Supplementary Material

Paleogeographic uncertainty imposes first-order constraint on understanding Earth's climate evolution

Werner et al. (2026)

Supplementary Information

Differences between Paleogeographic reconstructions

PALEOMAP is the most interpretive of the three. It reconstructs past environments from geological proxies and maps them to broad elevation classes, with low-lying settings assigned modest elevations and collisional belts assigned high elevations; its bathymetry is a mere fixed depth at 4000 m for most of the Phanerozoic (Scotese, 2021)

MERDITH is a full-plate reconstruction with explicit plate boundaries and kinematic consistency through time. Its topography is generated dynamically from tectonic environments such as arcs, collisions, rifts, and large igneous provinces, while its bathymetry is calculated separately from synthetic oceanic crustal age grids using a plate-cooling age-depth model, then merged with the topography (Merdith et al., 2025, preprint)

PANALESIS is also a full-plate model covering 100% of Earth's surface, and it uses a dual-control framework. In that framework, each reconstruction is constrained both by present-day geological evidence moved back to its past position and by dynamic coherence with the preceding

reconstruction, which then controls the next one. Its topography is generated from plate-model features through TopoChronia (Franziskakis et al., 2025) by converting geological setting polylines into points with assigned elevation values, while ocean depths are derived from mid-ocean ridges and isochrons using plate-cooling equations, with sediment loading and water-load subsidence corrections applied separately.

In short, PALEOMAP is the most semi-quantitative, MERDITH is the most process-explicit, and PANALESIS combines explicit plate-tectonic control with a dual-control reconstruction workflow for deriving synthetic topography and bathymetry.

Supplementary Methods

GMST calculation and uncertainty bounds

Global mean surface temperature (GMST) for each time-slice was derived from the model's near-surface air-temperature field by first converting the raw monthly output to an annual mean and then applying a cosine-latitude weighting to account for the spherical geometry of the Earth (the weight at each latitude = $\cos \phi$, normalized by the total number of longitudes) before summing over all latitude–longitude cells. This latitude-weighted average was performed for each simulation, and the resulting GMST values (in kelvin) were subsequently converted to degrees Celsius (°C) by subtracting 273.15 K. Because the four hierarchical experiments (continents-only, +topography, +bathymetry, full) each provide an independent estimate of GMST for the same paleogeographic slice, the spread among them was taken as a measure of structural uncertainty due to paleogeographic complexity. To quantify statistical confidence, a bootstrap-resampling procedure was applied across the four experiment tiers: for each of 400 replicates the GMST values at a given

age were randomly drawn with replacement from the available tiers, the mean was recomputed, and a moving-average pre-filter was applied before interpolating the bootstrap series onto a uniform 1 Ma age grid using a piecewise-cubic Hermite interpolant. The 16-th/84-th and 2.5-th/97.5-th percentiles of the 400 bootstrap realizations defined the central 68 % and 95 % uncertainty envelopes that are shown as shaded bands around the smoothed GMST curves. Finally, the inter-reconstruction GMST spread at each age was calculated as the difference between the maximum and minimum GMST across the three paleogeographic families on the common age grid, providing a regime-by-regime metric of paleogeographic-driven temperature variability.

Statistical Analysis

For regime-based statistical analyses, we first computed the reconstruction-induced GMST spread $\Delta\mathrm{GMST}(t)$ at each age t on the common 1 Ma grid as the difference between the maximum and minimum smoothed GMST across the three paleogeographic reconstructions (PANALESIS, MERDITH, PALEOMAP). Ages were then assigned to tectonic regimes using fixed age boundaries (Pre-Pangea: 320–545 Ma; Pangea: 200–320 Ma; Post-Pangea: 0–200 Ma), and ages falling outside these bounds were excluded from regime statistics. Within each regime, we summarized the distribution of $\Delta\mathrm{GMST}$ using the sample size n, mean, median, standard deviation, interquartile range (25th–75th percentiles), and full range (minimum–maximum). To test for differences in $\Delta\mathrm{GMST}$ among regimes, we used a Kruskal–Wallis test (non-parametric ANOVA) across the three regimes, reporting the Kruskal–Wallis statistic (H), two-sided (p)-value, and the associated epsilon-squared effect size ε^2, computed as $\varepsilon^2 = (H - k + 1)/(n - k)$, where k is the number of regimes and n the total number of slices included in the test. Pairwise regime differences were assessed with two-sided Mann–Whitney U tests, for each pair of regimes, and we quantified the effect size

using the rank-biserial correlation, $r_{rb} = 1 - 2U/(n_1 n_2)$), where U is the Mann–Whitney statistic and n_1, n2) the sample sizes in each regime; to control the false discovery rate across pairwise comparisons, we applied the Benjamini–Hochberg procedure to the set of pairwise p-values and report both raw and FDR-adjusted p-values. Because adjacent ages on the 1 Ma grid are not strictly independent, we evaluated the robustness of our conclusions to temporal autocorrelation by "thinning" the series (retaining every 1, 5, 10, or 20 Ma point), recomputing the Kruskal–Wallis test and ε^2 for each thinning step, and confirming that the inferred regime differences and effect sizes are qualitatively unchanged. In addition, we characterized the stability of the reconstruction ranking through time by computing, at each age, which reconstruction is warmest and coldest, the fraction of ages with the strict ordering MERDITH > PALEOMAP > PANALESIS, and Kendall's coefficient of concordance W across reconstructions: we first ranked GMST values within each age (1 = coldest, 3 = warmest), then calculated W from the summed ranks per reconstruction and assessed its significance using a chi-square approximation with $\chi^2 = n(k\text{-}1)W$ and k-1 degrees of freedom.

Regime ΔGMST diagnostics

Pre-Pangea: n=221, mean=5.61 °C, median=5.15 °C, IQR=4.72–6.11 °C, range=4.26–8.49 °C

Pangea: n=120, mean=4.30 °C, median=4.63 °C, IQR=3.01–5.58 °C, range=1.68–6.40 °C

Post-Pangea: n=200, mean=4.37 °C, median=4.34 °C, IQR=2.92–5.99 °C, range=0.76–7.39 °C

Kruskal–Wallis across regimes: H=63.879, p=1.35e-14, ε^2=0.115 (k=3, n=541)

Pairwise Mann–Whitney U (two-sided), BH-FDR corrected; effect size = rank-biserial:

Pre-Pangea vs Pangea:

U=18684.0, p=4.42e-10, p_{BH}=6.63e-10, RBC=-0.409 (nA=221, nB=120)

Pre-Pangea vs Post-Pangea:

U=30945.0, p=1.3e-12, p_BH=3.91e-12, RBC=-0.400 (nA=221, nB=200)

Pangea vs Post-Pangea:

U=11572.0, p=0.594, p_BH=0.594, RBC=0.036 (nA=120, nB=200)

Sensitivity to thinning (addresses autocorrelation by reducing effective n):

thin_step_Ma	n	H	p	ε^2
1	541	63.88	1.345e-14	0.115
5	109	12.77	0.001689	0.1016
10	55	6.908	0.03161	0.09439
20	28	3.469	0.1765	0.05876

Geoindicators

To relate large-scale climate responses to the underlying paleogeography, we compute a set of quantitative geoindicators for each reconstruction and time slice. These metrics and calculations are based on the geo_indicators repository https://github.com/florianfranz/geo_indicators. These metrics summarize land–sea configuration, continental distribution by latitude, shelf extent, topographic relief, and ocean-gateway geometry on the same grid used by the climate model.

