Along-Strike Heat Flow Continuity Near Subduction Zones Is Mostly Ambiguous, Not Globally Uniform

Buchanan Kerswell[1], Matthew J. Kohn[2]

[1]Department of Earth, Oceans and Ecological Sciences, University of Liverpool, Liverpool L69 3BX, UK
[2]Department of Geosicences, Boise State University, Boise, ID 83725

Key Points:

- Along-strike surface heat flow continuity is geographically limited; four domains show good continuity, three show clear discontinuity
- Kriging and Similarity exhibit complementary failure modes; method consensus does not guarantee accuracy
- Nearly 3/4 of all transects had ≤ 25 observations; such sparse observations are the dominant constraint on geodynamic interpretation

Corresponding author: Buchanan Kerswell, b.kerswell@liverpool.ac.uk

Abstract

A common class of subduction zone thermal models assumes slab-mantle coupling near 70–80 km depth and thin backarc lithospheres, based on surface heat flow profiles from a small number of well-instrumented subduction zone segments. These models implicitly assume along-strike continuity over hundreds of kilometers. We tested this assumption by applying ordinary Kriging to 60,316 surface heat flow measurements across 34 subduction zone domains, comparing Krige and Similarity interpolations point-by-point, and applying cross-correlation function (CCF) analysis to 201 adjacent trench-perpendicular transect pairs. Four well-sampled domains (Japan, Cascadia, Sumatra, and Kermadec–Hikurangi) show sustained continuity consistent with observations and both interpolation methods. Three other well-sampled domains (Kurils–Kamchatka, Central America–Panama, and Ryukyus–Sagami) show abrupt along-strike changes at 100–150 km scales, rendering single-profile characterizations misleading. Two domains (Caribbean and Izu–Bonin) exhibit systematic disagreement between methods. Most transects (168/235; 71%) have ≤ 25 observations, which is too sparse for independent assessment. In well-sampled settings, Similarity mischaracterizes along-strike correlations while Kriging tracks local observations closely. In sparse settings, only Similarity provides plausible surface heat flow estimates. Critically, method agreement does not guarantee accuracy. For Izu–Bonin, 190 local observations contradict both interpolation methods. For the Caribbean, the two methods displace a structural boundary by 100–150 km relative to its inferred location. This analysis shows that the accuracy of thermal models cannot be determined from the existing surface heat flow record. Improvement will require strategically targeted surveys and hybrid Kriging–Similarity approaches guided by the failure-mode diagnostics defined here.

Plain Language Summary

When one tectonic plate dives beneath another at a subduction zone, the two plates lock together and drive flow of warm rock in Earth's mantle. Scientists measure the heat escaping through Earth's surface to locate this "sticking-point" because its depth controls fault behavior and earthquake hazards. For decades, a few intensely studied regions were used to identify a sticking-point depth of 70–80 km, assuming that a single cross-section could represent the thermal structure for hundreds of kilometers laterally. We tested this assumption by analyzing more than 60,000 heat flow measurements across

34 subduction zone segments worldwide using two independent mapping methods. Thermal structure was consistent in four well-studied regions, supporting the classic model, but thermal structure changed abruptly over distances of 100–150 km in three other well-sampled regions, challenging the classic model's assumptions. Two additional regions gave contradictory results, while nearly three quarters of all regions we analyzed had too few measurements to evaluate thermal structure. Importantly, in one region, both methods indicate a uniform thermal structure while 190 measurements directly contradict that conclusion; agreement does not prove accuracy. Improving the subduction zone thermal models requires strategically placed field surveys combined with mapping strategies that integrate local measurements and global geological patterns in a principled way.

Definition of Symbols

Parameter	Symbol	Unit	Equations
Spatial Field			
Spatial location	u	km	1, S1, S3, S4, S5
Observation (surface heat flow) at u	$Z(u)$	mW m^{-2}	1, S1, S5
Mean surface heat flow	μ	mW m^{-2}	S1
Covariance function	$C(h)$	mW^2 m^{-4}	S1, S5
Raw sample count	N_{obs}	-	-
Variogram Model			
Separation distance (lag) between observations	h	km	1, 2, S1, S2, S6
Maximum separation distance	$\max(h)$	km	2, S6
Lag cutoff constant	κ	-	2, S6
No. of lag intervals	n	-	2, S6, S7
Lag bin width	δ	km	S6
Observation pair count at lag h	$N(h)$	-	1, 2
Experimental variogram	$\hat{\gamma}(h)$	mW^2 m^{-4}	1
Variogram model	$\gamma(h)$	mW^2 m^{-4}	1, S2
Variogram model type	Sph, Exp, Gau	-	S2
Nugget variance	τ^2	mW^2 m^{-4}	S2
Sill variance	σ^2	mW^2 m^{-4}	S2
Effective range	a	km	S2
Kriging Estimation			
Krige estimate at u	$\hat{Z}(u)$	mW m^{-2}	S3, S4, S5
No. of observations used for Kriging	n_K	-	S3, S4, S5
Kriging weight	λ_i	-	S3, S4, S5
Kriging variance	$\sigma_K^2(u)$	mW^2 m^{-4}	S5

Parameter	Symbol	Unit	Equations
Maximum neighborhood observations	m	-	-
Maximum neighborhood size	d_{max}	km	-
Optimization			
Optimization objective	J	-	S7
Variogram weighted least squares error	$\mathrm{WLS}_{\mathrm{error}}$	$\mathrm{mW}^4\ \mathrm{m}^{-8}$	S7
Standard deviation of experimental variogram	σ_{γ}	$\mathrm{mW}^2\ \mathrm{m}^{-4}$	S7
Kriging Root Mean Square Error	$\mathrm{RMSE}_{\mathrm{Krg}}$	$\mathrm{mW\ m}^{-2}$	S7
Standard deviation of cross-validation	σ_{cv}	$\mathrm{mW\ m}^{-2}$	S7

1 Introduction

Surface heat flow reflects the integrated thermal state of the lithosphere and contributions from hydrothermal circulation, radioactive decay, fault motion, and mantle convection (Fourier, 1827; Furlong & Chapman, 2013; Hutnak et al., 2008; Kelvin, 1863; Stein & Stein, 1994). Observations of surface heat flow are therefore central for understanding lithospheric evolution (Hasterok, 2013; Parsons & Sclater, 1977; Pollack & Chapman, 1977; Rudnick et al., 1998; Stein & Stein, 1992), subduction dynamics and seismic hazards (Currie et al., 2004; Currie & Hyndman, 2006; Furukawa, 1993; Gao & Wang, 2014; Kohn et al., 2018; Wada & Wang, 2009), groundwater hydrology (Hutnak et al., 2008; Stein & Stein, 1994), and exploration of economic resources including geothermal energy (Batir et al., 2016; Majorowicz & Grasby, 2010). Decades of community effort have produced tens of thousands of continental and oceanic measurements compiled into global databases (Hasterok & Chapman, 2008; Jennings et al., 2021; Lucazeau, 2019; Pollack et al., 1993) that underpin multidisciplinary investigations of lithospheric and crustal processes.

A conventional model of subduction zone thermal structure has emerged from studies that calibrate numerical thermomechanical models against individual trench-perpendicular surface heat flow profiles (Currie et al., 2004; Currie & Hyndman, 2006; Furukawa, 1993; Hyndman et al., 2005; Wada & Wang, 2009), concluding that slab-mantle coupling near 70–80 km depth and a thin backarc lithosphere are common features of active margins (Abers et al., 2017; Holt & Condit, 2021; Syracuse et al., 2010; van Keken et al., 2011, 2018). The model relies on the implicit but critical assumption that a small number of well-instrumented transects are spatially representative for hundreds of kilometers along strike. If proximal transects instead show markedly different surface heat flow profiles, as we test here, this assumption breaks down. This raises a fundamental question: do subduction dynamics genuinely vary over tens to hundreds of kilometers along strike, or do near-surface perturbations and observational gaps merely obscure a continuous thermal signature?

Investigating this geodynamic problem requires evaluating transect-scale surface heat flow patterns across broad regions along strike. The most recent global compilation from the International Heat Flow Commission (IHFC 2024) contains 91,182 surface heat flow observations (Global Heat Flow Data Assessment Group et al., 2024), yet regional coverage near subduction zones remains sparse and highly uneven at scales of 10–1000 km. High-resolution local surveys, often collocated with seismic monitoring networks, capture fine-scale variations at individual sites but do not resolve along-strike variability across subduction zone segments. The critical methodological challenge is therefore evaluating sparse, irregularly spaced observations separated by distances ranging from 0.1–1000 km.

To address both the geodynamic problem and the methodological challenge, we evaluated two independent interpolation methods that represent end-member approaches built on fundamentally different principles: *Kriging* (Krige, 1951) and *Similarity* (Goutorbe et al., 2011). Their differences are highly consequential for understanding lithospheric thermal structure, detecting surface heat flow anomalies, assessing the practical limitations of each approach, and motivating the development of hybrid methods. The rationale for applying both methods follows from the *Three Laws of Geography*:

1. Everything is related, but near things are more related than distant things (Matheron, 1963)
2. Geographic phenomena are inherently heterogeneous (Goodchild, 2004)
3. Localities with similar geographic configurations share other attributes (Zhu et al., 2018)

Kriging exploits the First Law by estimating surface heat flow in data-sparse regions using a linear combination of nearby observations, weighted by their spatial correlations. This reliance on spatial correlation matches the physics of surface heat flow, which reflects large-scale lithospheric thermal structure and heat-transfer processes. For instance, broad regions of low surface heat flow on continents outline cratons (Nyblade & Pollack, 1993); anomalously low oceanic surface heat flow indicates heat extraction by hydrothermally circulating seawater (Fisher & Becker, 2000; Hasterok et al., 2011; Hutnak et al., 2008; Stein & Stein, 1994); and trench-perpendicular surface heat flow profiles have been interpreted as evidence for broadly uniform slab-mantle coupling depths (Furukawa, 1993; Wada & Wang, 2009) and upper-plate lithospheric thickness (Currie et al., 2004; Currie & Hyndman, 2006; Hyndman et al., 2005).

