(\text{dist_min_km} + 1)$, where the additive constant +1 prevents a singularity at $\log_{10}(0)$ while preserving continuity at short distances.

As supporting robustness analyses, the global signal was also reassessed after spatial aggregation of observations into metric sites (~ 0.5 -2 km), using regressions weighted by the number of measurements per site and sensitivity checks with respect to spatial resolution. These analyses are intended to support, not replace, the primary inference based on the canonical point-by-point dataset.

The canonical intraplate dataset was constructed by applying the criterion $\text{dist_min_km} \geq 110$ km directly to the point-by-point catalog, without spatial aggregation at this stage. The resulting artifact was deterministically ordered and linked to a cryptographic integrity identifier, allowing independent verification of the exact dataset used in the main analyses.

S5. Multi-criterion filtering (A/B/C): implementation and diagnostic structure

The subset of intraplate anomalies was subjected to three independent rejection criteria, all applied strictly at point level in accordance with the canonical protocol summarized in the main text.

Criterion A (shallow seismicity): for each point, the distance to the nearest shallow seismic event and the annual event rate within a 100 km radius were calculated (magnitude ≥ 3.5 ; depth ≤ 35 km). Rejection was assigned when the minimum distance to a shallow event was < 50 km or when the local rate was ≥ 2 events/year.

Criterion B (lithospheric deformation): values of the *sir* field (GSRM v2.1; Kreemer et al., 2014) were obtained by direct spatial interpolation, and an operational threshold of 1×10^{-9} was adopted to identify regions of elevated deformation.

Criterion C (Bouguer gravity): for each point, Bouguer gravity values (World Gravity Map; Bonvalot et al., 2012) were used to construct a local reference distribution within a 300-500 km annulus, summarized by the median and scaled MAD. A robust z-score was then calculated, with rejection defined by $|z| \geq 2$.

The three criteria were applied independently. The final retained subset therefore consists of observations not eliminated by operational proxies for shallow seismicity, elevated lithospheric deformation, or anomalous gravity context, reducing the likelihood that the retained anomalies are dominated by a single local geophysical control.

S5.1 Exploratory clustering-concordance analysis

The clustering-concordance diagnostic was conducted on all 53,420 canonical intraplate observations using the standardized two-variable feature space $\log_{10}(q)$ and $\log_{10}(\text{dist_min_km} + 1)$. K-means clustering (MacQueen, 1967) provided the reference partition with $k = 4$, $\text{random_state} = 42$, and $n_init = 50$. Alternative partitions were generated using Ward hierarchical clustering with $k = 2$ (Ward, 1963) and DBSCAN with $\text{eps} = 0.833$ and $\text{min_samples} = 10$ (Ester et al., 1996).

Because cluster labels are arbitrary, Ward and DBSCAN labels were aligned to the K-means label space by optimal assignment before pointwise comparison. Agreement was summarized using the adjusted Rand index and normalized mutual information, with DBSCAN noise retained explicitly; the adjusted Rand index corrects pairwise agreement for chance (Hubert and Arabie, 1985). Each observation was then classified as showing full, partial, or divergent agreement across methods. This analysis is exploratory and tests the stability of discrete groupings; it does not replace the regression-based inference.

S6. Reproducibility framework

All analyses were performed in a deterministic computational environment. The canonical intraplate dataset was ordered using fixed structural keys and associated with a cryptographic integrity signature, allowing independent verification of the identity and consistency of the analytical artifact used in the study.

Throughout the pipeline, exploratory and canonical artifacts were kept strictly separate, and protocol parameters were explicitly recorded. This framework provides traceability, reproducibility, and consistency between the results summarized in the main text and the technical material documented here.

References

- Bird, P. (2003). An updated digital model of plate boundaries, *Geochem. Geophys. Geosyst.*, 4, 3, 1027, doi:10.1029/2001GC000252.
- Bonvalot, S., G. Balmino, A. Briais, M. Kuhn et al. (2012). World Gravity Map, BGI-CGMW-CNES-IRD, Paris, doi:10.18168/bgi.23.
- Ester, M., H. P. Kriegel, J. Sander and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231.
- Fuchs, S., B. Norden, F. Neumann, N. Kaul et al. (2023). Quality assurance of heat-flow data: The new structure and evaluation scheme of the IHFC Global Heat Flow Database, *Tectonophysics*, 863, 229976, doi:10.1016/j.tecto.2023.229976.
- Hubert, L. and P. Arabie (1985). Comparing partitions, *J. Classif.*, 2, 193-218, doi:10.1007/BF01908075.
- Kreemer, C., G. Blewitt and E. C. Klein (2014). A geodetic plate motion and Global Strain Rate Model, *Geochem. Geophys. Geosyst.*, 15, 10, 3849-3889, doi:10.1002/2014GC005407.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.
- Styron, R. and M. Pagani (2020). The GEM Global Active Faults Database, *Earthq. Spectra*, 36, 1, 160-180, doi:10.1177/8755293020944182.
- Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, 58, 301, 236-244, doi:10.1080/01621459.1963.10500845.