All geoindicators are derived from the gridded land–sea mask, surface elevation, and ocean bathymetry fields for PALEOMAP, PANALESIS, and MERDITH, using consistent geometric definitions across reconstructions and through time. The resulting scalar predictors are then used in the statistical dependency analysis linking paleogeography to GMST, land-mean (P-E), and equatorial ocean heat transport.

We group the geoindicators into the following categories.

Global land–sea metrics

- Total land area (m^2) and total ocean area (m^2) computed from the land–sea mask.
- Ocean volume (m^3), obtained by integrating bathymetric depth over all ocean grid cells.
- Land fraction and ocean fraction, expressed as the global area fractions of land and ocean and used in some analyses as “total land area” or “total land fraction”.

Latitudinal land-distribution metrics

To capture where continents reside relative to the circulation and insolation field, we compute land area within fixed latitude bands:

- Tropical land area (m^2) and subtropical land area (m^2).
- Temperate land area (m^2).
- Polar land area (m^2) and polar land percentage, defined as the land fraction within the polar latitude bands.
- Northern-hemisphere land area and southern-hemisphere land area (m^2), characterizing hemispheric asymmetry in continental distribution.

These latitudinal metrics appear in the importance analysis as “tropical”, “subtropical”, “temperate”, “polar”, “northern-hemisphere”, and “southern-hemisphere” land area, and they consistently rank among the most influential predictors for GMST and land (P-E).

Shelf and coastline metrics

Continental margins and shelf seas exert a strong control on ocean circulation and air–sea exchange. We therefore diagnose:

- Continental shelf area (m^2), defined as the area of ocean grid cells within a prescribed shallow-water depth range, and its normalized variant shelf fraction.
- Coastline length (m), computed as the total length of land–sea boundaries on the model grid.

Shelf fraction and coastline length are used both directly as predictors and indirectly in composite metrics such as shelf fraction and land asymmetry in the mutual-information analysis with GMST.

Topographic metrics

To represent the climatic influence of orography and elevated plateaus, we use:

- High-elevation land area (m^2), the area of land grid cells above a prescribed elevation threshold, and a more generic high-altitude area (m^2)
- Total land area (m^2) (as in section 1), which combines with high-elevation area to encode the relative extent of mountainous versus low-lying terrain.

High-elevation land area emerges as an important, though reconstruction-contingent, control particularly for circulation-sensitive metrics such as equatorial ocean heat transport.

Connectivity and gateway metrics

Ocean-gateway geometry influences basin connectivity and meridional heat transport. We quantify this using:

- Minimum ocean-gateway width (km): the narrowest effective connection between major ocean basins, regardless of orientation.
- Minimum east–west (E–W) gateway width (km): the minimum zonal width of any low-latitude ocean passage, a proxy for the potential strength and structure of zonal currents.
- Passage minimum width (km): a more general metric capturing the smallest width of any continuous navigable ocean corridor.

These gateway metrics appear in the importance heatmaps as “minimum ocean-gateway width” and “minimum E–W gateway width” and show particularly high, and reconstruction-dependent, importance for equatorial ocean heat transport.

Tectonic aggregation metrics

Finally, we include simple measures of continental aggregation:

- Number of continents**,** defined as the number of distinct land clusters separated by ocean at the model resolution.

- Derived measures such as land asymmetry**,** constructed from the difference between northern- and southern-hemisphere land area, and used in some sensitivity analyses of GMST.

All geoindicators are computed independently for each reconstruction family and time slice and compiled into predictor tables that are archived alongside the analysis scripts. These tables form the input to the cross-validated statistical models that quantify the predictability of GMST, land (P-E), and equatorial ocean heat transport from paleogeography and underpin the importance analyses shown in Fig. S6.

Predictability of GMST from geography

Predictability of large-scale climate metrics—GMST, land-mean P–E, and OHT_{eq}—was quantified using cross-validated statistical dependency models relating climate responses to a suite of quantitative paleogeographic indicators. These indicators describe continental distribution by latitude, coastline geometry, continental shelf extent, land elevation, ocean basin geometry, and gateway width.

Model skill was evaluated using out-of-sample R^2, computed separately for each paleogeographic reconstruction and for a pooled dataset combining all reconstructions. This comparison distinguishes reconstruction-specific predictability from signals that are robust across alternative geographic realizations.

Predictor importance was assessed using permutation-based importance scores, defined as the reduction in cross-validated R^2 from random permutation of each predictor. Importance scores were computed independently for each reconstruction to allow direct assessment of reconstruction contingency.

To identify robust versus reconstruction-contingent geographic controls, we analyzed (i) the mean permutation importance of each predictor across reconstructions and (ii) the coefficient of variation (CV = standard deviation / mean) of importance across reconstructions. Pairwise Spearman rank correlations between reconstruction-specific importance vectors were used to quantify agreement in inferred controls. Only predictors with positive median importance were retained to avoid spurious contributions from noise-dominated variables.

3. Supplementary Tables

PlaSim-GENIE parameterization

Supplementary Table 1: PlaSim–GENIE configuration and parameter values used in the paleogeographic ensemble experiments. The table lists XML parameter names, numerical values, and brief functional descriptions for (i) CONTROL settings that define integration length, time stepping, and coupling frequencies; (ii) the GOLDSTEIN ocean module, including diffusivities, drag, equation-of-state and mixing parameters; (iii) the GOLDSTEINSEAICE module, including sea-ice diffusivity and thickness/area thresholds; and (iv) the PLASIM atmosphere–land component, including radiation coefficients, dissipation and diffusion parameters, CO_2 concentration, solar constant, and surface albedo settings. These values define the common baseline model configuration applied to all simulations reported in this study.