Similarity, in contrast, does not consider spatial correlations. It predicts surface heat flow at any target location by comparing that location's *geographic configuration* (the combination of geologic and geophysical attributes) against the entire global database of measured surface heat flow (Goutorbe et al., 2011). This reflects the Third Law: localities sharing similar bathymetry, lithology, proximity to orogens, seafloor age, and upper-mantle velocity tend to produce similar surface heat flow (Chapman & Pollack, 1975; Davies, 2013; Lee & Uyeda, 1965; Lucazeau, 2019; Sclater & Francheteau, 1970; Shapiro & Ritzwoller, 2004). Similarity estimates are therefore geologically reasoned estimates rather than spatial interpolations.

These considerations motivate a direct test of the assumption underlying the conventional subduction zone thermal model: that single surface heat flow profiles are regionally representative of thermal structure, and that profiles from different subduction systems share common features. Here we asked three specific questions: a) How spatially continuous is surface heat flow near active subduction zones? b) How do Krige and Similarity estimates compare along individual transects and between adjacent transects? c) What do methodological agreements and disagreements imply about geodynamic variability and observational gaps?

Our implementation applied Kriging to IHFC 2024 data near 235 predefined trench-perpendicular transects (organized into 34 transect sets) from the SubMap tool (Lallemand et al., 2026; Lallemand & Heuret, 2017), spanning four major subduction zone regions across the globe. We compared these Krige estimates point-by-point to global Similarity estimates from Lucazeau (2019), and used cross-correlation function (CCF) analysis on adjacent transects to quantify along-strike spatial continuity. The resulting analysis reveals both region-specific and method-specific patterns with implications for how surface heat flow observations constrain subduction zone dynamics.

2 Methods

2.1 Spatial Datasets

2.1.1 IHFC 2024 Surface Heat Flow Database

We used the IHFC 2024 database from the Global Heat Flow Data Assessment Group et al. (2024) for Kriging and CCF analysis. The database contains 91,182 surface heat flow observations from continental and oceanic settings (Figure 1), together with geographic coordinates, elevation, unique observation identifiers, and other metadata. Measurement quality metadata are encoded using the IHFC uncertainty (U) and measurement (M) classification scheme described by the Global Heat Flow Data Assessment Group et al. (2024).

We preprocessed the database prior to interpolation. First, we filtered observations using a simple range criterion, retaining only surface heat flow values within the interval [0, 250] mW m^{-2}. This step excluded non-physical values and extreme observations commonly associated with localized geothermal systems or measurement artifacts (Lucazeau, 2019). We applied no additional filtering or weighting based on IHFC quality classifications.

Second, we identified and resolved duplicate observations located at identical coordinates. Where duplicate records had different quality classifications, we retained the observation with the higher combined uncertainty and measurement quality score (U + M). When duplicate observations had equal quality scores, we selected one observation randomly.

After duplicate removal and range filtering, the processed IHFC 2024 dataset contained 60,316 observations used for Kriging and CCF analysis.

2.1.2 SubMap Transects and Interpolation Domains

We used SubMap transects as a standardized geographic framework for sampling and comparing surface heat flow. SubMap transects are predefined trench-perpendicular cross-sections through active subduction zones (Lallemand et al., 2026; Lallemand & Heuret, 2017). Transects were organized into 34 geographically contiguous sets spanning four regional groups: the N Pacific (NPA; 8 sets, 47 transects), S America (SAM; 10 sets, 66 transects), SE Asia (SEA; 12 sets, 90 transects), and the SW Pacific (SWP; 4 sets, 32 transects), for a total of 235 transects (Figure 1; Tables S1 and S2).

For each transect set, we defined an interpolation domain as a convex polygon enclosing all transects in the set. We sampled IHFC 2024 observations and interpolated surface heat flow grids within each domain (Figure 2: top row) for CCF analysis. We projected surface heat flow data onto each transect (Figure 2: bottom row) by computing the orthogonal projection distance from each data point to the transect line. A lateral buffer of 75 km on each side of each transect (150 km total buffer width) defined the projection window, limiting contributions from points far from the transect. We sam-

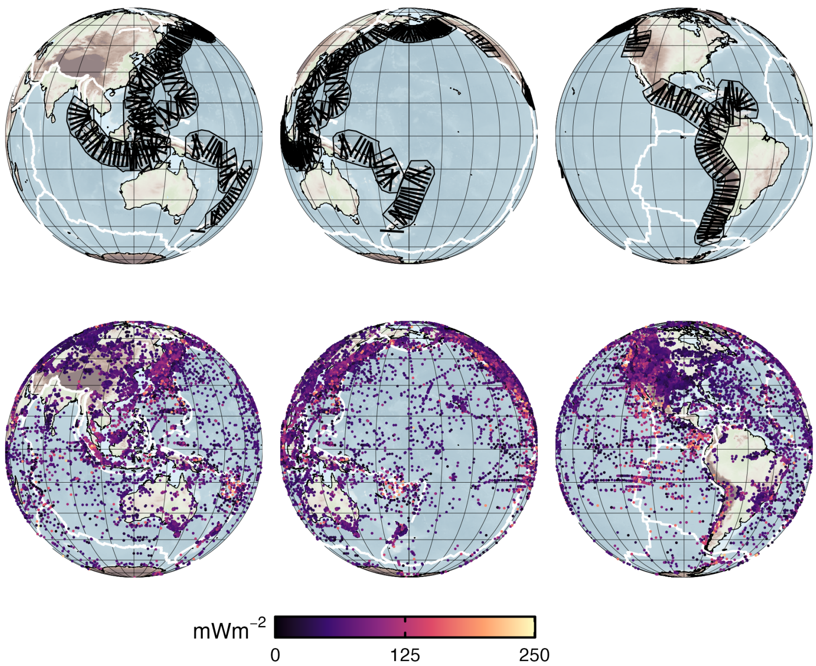

Figure 1: Subduction zone segments and global surface heat flow observations. A total of 235 trench-perpendicular transects (top row: bold black lines) and 91,182 raw, unfiltered surface heat flow observations (bottom row) were used for spatial analysis. Convex polygons drawn around 34 sets of transects (top row: shaded areas) defined spatial domains used for sampling and Kriging. Filtered and deduplicated data (N_{obs} = 60,316 total) show uneven, irregular distributions regionally and globally (Tables S1 and S2). Note the relatively high observation density in N America and around the N Pacific compared to the S Pacific. Plate boundaries (bold white lines) were defined by Coffin et al. (2018). Surface heat flow data are from the Global Heat Flow Data Assessment Group et al. (2024). Transects were defined by Lallemand et al. (2026).

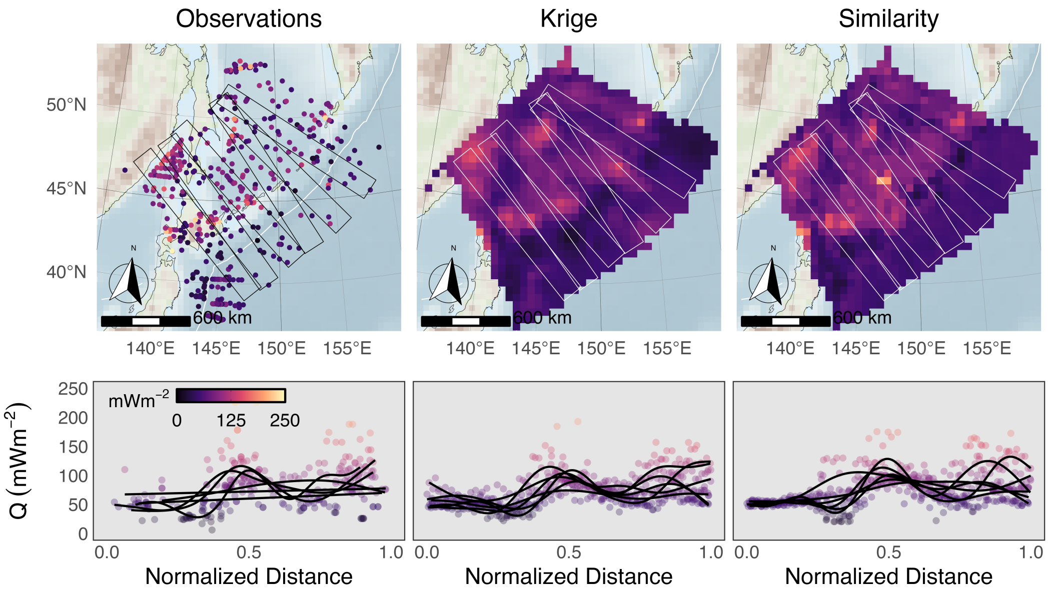

Figure 2: Example of surface heat flow interpolation and sampling within the Kurils transect set domain (NPA_SET2). IHFC 2024 observations (top left) were cropped within a convex polygon (not shown) enclosing the transect set. Krige estimates (top center) were made point-by-point across the same global 0.5° × 0.5° grid defined by Lucazeau (2019) for Similarity estimates (top right). Both interpolation grids were cropped within the same convex polygon as the IHFC 2024 observations. The 75 km buffers drawn around individual transects (top row: thin black or white rectangles) defined the projection window for the trench-perpendicular profiles used for CCF analysis (bottom row). Plate boundaries (white lines) were defined by Coffin et al. (2018). Surface heat flow data are from the Global Heat Flow Data Assessment Group et al. (2024). Transects were defined by Lallemand et al. (2026).

pled a total of 18,635 IHFC 2024 observations within the 34 transect set domains (Table S1), with 9,325 observations located within the projection windows of the 235 transects (Table S2).

2.1.3 Map Projection and Interpolation Grid

We processed spatial datasets using the R package `sf` (Pebesma, 2018). We imported and stored the IHFC 2024 observations, the Lucazeau (2019) Similarity interpolation, and the SubMap transects in geographic coordinates referenced to WGS84 (EPSG:4326).

We applied local azimuthal equidistant projections (ESRI:54032) where required for geometric buffering operations during construction of convex polygons, then transformed geometries back to geographic coordinates referenced to WGS84 (EPSG:4326).

We defined target locations for Kriging using the geometry of the global surface heat flow grid distributed with the Similarity interpolation of Lucazeau (2019), enabling direct point-by-point comparison between Krige estimates and Similarity estimates (e.g., see Figure 2: top row).

2.2 Kriging

We applied ordinary Kriging to IHFC 2024 observations within each convex polygon domain using the R package `gstat` (Gräler et al., 2016; Pebesma, 2004). Kriging is a three-step process that involves: 1) computing an experimental variogram $\hat{\gamma}(h)$ to characterize spatial correlations, 2) fitting a variogram model $\gamma(h)$ to the experimental variogram $\hat{\gamma}(h)$, and 3) solving a linear system of equations to predict surface heat flow at target locations (Cressie, 2015; Krige, 1951; Matheron, 1963). We computed the experimental variogram as:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i,j \in N(h)} \Big[Z(u_i) - Z(u_j) \Big]^2 \tag{1}$$

where $Z(u_i)$ and $Z(u_j)$ are observations at locations u_i and u_j separated by lag $h = \|u_i - u_j\|$, and $N(h)$ is the number of observation pairs separated by that lag distance. For the irregularly spaced data in this study, we defined $N(h)$ using a bin width $\delta = \max(h) \,/\, \kappa\, n$ such that:

$$N(h) = \# \left\{ (i,j) : h - \frac{\delta}{2} \leq h_{ij} < h + \frac{\delta}{2} \right\} \tag{2}$$

where $\max(h)$ is the maximum separation distance within the Kriging domain, n is the number of lag intervals, and κ is the lag cutoff constant that limits the maximum lag distance h used to construct the experimental variogram $\hat{\gamma}(h)$ (conventionally $\kappa = 3$). We applied Kriging locally by restricting each estimate to the m nearest observations that fell within a predefined maximum neighborhood size $d_{\max}$, allowing the effective interpolation radius to adapt to local data density.

Rather than choosing variogram parameters by eye, a common but subjective practice, we simultaneously optimized all parameters using constrained nonlinear optimization (after Li et al., 2018). We tuned the parameters $\{n, \kappa, m, d_{\max}\}$ to minimize a weighted cost function combining variogram-fitting error and cross-validation interpolation error. Global optimization used the MLSL algorithm with Nelder-Mead local refinement. A subsequent fine-tuning step searched a restricted parameter neighborhood, and we implemented both stages in the R package `nloptr`. Full optimization details and results for all transect sets are presented in the Supplementary Information (Text S1; Figures S5–S38; Table S9).

2.3 Cross-Correlation Function Analysis

CCF analysis measured how consistently surface heat flow patterns varied along strike within each transect set. First, we projected surface heat flow data from IHFC 2024 observations, Similarity interpolation, and Krige interpolation onto each transect. To extract continuous trench-perpendicular trends, we smoothed these projected profiles using generalized additive models (GAMs, Wood, 2017) via the `mgcv` package in R. We modeled smooth profile curves as a function of along-transect distance using thin plate regression splines, with the smoothing parameter optimized via Restricted Maximum Likelihood (REML). We applied the default basis dimension constraint, allowing the algorithm to dynamically penalize overfitting based on sample size. We restricted trend fitting to transects containing a minimum of 10 projected data points. Finally, we resampled the smoothed mean profiles to 500 evenly spaced points along each transect, running from the subducting plate to approximately 500 km into the upper plate (e.g., see Figure 2: bottom row).

For each pair of adjacent transects, we quantified the similarity between surface heat flow profiles using the `stats::ccf()` function in R. The analysis allowed small horizontal shifts between profiles (up to ± 10% of the total transect length) to account for minor offsets in the positions of surface heat flow anomalies. We retained the maximum correlation value as a measure of profile similarity. Correlation coefficients near 1 indicate that adjacent transects have very similar surface heat flow patterns, whereas values near 0 or negative indicate weakly related or contrasting patterns.

We performed CCF analysis for all three surface heat flow datasets: IHFC 2024 observations (where data coverage was sufficient), Similarity interpolation, and Krige interpolation. We computed additional comparisons between methods (Obs–Sim, Obs–Krg, and Sim–Krg) for individual transects to evaluate agreement between interpolation approaches. In total, we analyzed 201 adjacent transect pairs across the four regional groups (39 in NPA, 56 in SAM, 78 in SEA, and 28 in SWP).

Following the pair-wise calculations, we classified transect sets into four spatial regimes based on data density and profile correlation thresholds. Sets with a mean density of fewer than 25 observations per transect were classified as ***ambiguous***. For sets meeting the 25-observation threshold, classifications were determined by summarizing cross-method agreement and along-strike continuity:

- ***Inferred continuous***: required high method agreement within individual transects (mean Obs–Krg CCF ≥ 0.70) and high along-strike agreement across all data types (minimum Obs CCF and Krg CCF and Sim CCF ≥ 0.50)
- ***Inferred discontinuous***: assigned when methods agreed within individual transects (mean Obs–Krg CCF ≥ 0.70) but adjacent profiles revealed abrupt spatial shifts in either the observations (minimum Obs CCF < 0) or Krige estimates (minimum Krg CCF < 0.30)
- ***Method disagreement***: assigned when estimates conflicted, either through poor agreement between observations and Krige estimates across individual transects (mean Obs–Krg CCF < 0.30) or when observations captured a structural boundary (minimum Obs CCF < 0) that the Similarity method smoothed over (maximum Sim CCF > 0.50)

Any remaining sets that did not fit these definitive criteria were designated as ***ambiguous***.

2.4 Interpolation Accuracy

We assessed interpolation accuracy by pairing estimated surface heat flow values from the $0.5^\circ \times 0.5^\circ$ interpolation grid directly with the IHFC 2024 observations (after Lucazeau, 2019). To minimize spatial mismatch, we established point pairs using a bidirectional nearest-neighbor approach: the nearest observation to each grid node, and the nearest grid node to each observation, were cross-referenced. We excluded pairs separated by a Euclidean distance exceeding a strict 10 km threshold to ensure localized validity.

The final evaluation dataset mapped each qualified interpolated node to its nearest valid observation. We summarized estimation errors across the SubMap transect sets using standard descriptive metrics, including the Root Mean Square Error (RMSE), mean error, and interquartile range (IQR).

We applied the same evaluation procedure to both Similarity and Krige interpolations to ensure consistent comparison between methods. We did not adopt cross-validation approaches commonly used for Kriging because the Similarity method uses the full global dataset and is less sensitive to removal of individual observations (Goutorbe et al., 2011). Direct comparison with observations therefore provided the most consistent basis for evaluating both interpolation methods.

3 Results

3.1 Along-Strike CCF Analysis Comparing Adjacent Transect Pairs

CCF analysis across 201 adjacent transect pairs reveals substantial variation in along-strike surface heat flow continuity among subduction systems (Figure 3; Table S7). The classifications shown in Figure 3 summarize the dominant behavior of each transect set rather than requiring perfect agreement at every transect pair. A classification therefore reflects the prevailing surface heat flow pattern across a region, while local departures are examined separately in Section 3.2.

Four domains exhibit sustained along-strike continuity over hundreds of km. Japan (NPA_SET1), Cascadia (NPA_SET8), Sumatra (SEA_SET2), and Kermadec–Hikurangi (SWP_SET4) all show moderate-to-high observation density with strong agreement among observations, Kriging, and Similarity estimates (Figure 3: green shaded areas; Table S7). In these regions, neighboring transects consistently reproduce similar surface heat flow patterns despite spanning large distances along strike. These domains therefore provide the clearest evidence that spatially continuous thermal structure can persist across substantial portions of some subduction segments.

Conversely, three similarly well-sampled domains exhibit abrupt changes in apparent thermal structure between neighboring transects. Kurils–Kamchatka (NPA_SET3), Central America–Panama (SAM_SET2), and Ryukyus–Sagami (SEA_SET9) all contain adjacent transects with weak or negative correlations that are supported by both observations and Krige profiles (Figure 3: red shaded areas; Table S7). In Central America–Panama, strongly correlated transects occur immediately adjacent to discontinuous pairs, demonstrating that surface heat flow continuity can change over distances comparable to a single transect spacing. Manila (SEA_SET7) displays a similar transition from continuity to discontinuity, although relatively sparse observations limit independent validation. Together, these domains indicate that along-strike thermal structure is not uniformly continuous and may change abruptly over distances of roughly 100–150 km.

Two domains exhibit mixed continuity patterns that cannot be attributed confidently to either continuity or discontinuity. The Caribbean (SAM_SET4) and Izu–Bonin (SEA_SET10) both contain alternating patterns of high and low adjacent-transect similarity, and observations and interpolation methods support different interpretations of the same region (Figure 3: yellow shaded areas; Table S7). These patterns suggest either unresolved thermal heterogeneity at scales smaller than the interpolation grid or methodological limitations in reconstructing local surface heat flow structure. Consequently, the location and significance of apparent thermal regime boundaries remain uncertain in both domains.