SECTION	PARAMETER	VALUE (AS IN XML)	PURPOSE / DESCRIPTION
CONTROL	koverall_total	34 560 000	Total number of atmospheric timesteps (defines simulation length).

	genie_timestep	2739.375 s	Length of one atmospheric timestep (derived from a 365.25 day year).
	par_data_save_slice_n	180	Number of timesteps per output sub-interval (monthly-like saving).
	kocn_loop	16	Gear ratio: atmosphere timesteps per ocean timestep (45 min × 16 = 12 h).
	ksic_loop	16	Gear ratio for sea-ice timesteps (same 12 h coupling).
	conv_kocn_katchem	10	Coupling interval for ocean ↔ atmospheric chemistry.
	conv_kocn_kbiogem	10	Coupling interval for ocean ↔ biogeochemistry.
	conv_kocn_krokgem	10	Coupling interval for ocean ↔ rock-weathering.
	conv_kocn_ksedgem	180	Coupling interval for ocean ↔ sediment module.
	par_grid_lon_offset	–2.8125	Longitudinal offset for the GENIE grid.
	atm_select (indices 3,4,5,6,18,19)	.true.	Activate specific atmospheric diagnostics/components.

	ocn_select (indices 3,4,5,8,10,12,15,16,17,20,45,46)	.true.	Activate chosen oceanic processes/modules.
	sed_select (indices 3,4,5,8,14,15,16,22,32,33,34,36)	.true.	Enable selected sediment-transport tracers.
GOLDSTEIN (OCEAN)	debug_loop	.true.	Turn on diagnostic output for debugging.
	saln0	34.9 psu	Reference salinity for the ocean.
	igrid	2	Flag for T21 latitude spacing.
	yearlen	360	Length of a model year in days (360-day calendar).
	nyear	720	Number of ocean timesteps per model year (12-h step).
	iwstp	72 000	Restart output frequency (every 100 yr).
	ianav	72 000	3-D field output frequency (every 10 yr).
	itstp	72 000	Integral diagnostics output frequency (every 1 yr).
	npstp	72 000	Screen diagnostics output frequency (every 1 yr).
	world	0155pmCO	Identifier for the set of boundary-condition files (land-sea mask, bathymetry, paths).

	tdatafile	pl16_tempann.silo	Observational ocean temperature climatology file.
	sdatafile	pl16_saliann.silo	Observational ocean salinity climatology file.
	temp1 / temp0	0.0 °C	Initial ocean temperature for NH/SH (0 °C indicates use of observational init).
	diff(1)	2005.240135 $m^2 s^{-1}$	Isopycnal (horizontal) diffusivity – tuned parameter OHD.
	diff(2)	0.000135386 $m^2 s^{-1}$	Diapycnal (vertical) diffusivity – tuned parameter OVD.
	adrag	2.554625913 days	Inverse frictional drag coefficient – tuned.
	scf	2.44653955	Wind-stress scaling factor – tuned.
	ediffpow1	1.077402066	Depth-dependence exponent for diapycnal diffusivity – tuned.
	ieos	1	Enable thermobaricity in the equation of state (1 = on).
	iediff	1	Enable stratification-dependent mixing (1 = on).
GOLDSTEINSEAICE	debug_loop	.true.	Debug output for sea-ice module.
	igrid	2	Same T21 grid flag as ocean.

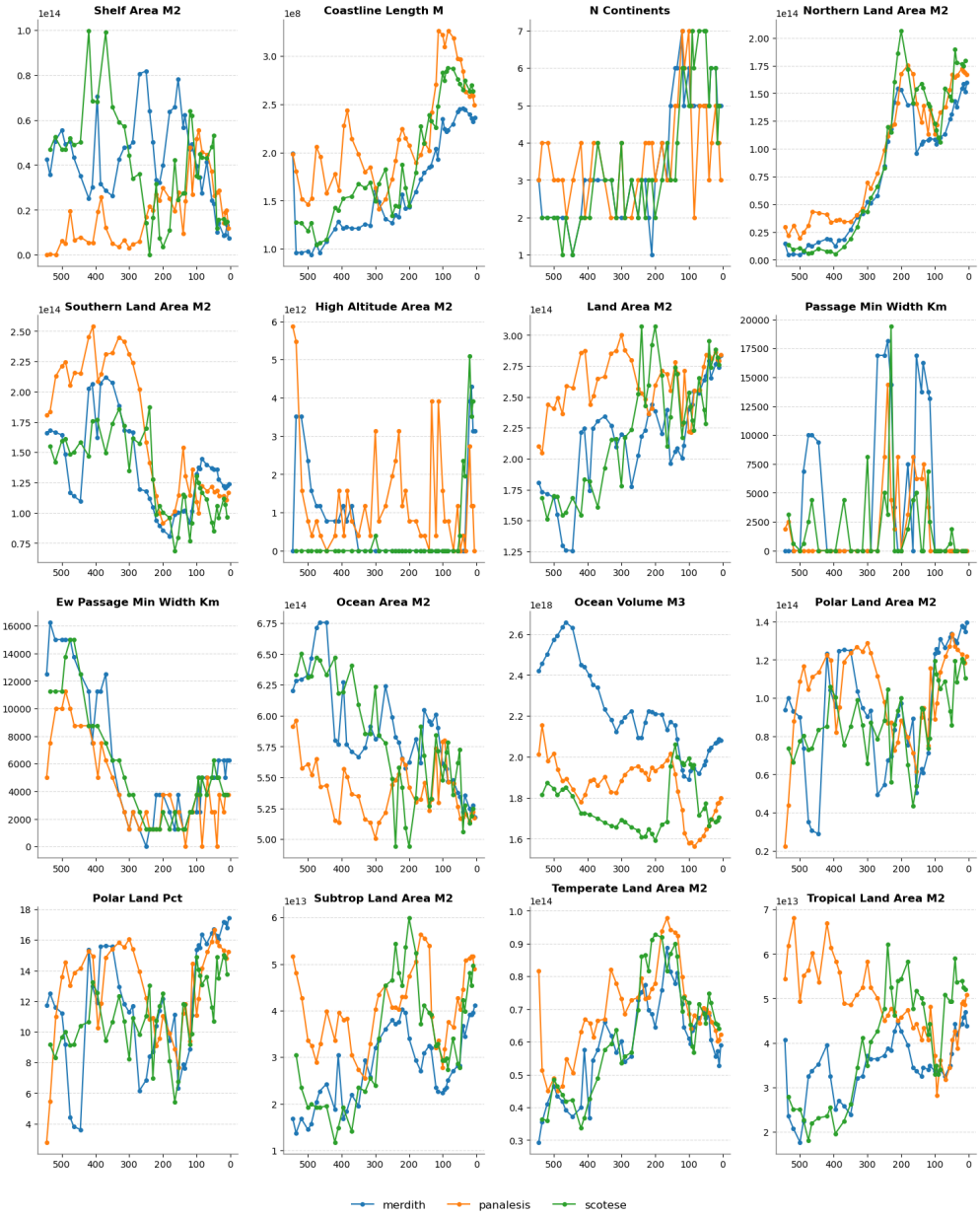
	nyear	720	Sea-ice timestep frequency (12 h).
	iwstp / ianav / npstp / itstp	72 000	Output frequencies matching ocean module.
	world	0155pmCO	Same boundary-condition label as ocean.
	diffsic	15000 $m^2 s^{-1}$	Sea-ice diffusivity (non-default SID).
	par_sich_thresh	30.0 m	Maximum sea-ice thickness allowed for advection into a cell (prevents non-convergent growth).
	par_sica_thresh	1.0 %	Fractional sea-ice cover threshold that stops further sea-ice addition by advection.
PLASIM (ATMOSPHERE-LAND)	ngear	1	Gear ratio for atmospheric internal loops (no sub-stepping).
	ngear_years_plasim	1	Years per gear cycle (baseline).
	ngear_multiple	10	Multiplier for diagnostic output frequency.
	nafter	1 152 000	Number of atmospheric timesteps between averaged output (≈10 yr).
	albgmax	0.8	Maximum surface albedo (e.g., for snow/ice).