Most transect sets are ambiguous because observation density is currently too low to reliably evaluate continuity. Transects along large portions of Alaska, the Caribbean–Andean margin, SE Asia, and the SW Pacific have few or no observations (Tables S1 and S2), preventing independent validation of interpolated profiles. In these regions, Krige estimates are highly sensitive to sparse local data, while Similarity provides the only spatially complete surface heat flow estimate. As a result, continuity, discontinuity, and method disagreement cannot be inferred confidently across 71–75% of the global subduction zone system examined here (Table S7).

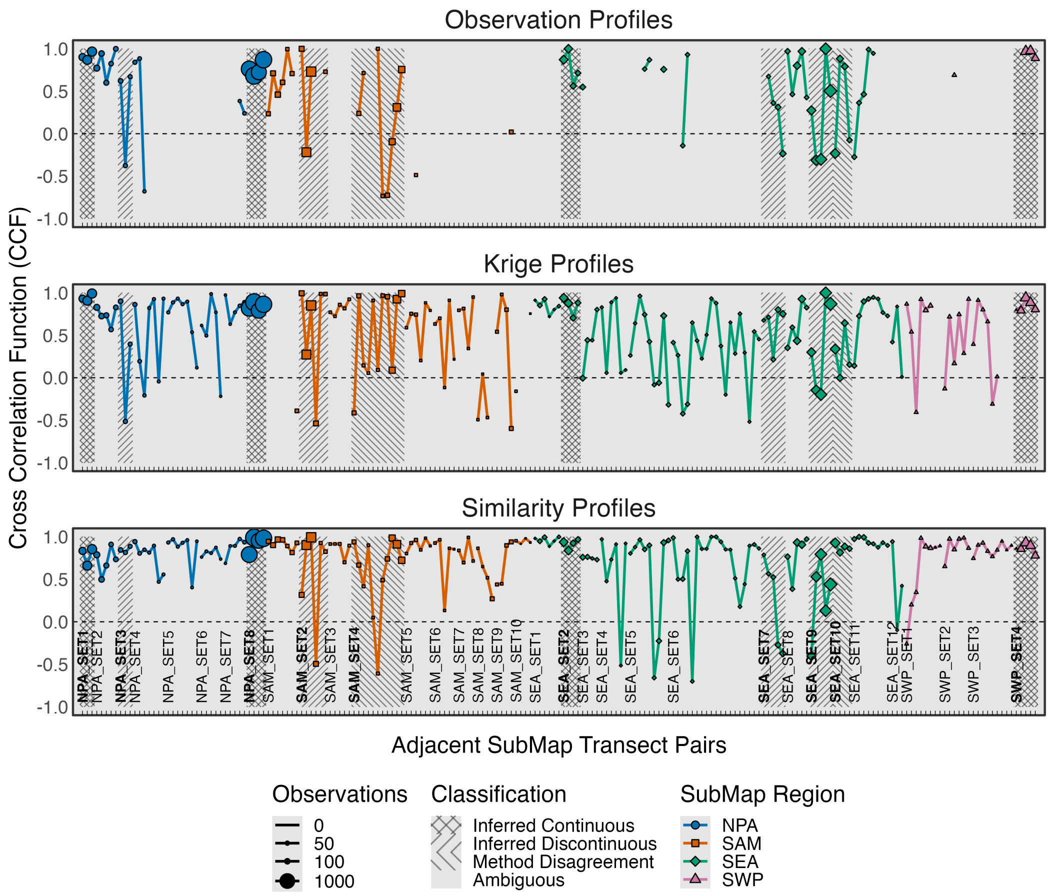

Figure 3: Cross-correlation function (CCF) analysis comparing observation, Krige, and Similarity profiles between adjacent SubMap transect pairs. Maximum CCF values (y-axis) are plotted across 201 adjacent transect pairs arranged by region and set (x-axis), with symbol size scaled by the number of total observations contributing to each pair. Along-strike continuity varies substantially within and across subduction systems. Background shading highlights distinct thermal regimes and method failure modes discussed in the text. Green areas for inferred continuity (NPA_SET1: Japan; NPA_SET8: Cascadia; SEA_SET2: Sumatra; SWP_SET4: Kermadec–Hikurangi); yellow areas for cross-method disagreement or sub-grid heterogeneity where methods contradict data or give false confidence (SAM_SET4: Caribbean, SEA_SET10: Izu–Bonin); red areas for inferred discontinuity along strike (NPA_SET3: Kurils–Kamchatka; SAM_SET2: Central America–Panama; SEA_SET7: Manila; SEA_SET9: Ryukyus–Sagami); and gray areas for sparse, unconstrained domains where data density is insufficient to infer continuity.

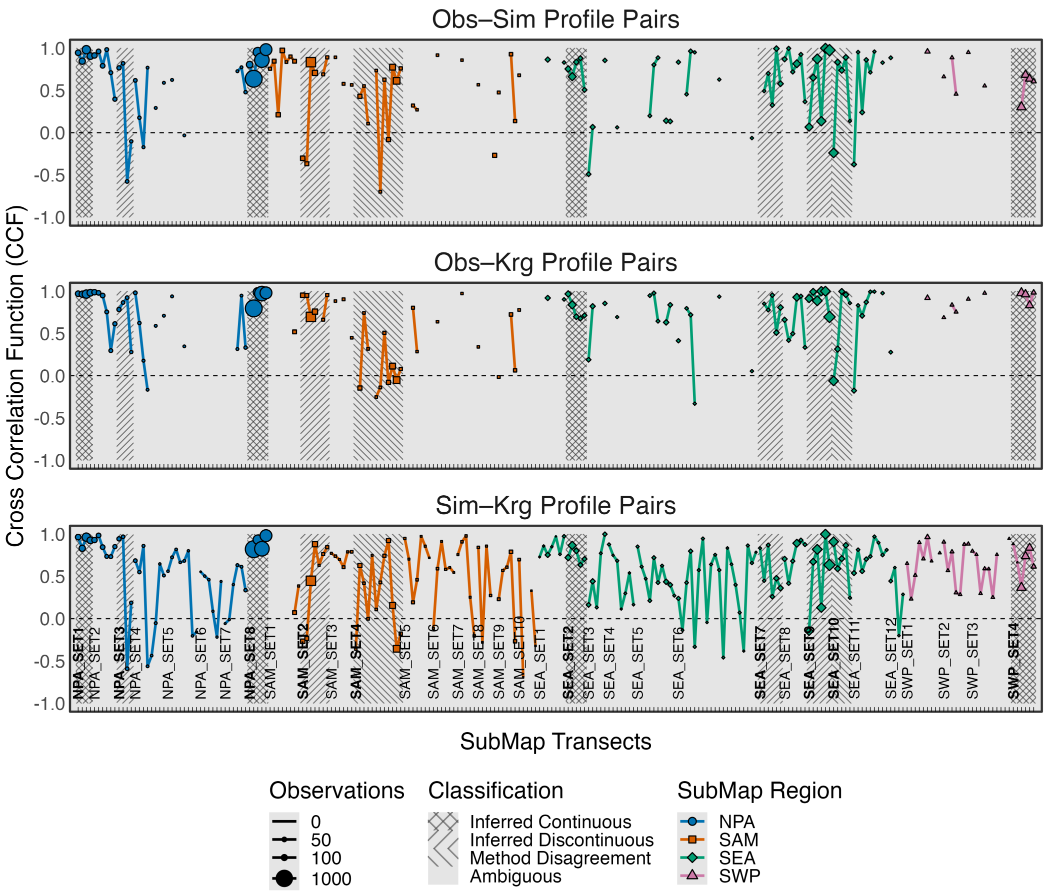

Figure 4: Cross-correlation function (CCF) analysis comparing Obs–Sim (Observed–Similarity), Obs–Krg (Observed–Krige), and Sim–Krg (Similarity–Krige) profile pairs for individual SubMap transects. Maximum CCF values (y-axis) are plotted across individual transects ordered by region and set (x-axis), with symbol size scaled by the number of observations contributing to each transect. High Sim–Krg CCF values identify transects where both interpolation methods recover matching thermal structures, while low or negative values expose systemic method disagreement. Background shading highlights distinct thermal regimes and method failure modes as described in Figure 3.

3.2 Transect-Level CCF Analysis Comparing Observations, Kriging, and Similarity

The classifications in Figure 3 describe the dominant continuity pattern within each transect set, but they do not imply perfect agreement among observations and interpolation methods at every transect. Examination of individual transect pairs (Figure 4) reveals three recurring sources of uncertainty: 1) disagreement between Similarity and locally constrained data, 2) Kriging failure despite dense observations, and 3) increasing interpolation uncertainty in sparsely sampled regions.

The most common disagreement occurs when Similarity produces continuity estimates inconsistent with both observations and Krige profiles. A particularly clear example is Ryukyus–Sagami (SEA_SET9), where observation density is high ($N_{\mathrm{obs}} = 83$–279) and Obs–Krg agreement is consistently strong (Obs–Krg CCF = 0.70–1.00; Figure 4; Table S8). Across five consecutive transect pairs, observations and Kriging agree on both the sign and magnitude of adjacent-transect correlations, while Similarity returns a different interpretation for four of the five pairs (Figure 5; Table S7). In several cases, Similarity indicates strong continuity where observations and Kriging indicate discontinuity, whereas in another pair it fails to reproduce a nearly perfect correlation identified by both locally-constrained datasets. Because observations and Kriging converged across multiple neighboring transects, we attribute the disagreement to Similarity's failure to resolve the apparent local thermal structure. Ryukyus–Sagami therefore illustrates a failure mode in which geologically derived Similarity estimates diverge from observationally constrained surface heat flow patterns.