	qthresh	0.1	Moisture-content threshold for photosynthesis (disabled in this study).
	scale_apm	0	Scaling factor for Atlantic-Pacific moisture flux adjustment (tuned).
	co2	1120 ppm	Atmospheric CO_2 concentration (fixed for all runs).
	albseamax	0.449923169	Latitude-dependent ocean albedo parameter.
	gsol0	1365	Incoming solar radiation
	tdissd	0.012448584	Dissipation coefficient for zonal wind component.
	tdissz	0.046267461	Dissipation coefficient for meridional wind component.
	tdisst	1.03201731	Dissipation coefficient for temperature tendency.
	tdissq	0.061880547	Dissipation coefficient for specific humidity tendency.
	vdiff_lamm	12.95755067	Vertical diffusion coefficient for momentum.
	tswr1	0.324030081	Short-wave radiation scaling factor (band 1).
	tswr2	0.032974609	Short-wave radiation scaling factor (band 2).

	acllwr	0.501520425	Long-wave radiation scaling factor.
	th2oc	0.023572382	Temperature-to-ocean coupling coefficient.
	rcritmin	0.948665442	Minimum critical relative humidity for cloud formation.
	gamma	0.007991045	Convective adjustment strength parameter.
	k17	0.842488946	Tuned coefficient for cloud-radiative feedback.
	k18	1.204×10^{-7}	Tuned coefficient for cloud optical depth.
	k26	2.41966×10^{-9}	Small-scale turbulence parameter.
	k32	218.3556628	Surface drag coefficient.
BUILD / COMPILATION	IGCMATMOSDP / GENIEDP	TRUE	Enable double-precision arithmetic for atmosphere and GENIE cores.
	GENIENX / GENIENY	64 / 32	Atmospheric horizontal resolution (64 × 32 grid).
	GOLDSTEINMAXISLES	15	Maximum number of islands allowed in the sea-ice/ ocean mask.
	GOLDSTEINNLONS / GOLDSTEINNLATS	64 / 32	Ocean horizontal resolution (64 × 32).
	GOLDSTEINNLEVS	16	Number of vertical ocean levels.

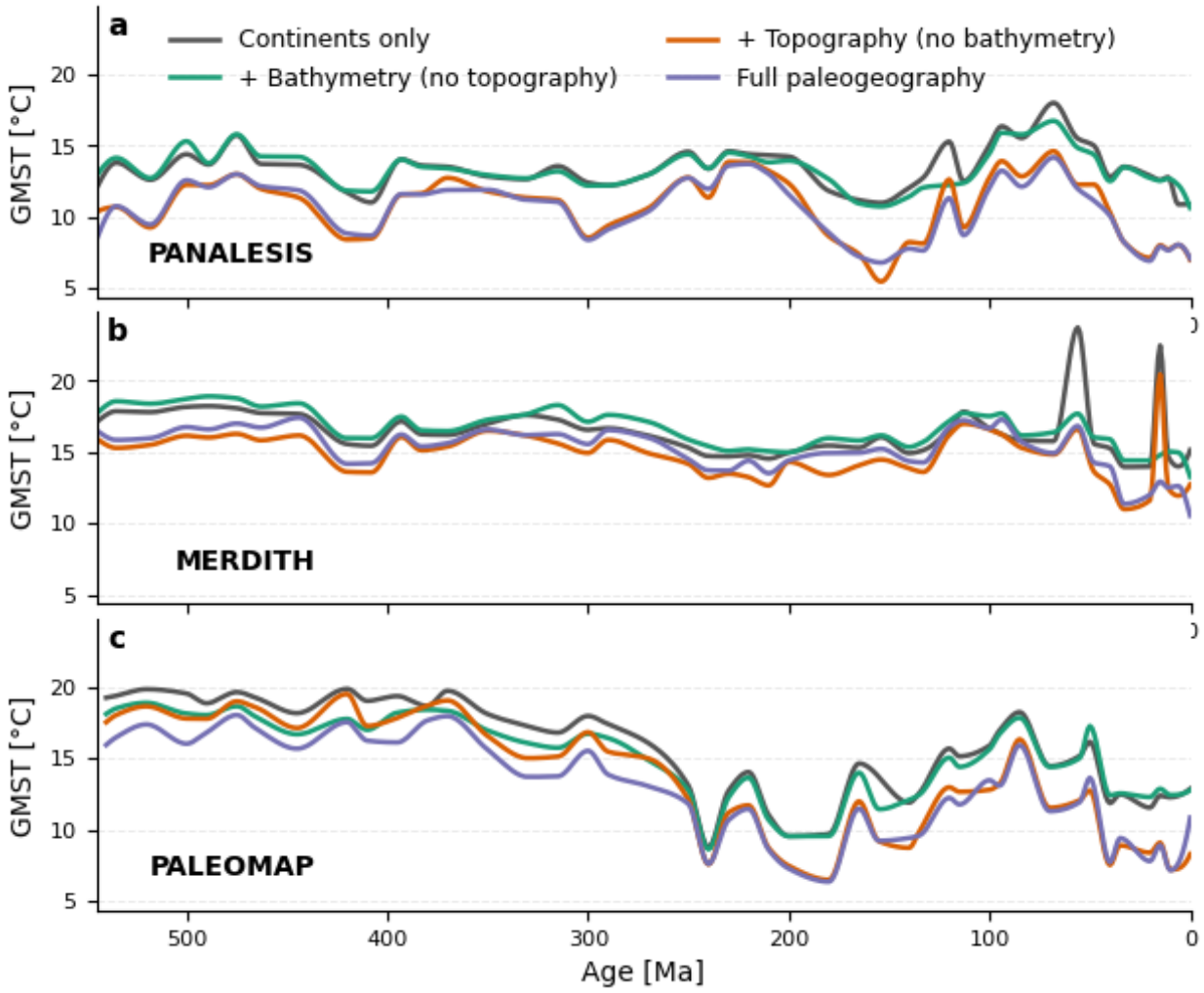
	SEDGEMNLONS / SEDGEMNLATS	64 / 32	Sediment-GEM horizontal resolution.
	ROKGEMNLONS / ROKGEMNLATS	64 / 32	Rock-GEM horizontal resolution.