The same failure mode appears in several independent subduction systems. In Central America–Panama (SAM_SET2), Similarity underestimates continuity in one transect pair and overestimates it in the immediately adjacent pair, despite strong support from observations and Kriging for opposite outcomes. Similar discrepancies occur in Japan (NPA_SET1), where Kriging closely tracks observations while Similarity consistently returns weaker correlations, and in Kurils–Kamchatka (NPA_SET3), where Similarity indicates strong continuity even though both observations and Kriging indicate discontinuity (Figures 3 and 4; Tables S7 and S8). The recurrence of these disagreements across multiple regions suggests that Similarity sometimes fails to reproduce along-strike thermal structure even where local observations are sufficient to constrain it.

A second failure mode occurs when dense observations do not produce reliable Krige profiles. In most well-sampled domains, Kriging closely reproduces observed thermal structure and Obs–Krg CCF values are high. However, two domains depart from this pattern. In the Caribbean (SAM_SET4), the most densely sampled transects in the region (SAM31, SAM32; $N_{\mathrm{obs}} = 129, 175$) exhibit Obs–Krg CCF values near zero, indicating that the interpolated Krige profiles bear little resemblance to the underlying observations (Figure 4; Table S8). A similar result occurs in Izu–Bonin (SEA_SET10), where one densely sampled transect (SEA73; $N_{\mathrm{obs}} = 190$) also produces near-zero Obs–Krg agreement. For the adjacent pair SEA73–SEA74, both interpolation methods indicate positive continuity, yet observations indicate a weak discontinuity (Table S7). These examples demonstrate that high observation density alone does not guarantee accurate Krige-based profiles or continuity estimates.

Interpolation uncertainty increases markedly where observations are sparse. Point-by-point Sim–Krg prediction differences are centered near zero in most domains, indicating little systematic bias between methods, but the spread of those differences varies substantially among regions (Figure S1; Table S3). The largest discrepancies occur in Andean domains SAM_SET9 and SAM_SET10 and in New Hebrides (SWP_SET2), where transects contain few observations and Krige variance is correspondingly large. Similarity uncertainty remains lower than Krige uncertainty across most domains (Figure S2; Table S4), whereas in well-sampled regions Kriging generally produces narrower residual distributions and stronger agreement with observations (Figures S3 and S4; Tables S5 and S6). These patterns indicate that uncertainty in surface heat flow estimates increases systematically as observational constraints weaken.

3.3 Optimized Variogram Parameters Across the 34 Domains

Optimized Krige parameters vary widely among transect sets and show no evidence for a single characteristic spatial scale of surface heat flow continuity in subduction zones. All 34 domains converge to an optimal numerical solution (Figures S5–S38; Table S9). While this consistent convergence demonstrates the mathematical stability of the optimization algorithm across a full range of data densities, the resulting variograms in observation-poor domains (per-transect mean N_{obs} density ≤ 25) are highly unconstrained and subject to severe parameter volatility, reflecting localized sample noise rather than physically meaningful spatial structures. Nevertheless, the optimization reliably handles the

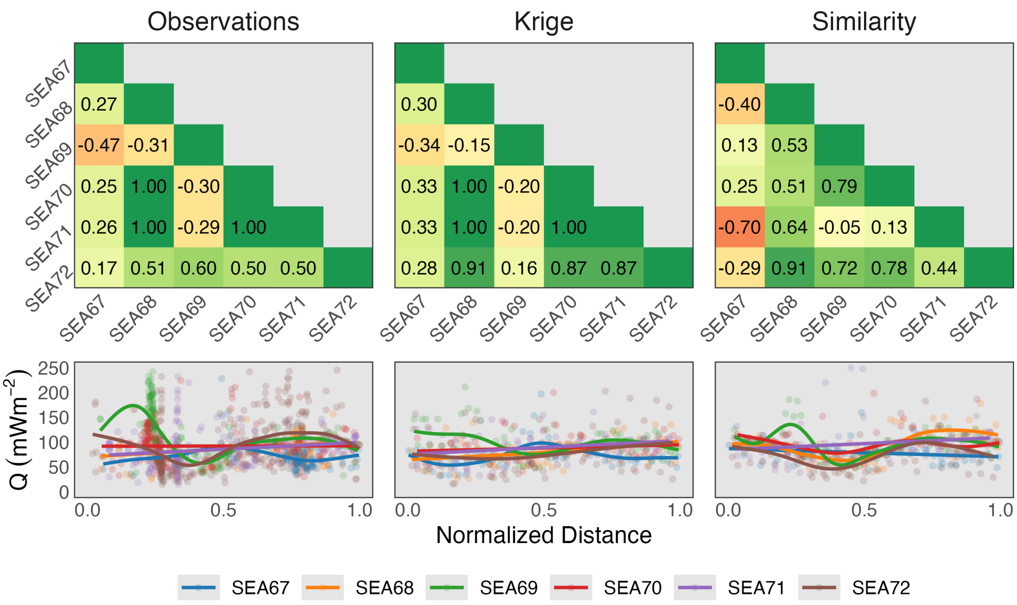

Figure 5: Cross-correlation function (CCF) analysis (top row) and surface heat flow profiles (bottom row) for neighboring Ryukyus–Sagami transects (SEA67–SEA72). Obs and Krg CCF values (left, center) agree in sign and magnitude for all five consecutive transect pairs, even when both are negative, while Similarity CCF (right) returns discordant values in four of the five consecutive pairs (compare first subdiagonals in CCF analysis matrices). Disagreement between observations and Krige estimates versus Similarity reveals Similarity's failure to resolve thermal structure continuity compared to local constraints.

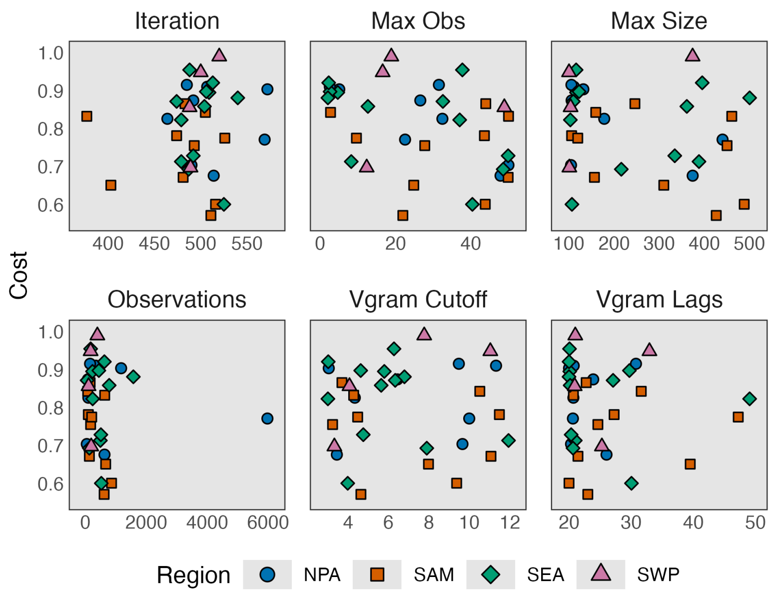

Figure 6: Summary of optimization iterations, observation counts, and optimized Krige parameters for all transect sets (colored by region). Parameter values (x-axes) span broad ranges and include exponential, spherical, and Gaussian variogram models (Table S9). No systematic relationship exists between individual parameter values and the optimized objective function (y-axis), indicating that characteristic spatial correlation scales vary among subduction zone domains. Optimized Krige parameter units are: maximum neighborhood observations (dimensionless), maximum neighborhood size (km), variogram lag cutoff constant (dimensionless), and number of variogram lag intervals (dimensionless).

full spectrum of sampling conditions, from dense networks like Cascadia (up to 1,077 observations per transect) to sparse regions with no immediate transect observations. Exponential, spherical, and Gaussian variogram models are each selected across multiple domains; maximum neighborhood sizes d_{max} range from approximately 100 to 500 km; lag cutoff constants κ range from 3–12; and the number of lag intervals n ranges from 20 to nearly 50 (Figure 6; Table S9). Domains with comparable observation counts frequently converge on different variogram models and spatial scales, and similarly low-cost solutions arise from a wide range of parameter configurations. No systematic relationship exists between individual parameter values and the total optimization cost.

4 Discussion

4.1 Along-Strike Surface Heat Flow Continuity is Ambiguous or Geographically Limited

Numerical thermomechanical models of subduction zones have historically used individual trench-perpendicular surface heat flow profiles to calibrate their initial thermal structures and boundary conditions. Wada & Wang (2009) synthesized models across 17 subduction zones and concluded that a common slab-mantle coupling depth near 70–80 km reconciles volcanic arc positions and forearc surface heat flow globally. Hyndman et al. (2005) and Currie & Hyndman (2006) identified broadly uniform backarc surface heat flow of approximately 75 ± 15 mW m^{-2} across ten circum-Pacific systems, attributing this uniformity to vigorous slab-driven corner flow and a thin, thermally eroded backarc lithosphere. These conclusions implicitly assume that the selected surface heat flow profiles all reliably represent thermal structure and are spatially continuous for hundreds of km along strike. Our CCF analysis directly evaluates that assumption across 201 adjacent transect pairs.