Supplementary Figures

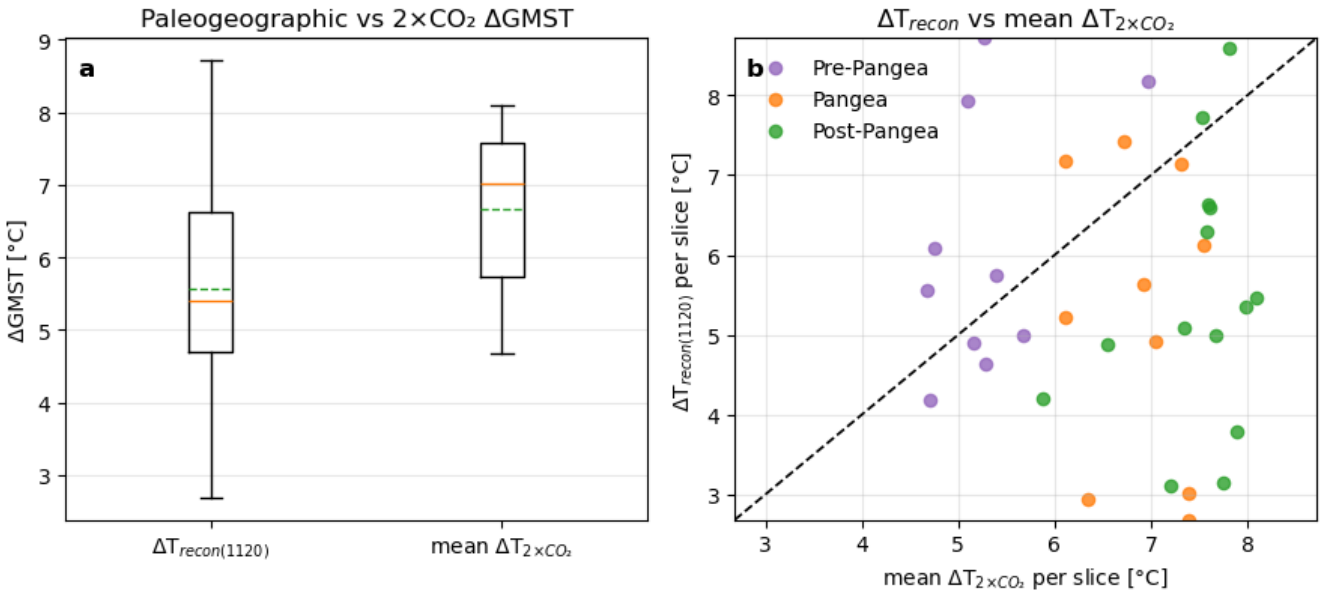


201

Supplementary Fig. S1: Time evolution of geographic indicators across Phanerozoic paleogeographic reconstructions. Sixteen indices describing global land–sea distribution and plate configuration are shown as a function of geological time (0–500 Ma), including shelf area, coastline length, number of continents, latitudinal land-area partitions (northern, southern, polar, subtropical, temperate, tropical), high-altitude area, total land and ocean area, ocean volume, and minimum widths of key passages (global and east–west). Each panel displays three reconstructions—MERDITH (blue), PANALESIS (orange), and SCOTESE (green)—computed with the geo_indicators framework (https://github.com/florianfranz/geo_indicators), illustrating both common trends and reconstruction-dependent differences in large-scale geography relevant for climate modeling.



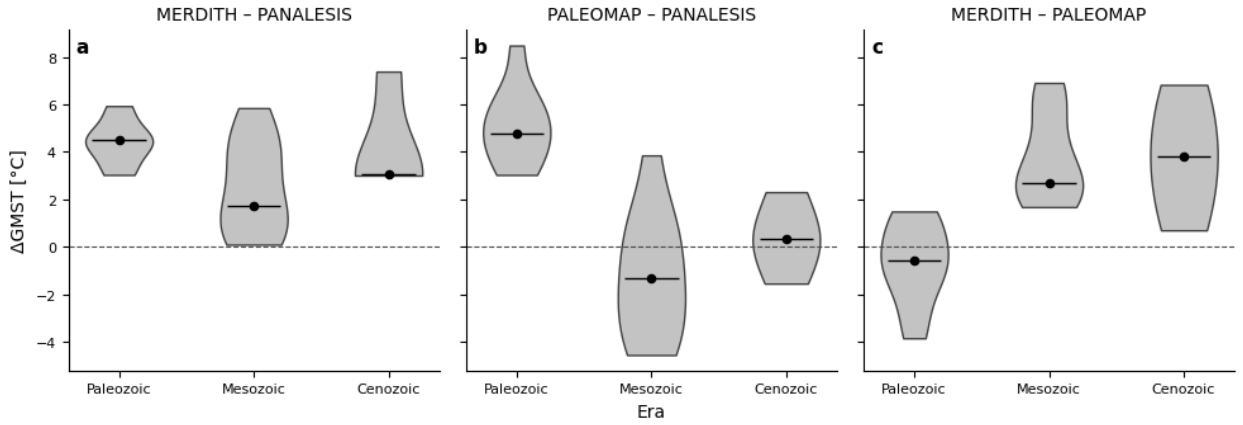
Supplementary Fig. S2: Global mean surface temperature (GMST) evolution for each paleogeographic reconstruction and geographic complexity tier. Panels show GMST versus age for (a) PANALESIS, (b) MERDITH, and (c) PALEOMAP reconstructions, with four experiment tiers per reconstruction: "continents only" (fixed 100 m land elevation and uniform 4000 m ocean depth), "+ bathymetry (no topography)" (reconstructed ocean bathymetry, flat continents), "+ topography (no bathymetry)" (reconstructed continental topography, uniform ocean depth), and "full paleogeography" (reconstructed land–sea mask, topography, and bathymetry). All simulations use the PlaSim–GENIE Earth system model forced with identical external boundary conditions (solar constant 1365 W m^{-2}, atmospheric (CO_2= 1120) ppm, no interactive vegetation), isolating the influence of paleogeographic complexity on GMST.