Our results support the conventional model of subduction zone thermal structure only for a minority of the segments where those models were originally calibrated. Japan (NPA_SET1), Cascadia (NPA_SET8), Sumatra (SEA_SET2), and Kermadec–Hikurangi (SWP_SET4) all show sustained along-strike continuity (Obs CCF = 0.56–1.00), indicating that only 24–40% of trench-perpendicular profiles underlying the Wada & Wang (2009), Hyndman et al. (2005), and Currie & Hyndman (2006) syntheses are regionally representative. The tacit assumption that a single profile captures the broader thermal state of a subduction segment is therefore geographically limited: elsewhere it either does not hold or cannot be confirmed. Three well-sampled domains (Kurils–Kamchatka: NPA_SET3, Central America–Panama: SAM_SET2, and Ryukyus–Sagami: SEA_SET9) show abrupt along-strike changes at 100–150 km scales, which render single-profile characterizations misleading. Two additional domains (Caribbean: SAM_SET4; Izu–Bonin: SEA_SET10) show abrupt CCF transitions whose locations cannot be reliably determined due to method disagreement, and 71–75% of the global dataset lacks sufficient observations for independent assessment (Figures 3 and 4; Tables S7 and S8). The conventional model of uniform coupling depths with thin backarc lithospheres is therefore better viewed as a locally validated description of a small number of well-characterized segments, rather than as a common feature of active subduction zones worldwide. A nearly equal number of well-characterized segments show the opposite pattern.

Several deep-seated tectonic mechanisms could explain along-strike deviations from the conventional model of subduction zone thermal structure. Numerical geodynamic modeling studies have shown that variations in slab geometry can concentrate hot sub-arc mantle at specific segments via dynamic pressure gradients (Araya Vargas et al., 2021), and that cessation of backarc spreading can trigger lithospheric instabilities and along-strike temperature heterogeneity over million-year timescales (Lee & Wada, 2021). Kerswell et al. (2021) showed that backarc lithospheric thickness and serpentinite stability jointly control the maximum slab-mantle coupling depth, offering another physical mechanism for the along-strike variability resolved by our CCF analysis. Peacock & Wang (2021) proposed a related mineralogical control involving the pressure-dependent breakdown of talc near 2.0–2.5 GPa (70–80 km depth). Subduction interface friction may also matter: Kohn et al. (2018) showed that forearc surface heat flow requires shear heating with friction coefficients of 0.03–0.1, so along-strike variation in friction [surface roughness; Gao & Wang (2014)] or convergence rate could similarly produce the surface heat flow variability documented here. Where slab geometry, metamorphic dehydration depth, or interface friction vary along strike, adjacent profiles can record different thermal regimes within a single subduction segment.

Yet each of these deeper mechanisms could also, in principle, be mimicked by near-surface hydrothermal perturbations. Hydrothermal circulation in young oceanic crust can reduce local heat flow to 10–40% of lithospheric predictions (Hutnak et al., 2008), and channelized basement fluid flow can create focused surface heat flow anomalies over lateral scales of tens of km (Fisher & Becker, 2000; Fisher & Harris, 2010). Without collocated seismic velocity models, fluid pressure data, or sediment thickness constraints, surface heat flow cannot discriminate between a slab-mantle coupling depth transition and a locally perturbed hydrothermal signal. This ambiguity inherently limits applications of surface heat flow as a geophysical observable in tectonically active settings.

4.2 Interpolation Failure Modes Depend on Observation Density and Proxy Resolution

The three uncertainty patterns identified in our CCF analysis (Sim–Obs disagreement, Kriging failure in dense settings, and sparsity-driven uncertainty; see Section 3.2) reflect a systematic dependence of method reliability on observation density. Where per-transect observations are sparse, Krige profiles overfit individual measurements, leaving Similarity as the only reliable estimate. Where observations are dense, Kriging closely reproduces local data while Similarity can mischaracterize along-strike correlations in specific settings. Understanding when and where each method can fail is essential for interpreting the surface heat flow record across the range of data densities represented in the IHFC 2024 dataset (Table S2).

In sparse domains, per-transect observation counts of roughly 25 or fewer produce Krige profiles sensitive to individual measurements rather than regional surface heat flow structure. The Aleutian domain (NPA_SET5) illustrates this clearly: 4–12 observations per transect yield highly variable Krg CCF across seven consecutive transect pairs, while Similarity produces high correlations for six of the seven (Figure 3; Table S7). In such domains, Similarity provides the only available geological interpolation. Although Lucazeau (2019) validated Similarity estimates at the global scale, systematic validation at the 100–150 km transect-pair scale has not been previously demonstrated. Whether Similarity accurately captures along-strike thermal structure within a single subduction segment therefore remains ambiguous. Where local observations are absent, any failure of the Similarity method at the transect scale would go undetected.

In well-sampled domains, the failure-mode pattern reverses and exposes a specific limitation of Similarity. In Ryukyus–Sagami (SEA_SET9), observations and Kriging converge on consistent along-strike CCF values across five consecutive transect pairs, yet Similarity returns discordant values for four of the five, including misidentifying discontinuous segments as continuous and substantially underestimating a near-perfect correlation inferred by both observations and Krige estimates (Figure 5; Table S7). Central America–Panama (SAM_SET2) demonstrates that Similarity failures can occur in both directions within immediately adjacent pairs. It underestimates strong continuity inferred by observations and Kriging in one pair, and in another pair returns strong continuity where observations indicate a structural discontinuity (Figures 3 and 4; Tables S7 and

S8). Similar discrepancies in Japan and Kurils–Kamchatka confirm that these failures are not confined to a single region.

These systematic failures reflect how Similarity constructs its estimates. The method identifies global analogs in multivariate proxy space and applies weighted averaging to their measured values (Goutorbe et al., 2011). Adjacent transects within a single subduction segment may share nearly identical proxy configurations while having distinct along-strike thermal structure. Such differences can arise where proxy variables (plate age, topography, seismic tomography, and others) fail to capture important properties or processes, for example fine-scale slab roughness (interface friction), localized fluid pathways, or short-wavelength variations in deeper mantle wedge flow. The $0.5° \times 0.5°$ grid spacing of the Lucazeau (2019) implementation (approximately 55 km at equatorial latitudes) may further prevent the proxy field from resolving along-strike thermal gradients at the transect scale. Prior validation by Shapiro & Ritzwoller (2004) demonstrated the viability of proxy-guided interpolation at regional scales but did not consider the resolution-dependent limitation identified here.

The Kriging failures in the Caribbean (SAM_SET4) and Izu–Bonin (SEA_SET10) illustrate a complementary limitation. At those transects, thermal structure varies at spatial scales shorter than any variogram model's correlation length can resolve. Long variogram ranges over-smooth the profiles, whereas short ranges are incompatible with sparse data coverage elsewhere in the domain. No single variogram configuration reconciles these competing constraints. The convergence of different Krige parameter configurations across domains therefore reflects regional differences in the scale of surface heat flow organization rather than degeneracy in the optimization algorithm (see Text S1 for full algorithm details). Across the 34 subduction domains examined here, the absence of a preferred variogram model or characteristic length scale (Figure 6; Table S9) further suggests that surface heat flow is heterogeneous at all resolvable scales and effectively unconstrained below a critical observation-density threshold. A large majority of transects ($> 70\%$) currently do not meet this threshold, which is why heat flow systematics are so ambiguous across much of the globe (Figures 3 and 4; Tables S7 and S8).

4.3 Method Agreement Does Not Guarantee Accuracy

The failure modes described in the previous section share an important diagnostic signal: method disagreement indicates a problem, but method agreement does not guarantee accuracy. Two additional patterns in our analysis make this clear and carry direct implications for how assessments of subduction zone thermal structure based on surface heat flow should be used.

The most consequential of these is false consensus between independent interpolation methods. In Izu–Bonin (SEA_SET10), Similarity and Kriging both return positive along-strike continuity (CCF = 0.34–0.93) for a transect pair (SEA73–SEA74) where 190 local observations indicate weak discontinuity (CCF = -0.23; Figure 3; Table S7). The Krige profile at the more densely sampled of the two transects (SEA73) fails to track observed surface heat flow structure despite the high observation count (Figure 4; Table S8), indicating that local thermal structure varies at scales shorter than the variogram can resolve. High Sim–Krg CCF agreement masks this failure rather than flagging it. Neither Similarity's global proxy resolution nor Kriging's regionally optimized variogram length scale captures finer-scale heterogeneity, and examining the observations directly is the only way to detect it. Analyses of this segment that rely solely on interpolated estimates would find consistent method agreement supporting a continuity that the measurements directly contradict.

The Caribbean (SAM_SET4) documents the distinct complication of apparent boundary displacement. Observations locate an abrupt thermal regime boundary at one transect pair (SAM28–SAM29; Obs CCF = -0.73), while Similarity places it at the immediately adjacent pair (SAM27–SAM28; Sim CCF = -0.61)—an apparent displacement of roughly 100–150 km (Figure 3). In thermal modeling contexts where surface heat flow profiles calibrate slab-mantle coupling depths, a positional error of this magnitude is geodynamically significant. It misrepresents the lateral extent of the coupled interface and the along-strike position of key frictional or metamorphic phase transitions (e.g., Gao & Wang, 2014; Kohn et al., 2018; Peacock & Wang, 2021). The Caribbean's multiple subduction geometries, polarity reversals, and active backarc basins (Lallemand & Heuret, 2017) generate short-wavelength thermal variability that proxy-based methods cannot resolve at the transect scale, and dense local measurements are required to accurately locate structural boundaries.

Together, these cases illustrate how method agreement narrows the range of possible failure modes but does not guarantee accuracy. In well-constrained domains, agreement between locally-validated Krige estimates and observationally-consistent Similarity estimates carries meaningful confidence. Where neither method can be independently checked against dense local observations, however, agreement may reflect coincident or correlated failures rather than shared accuracy. The diagnostic framework developed here, which evaluates per-transect observation counts alongside CCF analysis, provides a rubric for distinguishing these situations systematically. Interpolation comparison without observation-density context is insufficient for interpretation.

4.4 Sparse Observations Limit Geodynamic Interpretations

The transect-level observation counts reported here quantify a practical threshold for evaluating interpolation reliability. Nearly three quarters of all transects (168/235; 71%) lack sufficient observations for independent validation ($N_{\mathrm{obs}} \leq 25$), meaning that the surface heat flow record cannot currently support or refute geodynamic inferences about along-strike continuity for most of the global subduction system. This gap reflects a persistent sampling bias toward accessible continental margins, onshore monitoring networks, and geothermal exploration targets, leaving subduction zone forearcs and backarcs systematically undersampled at scales required for geodynamic interpretation.

Database growth has not yet resolved this issue. The IHFC 2024 compilation represents a substantial addition to prior datasets, adding roughly 21,000 measurements since the Lucazeau (2019) dataset, yet the new observations are concentrated in already well-sampled regions. The along-strike gaps identified here for the Aleutians, most of S America, and much of SE Asia and the SW Pacific persist, and the $>$ 1000 km nearest-neighbor distances documented by Pollack et al. (1993) across the S Pacific remain relevant for many of the transects examined here.

Where observations do exist, systematic bias in specific tectonic settings compounds the sparsity problem. Hydrothermal advection accounts for a substantial fraction of total oceanic heat loss in young oceanic crust, and conductive measurements in these settings systematically underestimate the true lithospheric heat flux (Stein & Stein, 1994). More than half of the plates entering the trenches analyzed here (127/235 transects; 54%) carry oceanic crust younger than the 65 $\pm$ 10 Ma sealing age beyond which hydrothermal circulation no longer measurably affects conductive surface heat flow (Stein & Stein, 1994). Basement fluid discharge along mid-plate seamount chains can locally reduce regional conductive measurements to a small fraction of lithospheric predictions (Hutnak et al., 2008). In forearcs and incoming-plate regions, the available observations are therefore doubly limited: there are too few measurements to constrain spatial patterns, and the measurements that do exist are potentially biased by hydrothermal extraction such that they underestimate the thermal signal that interpolation methods attempt to reconstruct.

These limitations directly constrain geodynamic inference from the surface heat flow record. The conventional model of subduction zone thermal structure proposed by Wada & Wang (2009) and Currie & Hyndman (2006) was built on a small number of well-sampled sites (Japan, Kamchatka, Cascadia, and Kermadec). Our CCF analysis shows that three of these sites do exhibit high along-strike continuity, confirming previous single-profile characterizations as regionally representative. However, our analysis also identifies one exception: Kurils–Kamchatka (NPA_SET3) shows along-strike discontinuity, indicating those profiles are less representative than previously assumed by Wada & Wang (2009). For the broader global system, whether the conventional model extends beyond the four inferred continuous segments in our CCF analysis (Figure 3: green shaded areas; Table S7) cannot be resolved from the existing surface heat flow record (Figure 3; Table S7). Discontinuities observed in several segments urge caution.

4.5 Targeted Surveys and Hybrid Approaches

The data limitations and failure-mode diagnostics discussed above together define specific, actionable survey priorities rather than a generic call for more data. Two target categories emerge, each requiring a different survey strategy and offering a different scientific return.

The most widespread case involves sparse domains where Similarity provides the only interpolation estimate and no independent validation is possible. These regions (Figure 3: gray areas; Table S7) currently have no basis for evaluating whether Similarity CCF values reflect genuine thermal continuity or simply the spatial smoothness of the proxy field. Even modest targeted surveys in these regions would provide the first independent test of Similarity estimates. The per-transect density thresholds identified here define the minimum observation density needed to cross from ambiguous to interpretable (roughly $N_{\mathrm{obs}} > 25$).

A more demanding strategy is required where fine-scale thermal heterogeneity exists at scales below global interpolation resolution. The Caribbean and Izu–Bonin segments demonstrate that short-wavelength thermal variability can produce abrupt along-strike CCF transitions within consecutive transect pairs (Figures 3 and 4; Tables S7 and S8), and that neither interpolation method reliably locates structural boundaries under these conditions. Resolving finer-scale structure requires closely spaced measurements rather than scattered additions. The tectonic complexity of these systems likely also demands collocated seismic imaging and geochemical fluid sampling to separate deep thermal signals from shallow hydrothermal overprinting.

Beyond survey design, the complementary failure modes of Similarity and Kriging motivate hybrid interpolation strategies. The most direct combination treats Similarity as a geologically informed prior for the regional surface heat flow field and applies Kriging to update that prior with local measurements, conceptually equivalent to Similarity as a trend model in universal Kriging or as a secondary covariate in a co-Kriging framework. The failure-mode diagnostics define where each approach should dominate: Similarity should govern where local observations are too sparse for Kriging to produce stable estimates, while Krige estimates should take precedence where per-transect counts and Obs–Krg CCF indicate strong local constraint. A framework incorporating the Lucazeau (2019) proxy variables as spatially varying covariates within a Kriging system would operationalize this without requiring separate comparison steps.

Kriging and Similarity also perform optimally at different spatial scales and can be deployed accordingly. The Ryukyus–Sagami analysis showed that Kriging outperforms Similarity at the transect-pair scale (Figure 5; Table S7). A scale-decomposed approach could assign Similarity to characterize the large-scale background field while Kriging recovers short-scale departures where local data constrain the variogram. Within the CCF framework, this decomposition would partition along-strike continuity patterns between the geological-analog signal in the Similarity field and the locally constrained spatial structure in Krige residuals, making failure-mode attribution explicit rather than inferred.

5 Conclusions

We applied ordinary Kriging to IHFC 2024 data across 235 transects in 34 transect-set domains and quantified along-strike continuity for 201 adjacent transect pairs using cross-correlation function (CCF) analysis. We compared Krige estimates with the global Similarity interpolation from Lucazeau (2019) to evaluate where both methods provide independent, locally constrained assessments of surface heat flow structure. The combined analysis yields four principal conclusions.

Along-strike surface heat flow continuity is ambiguous or geographically limited. Four domains (Japan, Cascadia, Sumatra, and Kermadec–Hikurangi) show sustained continuity consistent with observations and both interpolation methods, validating common models of subduction zone thermal structure for those segments (Currie & Hyndman, 2006; Wada & Wang, 2009). Three well-sampled domains (Kurils–Kamchatka, Central America–Panama, and Ryukyus–Sagami) show abrupt structural changes at 100–150 km along-strike scales independent of any interpolation method, contradicting assumptions of common subduction zone thermal models. Two other well-sampled domains (the Caribbean and Izu–Bonin) exhibit systematic method disagreement with abrupt along-strike CCF transitions whose locations could not be confirmed. A large majority of the global dataset lacks sufficient observations for any independent assessment.

Kriging and Similarity exhibit complementary failure modes. In sparse domains, Kriging is unreliable and Similarity provides the only substantive estimate. In well-sampled domains, Kriging tracks observations closely while Similarity mischaracterizes along-strike correlations in specific settings. Method consensus does not guarantee accuracy, but the diagnostic framework developed here provides a basis for distinguishing reliable from misleading interpretations.

No universal variogram model or characteristic spatial scale describes subduction zone surface heat flow continuity globally. Kriging optimization converges for all 34 domains, yet maximum search radii span approximately 100 to 500 km with no systematic relationship to optimization cost. Where Krige profiles fail to track local observations, thermal heterogeneity exists at spatial scales shorter than any variogram model can resolve. Convergence to different configurations across domains reflects real regional differences in the scale of surface heat flow organization rather than algorithmic degeneracy.

Sparse observations are the dominant limitation on geodynamic interpretation of subduction across most of the planet. Nearly three quarters of all transects (168/235; 71%) have 25 or fewer observations, precluding Krige-based validation. Additional observations in the IHFC 2024 database relative to prior compilations are concentrated in already well-sampled regions. In forearc regions, sparse observations may additionally be biased toward underestimating the true lithospheric heat flux due to hydrothermal extraction (Hutnak et al., 2008; Stein & Stein, 1994). Whether uniform slab-mantle coupling depths and thin backarc lithospheres are common features of subduction zones remains an open question. Answering it will require strategically targeted surface heat flow surveys guided by the failure-mode diagnostics and observation-density thresholds defined here.

Acknowledgments

We express our sincere gratitude to the communities, institutions, and research teams who develop and maintain the open-source datasets used in this study. Specifically, we thank the creators of the SubMap tool (Géosciences Montpellier); the University of Texas Institute for Geophysics for the plate boundary compilation; the NOAA National Centers for Environmental Information for the ETOPO 2022 global relief model; the Global Heat Flow Data Assessment Group for maintaining the Global Heat Flow Database; and F. Lucazeau for providing the Similarity interpolations. This work was supported by the National Science Foundation grants OIA 1545903 to M. Kohn, S. Penniston-Dorland, and M. Feineman and EAR 2118114 to M. Kohn.

Data Availability

All data, code, and relevant information for reproducing this work are archived on the OSF (Kerswell, 2026a) and Zenodo (Kerswell, 2026b) repositories. All code within these repositories is MIT Licensed and free for use and distribution (see license details). All spatial datasets used in this study are open-source: transects from the SubMap tool version 7.1 (Lallemand et al., 2026); present-day plate boundaries from Coffin et al. (2018); global relief model from NOAA National Centers for Environmental Information (2022); surface heat flow observations from the Global Heat Flow Data Assessment Group et al. (2024); and Similarity interpolations from Lucazeau (2019).

Conflict of Interest

The authors declare there are no conflicts of interest for this manuscript.

References

Abers, G., van Keken, P., & Hacker, B. (2017). The cold and relatively dry nature of mantle forearcs in subduction zones. *Nature Geoscience*, *10*(5), 333–337.

Araya Vargas, J., Sanhueza, J., & Yáñez, G. (2021). The role of temperature in the along-margin distribution of volcanism and seismicity in subduction zones: Insights from 3-d thermomechanical modeling of the central andean margin. *Tectonics*, *40*(11), e2021TC006879.