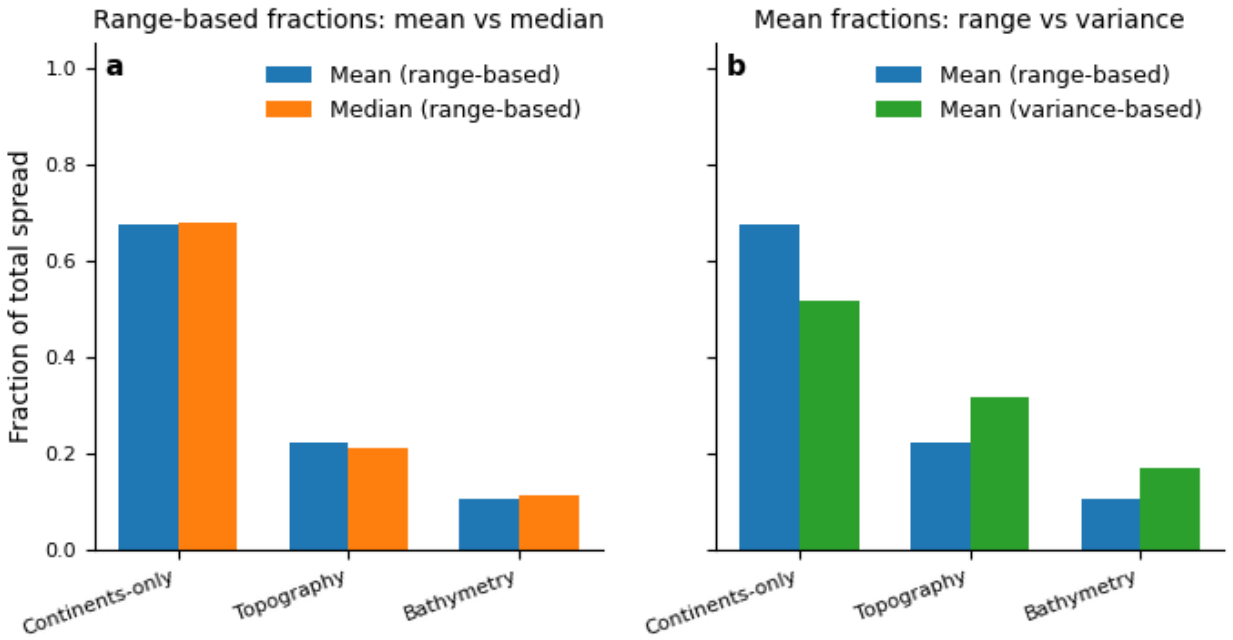
Supplementary Figure S3. Global-mean surface temperature (GMST) variability from paleogeographic uncertainty compared to CO_2 doubling. (a) Boxplots of slice-wise GMST differences for the strict subset of ages where all three reconstructions are available at both 1120 and 2240 ppm. The paleogeographic spread at fixed CO_2, $\Delta T_{recon(1120)}$ (max–min GMST across reconstructions), has a mean of 5.56 °C (median 5.41 °C, interquartile range 4.70–6.62 °C), while the corresponding mean CO_2-doubling response, mean $\Delta T_{2\times CO2}$ ($GMST_{2240}$–$GMST_{1120}$ averaged across reconstructions), has a mean of 6.66 °C (median 7.01 °C, interquartile range 5.73–7.58 °C). **(b)** Strict per-slice comparison of $\Delta T_{recon(1120)}$ versus mean $\Delta T_{2\times CO2}$, color-coded by tectonic regime (Pre-Pangea, Pangea, Post-Pangea); the dashed line marks equality. Across all time slices, the ratio $R=\Delta T_{recon(1120)}/\text{mean } \Delta T_{2\times CO2}$ has a mean and median of 0.86 (IQR 0.68–1.05), and 29.4 % of slices exhibit $\Delta T_{recon(1120)}>\text{mean } \Delta T_{2\times CO2}$. Regime-wise, paleogeographic spread exceeds or matches the CO_2-doubling response in most Pre-Pangea slices (ΔT_{recon} mean = 6.09 °C vs. mean $\Delta T_{2\times CO2}$ = 5.30 °C; 60 % of slices with $\Delta T_{recon}>\text{mean } \Delta T_{2\times CO2}$), while during Pangea (5.22 vs. 6.89 °C; 20 %) and Post-Pangea (5.42 vs. 7.46 °C; 14.3 %) the CO_2-doubling signal tends to dominate but remains of the same order as the paleogeographic uncertainty. Together, this demonstrate that inter-reconstruction paleogeographic differences routinely generate GMST variations comparable in magnitude to those produced by a doubling of atmospheric CO_2.