Batir, J., Blackwell, D., & Richards, M. (2016). Heat flow and temperature-depth curves throughout alaska: Finding regions for future geothermal exploration. *Journal of Geophysics and Engineering*, *13*(3), 366–378.

Chapman, D., & Pollack, H. (1975). Global heat flow: A new look. *Earth and Planetary Science Letters*, *28*(1), 23–32.

Coffin, M., Gahagan, L., & Lawver, L. (2018, July). Present-day plate boundary digital data compilation [Dataset]. University of Texas Institute for Geophysics. `https://doi.org/10.15781/T2ZP3WH84`

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Currie, C., & Hyndman, R. (2006). The thermal structure of subduction zone back arcs. *Journal of Geophysical Research: Solid Earth*, *111*(B8), 1–22.

Currie, C., Wang, K., Hyndman, R., & He, J. (2004). The thermal effects of steady-state slab-driven mantle flow above a subducting plate: The cascadia subduction zone and backarc. *Earth and Planetary Science Letters*, *223*(1-2), 35–48.

Davies, J. (2013). Global map of solid earth surface heat flow. *Geochemistry, Geophysics, Geosystems*, *14*(10), 4608–4622.

Fisher, A., & Becker, K. (2000). Channelized fluid flow in oceanic crust reconciles heat-flow and permeability data. *Nature*, *403*(6765), 71–74.

Fisher, A., & Harris, R. (2010). Using seafloor heat flow as a tracer to map subseafloor fluid flow in the ocean crust. *Geofluids*, *10*(1-2), 142–160.

Fourier, J. (1827). Mémoire sur les températures du globe terrestre et des espaces planétaires. *Mémoires de l'Académie Royale Des Sciences de l'Institut de France*, *7*, 570–604.

Furlong, K., & Chapman, D. (2013). Heat flow, heat generation, and the thermal state of the lithosphere. *Annual Review of Earth and Planetary Sciences*, *41*(1), 385–410.

Furukawa, Y. (1993). Depth of the decoupling plate interface and thermal structure under arcs. *Journal of Geophysical Research: Solid Earth*, *98*(B11), 20005–20013.

Gao, X., & Wang, K. (2014). Strength of stick-slip and creeping subduction megathrusts from heat flow observations. *Science*, *345*(6200), 1038–1041.

Global Heat Flow Data Assessment Group et al. (2024, April). The Global Heat Flow Database: Release 2024 [Dataset] (Version 1). GFZ Data Services. `https://doi.org/10.5880/fidgeo.2024.014`

Goodchild, M. (2004). The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, *94*(2), 300–303.

Goutorbe, B., Poort, J., Lucazeau, F., & Raillard, S. (2011). Global heat flow trends resolved from multiple geological and geophysical proxies. *Geophysical Journal International*, *187*(3), 1405–1419.

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal*, *8*, 204–218. Retrieved from `https://journal.r-project.org/archive/2016/RJ-2016-014/index.html`

Hasterok, D. (2013). A heat flow based cooling model for tectonic plates. *Earth and Planetary Science Letters*, *361*, 34–43.

Hasterok, D., & Chapman, D. (2008). Global heat flow: A new database and a new approach. In *AGU fall meeting abstracts* (Vol. 2008, pp. T21c–1985).

Hasterok, D., Chapman, D., & Davis, E. (2011). Oceanic heat flow: Implications for global heat loss. *Earth and Planetary Science Letters*, *311*(3-4), 386–395.

Holt, A., & Condit, C. (2021). Slab temperature evolution over the lifetime of a subduction zone. *Geochemistry, Geophysics, Geosystems*, *22*(6), e2020GC009476.

Hutnak, M., Fisher, A., Harris, R., Stein, C., Wang, K., Spinelli, G., et al. (2008). Large heat and fluid fluxes driven through mid-plate outcrops on ocean crust. *Nature Geoscience*, *1*(9), 611–614.

Hyndman, R., Currie, C., & Mazzotti, S. (2005). Subduction zone backarcs, mobile belts, and orogenic heat. *GSA Today*, *15*(2), 4–10.

Jennings, S., Hasterok, D., & Lucazeau, F. (2021). ThermoGlobe: Extending the global heat flow database. *Journal TBD*.

Kelvin, W. (1863). On the secular cooling of the earth. *Transactions of the Royal Society of Edinburgh*, *23*, 157–170.

Kerswell, B. (2026a, August). buchanankerswell/kerswell_kohn_heatflow: Publication Release [Dataset] (Version 1.0.0). OSF. `https://doi.org/10.17605/OSF.IO/CA6ZU`

Kerswell, B. (2026b, August). buchanankerswell/kerswell_kohn_heatflow: Publication Release [Software] (Version 1.0.0). Zenodo. https://doi.org/10.5281/zenodo.21790947

Kerswell, B., Kohn, M., & Gerya, T. (2021). Backarc lithospheric thickness and serpentine stability control slab-mantle coupling depths in subduction zones. *Geochemistry, Geophysics, Geosystems*, *22*(6), e2020GC009304.

Kohn, M., Castro, A., Kerswell, B., Ranero, C., & Spear, F. (2018). Shear heating reconciles thermal models with the metamorphic rock record of subduction. *Proceedings of the National Academy of Sciences*, *115*(46), 11706–11711.

Krige, D. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(6), 119–139.

Lallemand, S., & Heuret, A. (2017). Subduction zones parameters. *Reference Module in Earth Systems and Environmental Sciences*.

Lallemand, S., Cerpa, N., Peyret, M., Grosbeau, F., Heuret, A., Arcay, D., & van Rijsingen, E. (2026, April). SubMap: A tool for mapping subduction zones [Software] (Version 7.1). SubMap. Retrieved from https://submap.fr

Lee, C., & Wada, I. (2021). Volcano clustering promoted by the cessation of back-arc spreading and ensuing nascent lithospheric drips. *Geophysical Research Letters*, *48*(9), e2020GL091433.

Lee, W., & Uyeda, S. (1965). Review of heat flow data. *Terrestrial Heat Flow*, *8*, 87–190.

Li, Z., Zhang, X., Clarke, K., Liu, G., & Zhu, R. (2018). An automatic variogram modeling method with high reliability fitness and estimates. *Computers & Geosciences*, *120*, 48–59.

Lucazeau, F. (2019). Analysis and mapping of an updated terrestrial heat flow data set. *Geochemistry, Geophysics, Geosystems*, *20*(8), 4001–4024.

Majorowicz, J., & Grasby, S. (2010). Heat flow, depth–temperature variations and stored thermal energy for enhanced geothermal systems in canada. *Journal of Geophysics and Engineering*, *7*(3), 232–241.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*(8), 1246–1266.

NOAA National Centers for Environmental Information. (2022, October). ETOPO 2022 15 Arc-Second Global Relief Model [Dataset]. National Centers for Environmental Information. https://doi.org/10.25921/fd45-gt74

Nyblade, A., & Pollack, H. (1993). A global analysis of heat flow from precambrian terrains: Implications for the thermal structure of archean and proterozoic lithosphere. *Journal of Geophysical Research: Solid Earth*, *98*(B7), 12207–12218.

Parsons, B., & Sclater, J. (1977). An analysis of the variation of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research*, *82*(5), 803–827.

Peacock, S., & Wang, K. (2021). On the stability of talc in subduction zones: A possible control on the maximum depth of decoupling between the subducting plate and mantle wedge. *Geophysical Research Letters*, *48*(17), e2021GL094889.

Pebesma, E. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*, 683–691.

Pebesma, E. (2018). Simple features for r: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446.

Pollack, H., & Chapman, D. (1977). On the regional variation of heat flow, geotherms, and lithospheric thickness. *Tectonophysics*, *38*(3-4), 279–296.

Pollack, H., Hurter, S., & Johnson, J. (1993). Heat flow from the earth's interior: Analysis of the global data set. *Reviews of Geophysics*, *31*(3), 267–280.

Rudnick, R., McDonough, W., & O'Connell, R. (1998). Thermal structure, thickness and composition of continental lithosphere. *Chemical Geology*, *145*(3-4), 395–411.

Sclater, J., & Francheteau, J. (1970). The implications of terrestrial heat flow observations on current tectonic and geochemical models of the crust and upper mantle of the earth. *Geophysical Journal International*, *20*(5), 509–542.

Shapiro, N., & Ritzwoller, M. (2004). Inferring surface heat flux distributions guided by a global seismic model: Particular application to antarctica. *Earth and Planetary Science Letters*, *223*(1-2), 213–224.

Stein, C., & Stein, S. (1992). A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature*, *359*(6391), 123–129.

Stein, C., & Stein, S. (1994). Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *Journal of Geophysical Research: Solid Earth*, *99*(B2), 3081–3095.

Syracuse, E., van Keken, P., & Abers, G. (2010). The global range of subduction zone thermal models. *Physics of the Earth and Planetary Interiors*, *183*(1-2), 73–90.

van Keken, P., Hacker, B., Syracuse, E., & Abers, G. (2011). Subduction factory: 4. Depth-dependent flux of H2O from subducting slabs worldwide. *Journal of Geophysical Research: Solid Earth*, *116*(B1).

van Keken, P., Wada, I., Abers, G., Hacker, B., & Wang, K. (2018). Mafic high-pressure rocks are preferentially exhumed from warm subduction settings. *Geochemistry, Geophysics, Geosystems*, *19*(9), 2934–2961.

Wada, I., & Wang, K. (2009). Common depth of slab-mantle decoupling: Reconciling diversity and uniformity of subduction zones. *Geochemistry, Geophysics, Geosystems*, *10*(10).

Wood, S. (2017). *Generalized additive models: An introduction with R*. Chapman and Hall.

Zhu, A., Lu, G., Liu, J., Qin, C., & Zhou, C. (2018). Spatial prediction based on third law of geography. *Annals of GIS*, *24*(4), 225–240.