205

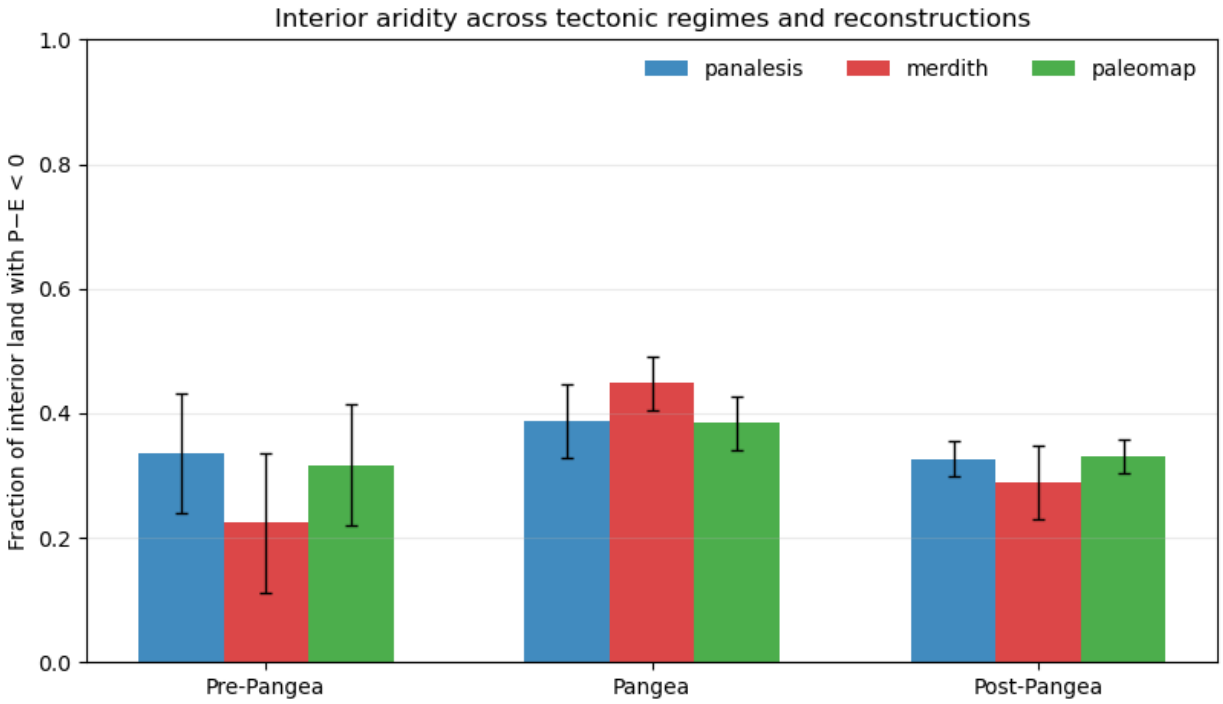
206



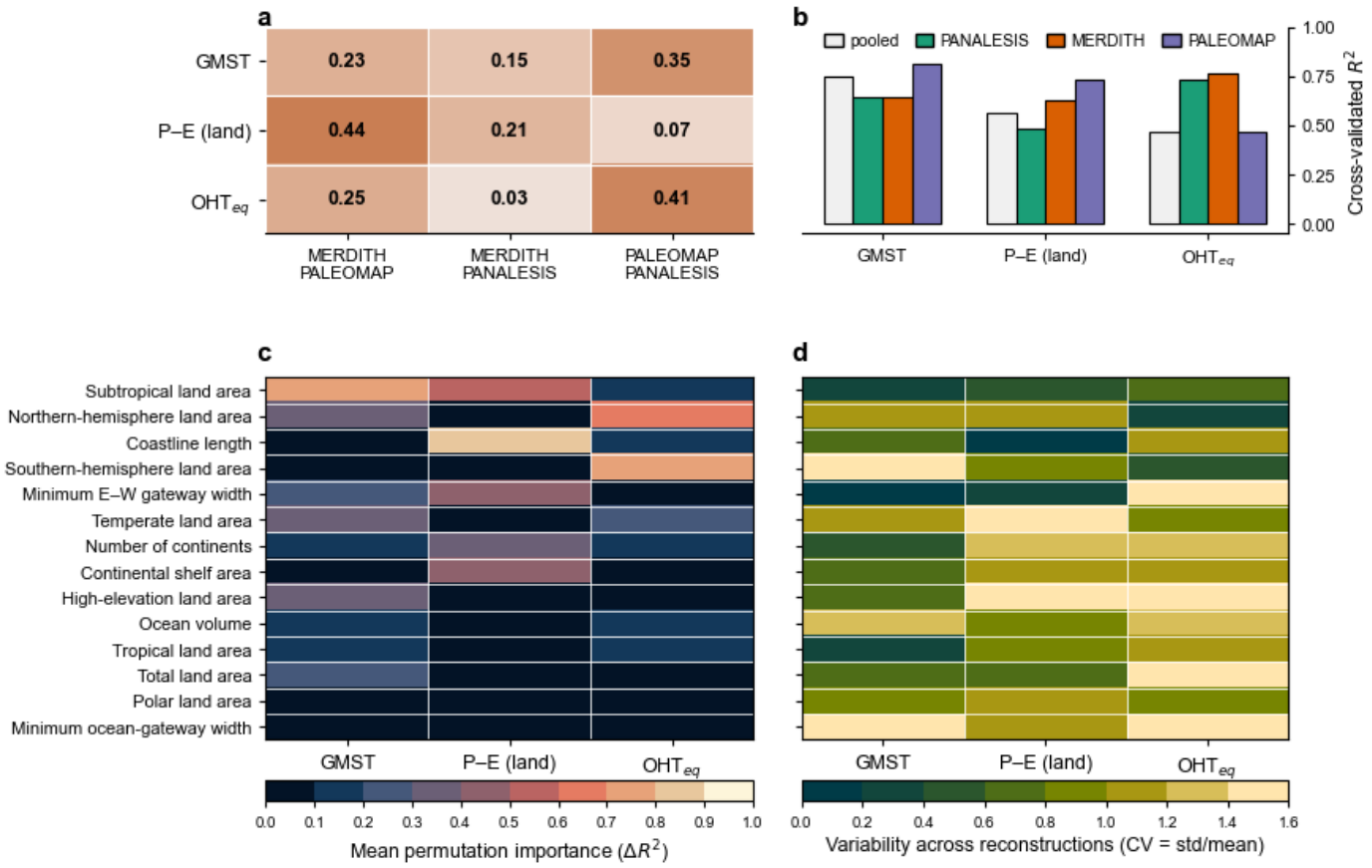
Supplementary Fig. S4: GMST spread diagnostics: pairwise reconstruction differences. Violin plots show the distribution of pairwise differences in global mean surface temperature (ΔGMST) between paleogeographic reconstructions, computed as across-tier means for each time slice and grouped by era (Paleozoic, Mesozoic, Cenozoic). Panels depict **(a)** MERDITH – PANALESIS, **(b)** PALEOMAP – PANALESIS, and **(c)** MERDITH – PALEOMAP. Positive values indicate that the reconstruction listed first is warmer than the second. The era-stratified distributions highlight that MERDITH is systematically warmer and PANALESIS systematically cooler, with PALEOMAP generally intermediate, and demonstrate that the ~5 °C mean spread and occasional >8 °C extremes reported in the main text arise from a broad range of time slices rather than a single outlier interval.



Supplementary Fig. S5: Sensitivity of the paleogeographic spread decomposition to analysis choices. (a) Comparison of mean (dark blue) and median (orange) fractional contributions of continental configuration ("Continents-only"), surface topography ("Topography"), and ocean bathymetry ("Bathymetry") to the total inter-reconstruction GMST spread, using the range across reconstructions as the spread metric and the same hierarchical decomposition as in Fig. 3b. Bars show the fraction of the total decomposed spread attributable to each component, averaged (or median-aggregated) over all matched Phanerozoic time slices. **(b)** Comparison of the mean range-based fractions (dark blue; as in panel a) with mean fractions obtained from an ANOVA-style variance decomposition (green), in which component contributions are defined from changes in the variance across reconstructions between successive tiers of geographic complexity. Across all analysis choices (mean vs median, range- vs variance-based), continental configuration consistently accounts for the majority of reconstruction-induced GMST spread, with topography providing a secondary contribution and bathymetry remaining comparatively minor, demonstrating that the dominance of continents is robust to the choice of diagnostic.

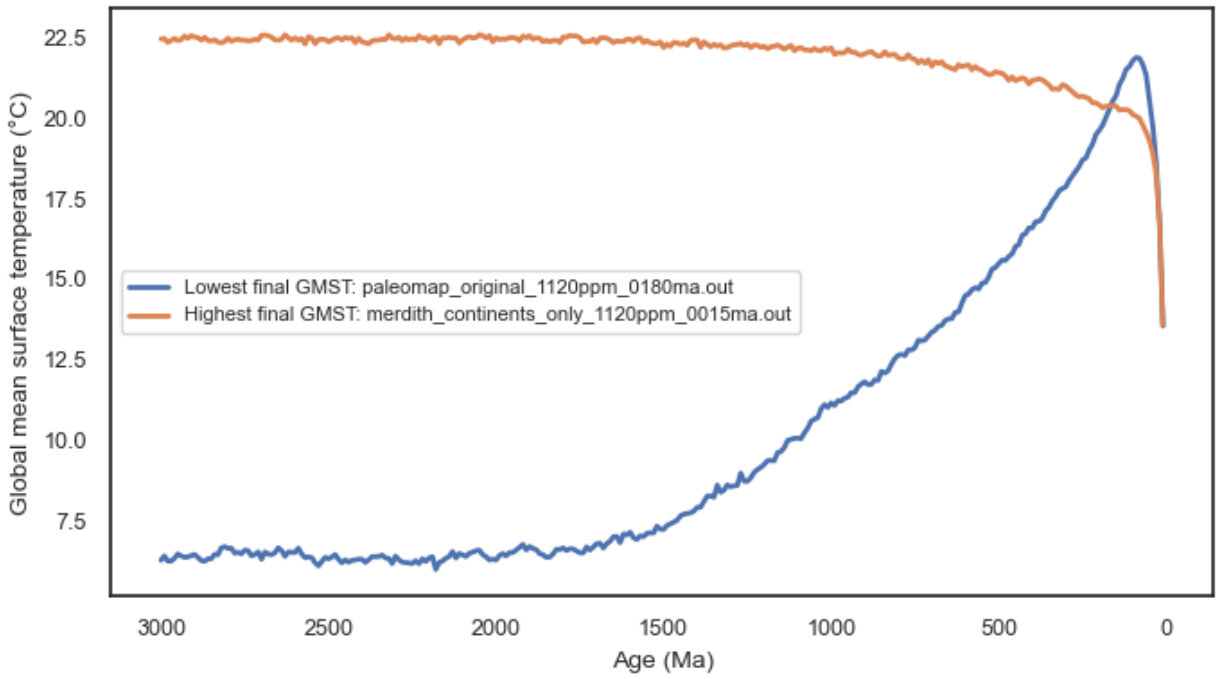


Supplementary Fig. S6: Interior aridity across tectonic regimes and paleogeographic reconstructions. Bars show the mean fraction of interior land area with negative moisture balance ($P - E < 0$) for pre-Pangea (545–320 Ma), Pangea (320–200 Ma), and post-Pangea (200–0 Ma) configurations, computed separately for the PANALESIS (blue), MERDITH (red), and PALEOMAP (green) reconstruction families. Error bars indicate the variability across time slices within each tectonic regime.



Supplementary Fig. S7. Statistical attribution of inter-reconstruction climate differences to large-scale geographic indicators. (a) Cross-validated R^2 values for linear models trained on one paleogeographic reconstruction and evaluated on another, for three climate diagnostics: GMST, continental moisture balance (land-mean P - E), and OHT_{eq}. Each cell summarizes model transferability between reconstruction pairs (MERDITH—PALEOMAP, MERDITH—PANALESIS, PALEOMAP—PANALESIS), with values ranging from ~0.07 to ~0.44, indicating that simple geographic predictors capture only part of the reconstruction-to-reconstruction spread and that the functional relationships between geography and climate are not fully interchangeable across reconstructions. **(b)** Cross-validated R^2 for models fit separately to each reconstruction (PANALESIS, MERDITH, PALEOMAP) and to the pooled multi-reconstruction dataset, for GMST, land-mean P - E, and OHT_{eq}. The pooled model performs comparably to, or slightly better than, most single-reconstruction fits for GMST up to $R^2 \approx 0.75$), reflecting robust large-scale controls of temperature by geography, whereas skill is lower and more reconstruction-dependent for hydrological and ocean-circulation metrics, especially land-mean P - E. **(c)** Mean permutation importance (change in R^2, ΔR^2) for individual geographic indicators when predicting GMST, land-mean P - E, and OHT_{eq}. Higher values indicate predictors whose random permutation most strongly degrades model skill. Subtropical land area, northern-hemisphere land area, and coastline length emerge among the most influential controls across variables, consistent with their roles in setting meridional temperature gradients, continental interiors, and land–sea thermal contrasts. Other indicators, such as the number of continents, minimum east–west gateway width, and high-elevation land area, show more moderate importance, suggesting geographically mediated but secondary impacts on the simulated climate diagnostics. **(d)** Coefficient of variation (CV = standard deviation/mean) of the same indicators across the three reconstructions, again shown separately for GMST, land-mean P - E, and OHT_{eq}. High CV values highlight geographic features that differ most strongly among PANALESIS, MERDITH, and PALEOMAP (e.g., temperate and subtropical land areas, minimum gateway widths), and thus contribute disproportionately to inter-reconstruction climate spread, whereas low-CV indicators (e.g., total land area, ocean volume) are comparatively stable and less responsible for the modeled divergence. Taken together, these panels show that a small set of large-scale geographic metrics both strongly influences climate and varies substantially among reconstructions, providing a compact explanation for why paleogeographic uncertainty alone produces multi-degree differences in Phanerozoic climate simulations.

221



Supplementary Fig. S8: Model spinup for highest and lowest final GMST value. Independent of final temperature value, all simulations reach a quasi-equilibrium state at 3000 years.

222

223 **Supplementary References**

224 Franziskakis, F., Vérard, C., Castelltort, S., & Giuliani, G. (2025). Comparing 545 Million Years
225 of Sea-Level Change: New Insights from the TopoChronia QGIS Plugin. *The International Ar-*
226 *chives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, 111-118.

227 Merdith, A. *et al.* Where were the mountains and how big were they? https://eartharxiv.org/re-
228 pository/object/10989/download/20094/ (2025).

229 Scotese, C. R. (2021). An atlas of Phanerozoic paleogeographic maps: the seas come in and the
230 seas go out. *Annual Review of Earth and Planetary Sciences*, *49*(1), 679-728.

231

232

233