

Supplementary Information for Interpretable Atmospheric Carbon Dioxide Growth Modeling with Hierarchical Station Effects and Transport-Aware Covariates

Austin Lubetkin

Florida Institute of Technology, Melbourne, Florida, USA

Corresponding author: bocaaust@gmail.com

This file contains supporting tables and diagnostic figures. All methodological descriptions are provided in the main manuscript, in accordance with the journal's policy that Supplementary Methods are not permitted.

Supplementary Tables

Table 1: Supplementary Table S1. Abbreviated SHA-256 hashes for principal downloaded and derived inputs. Prefixes identify the analysis files used to generate the supplied results. Full checksums accompany the reviewer data and code archive.

File	SHA-256 prefix
co2_mm_mlo.csv	11c3d997d0ef99e1...
co2_gr_mlo.csv	99d8050252da408b...
meiv2.csv	191cce369d350515...
GLB.Ts+dSST.csv	a2dc9f8e3ef7df02...
Global_Carbon_Budget_2025_v0.6.xlsx	ef7a68cb8f7d140a...
Co2_data_4.zip	e20e5006ac0d18f4...
noaa_gml_surface_flask_monthly_co2_panel.csv	090ebd87faac696e...
CT2022.flux1x1-monthly.nc	b0a107d12232740d...
obspack_co2_1_CARBONTRACKER_CT2022.2023-01-26.zip	42bdf1ff54744deb...
5a2a7ef998f3d512159e4fafb55566b.grib	30afba3e428bec82...
ct2022_station_grid_sampling_weights.csv	abcd835c02a548d3...

Table 2: Supplementary Table S2. Corrected-exponential parameter and block-bootstrap diagnostics. Local standard errors are Gauss–Newton diagnostics. Preferred intervals use 250 successful residual block-bootstrap refits with 12-month blocks.

Parameter or derived quantity	Estimate	Standard error or 95% interval
θ_0	313.959367	0.101428
θ_1	0.878475	0.011561
θ_2	135.774308	28.366887
θ_3	0.011518	0.000994
Annual amplitude A_1 (ppm)	2.8356	[2.7763, 2.8915]
Semiannual amplitude A_2 (ppm)	0.8033	[0.7699, 0.8322]
End growth (ppm yr ⁻¹)	2.7438	[2.6622, 2.8318]
End acceleration (ppm yr ⁻²)	0.0395	[0.0327, 0.0477]

Table 3: Supplementary Table S3. Monthly distributed-lag diagnostics. Covariates were standardized; cumulative effects sum lags 0–6. HAC denotes heteroskedasticity- and autocorrelation-consistent covariance with 12 monthly lags. Tests are two-sided.

Outcome	Covariate	n	R^2	RMSE	Cumulative effect	HAC s.e.	p	Peak lag
Residual	Temperature	563	0.087	0.689	-0.140	0.110	0.204	0
Residual	MEI.v2	563	0.087	0.689	0.047	0.096	0.621	6
Growth anomaly	Temperature	563	0.223	0.548	0.117	0.071	0.100	2
Growth anomaly	MEI.v2	563	0.223	0.548	0.167	0.067	0.013	6

Table 4: Supplementary Table S4. Global Carbon Budget identity recovery. The response is annual atmospheric growth in GtC yr⁻¹ for 1979–2024. The residual is numerical roundoff; this is a data-sign and parsing check, not a forecast or attribution model.

Term	Expected coefficient	Fitted coefficient
Intercept	0	1.78×10^{-15}
Fossil emissions	1	1.0000
Land-use-change emissions	1	1.0000
Ocean sink	-1	-1.0000
Land sink	-1	-1.0000
Cement carbonation sink	-1	-1.0000
Budget imbalance	-1	-1.0000
Temperature anomaly	0	-2.84×10^{-14}
MEI.v2	0	1.47×10^{-15}

Table 5: Supplementary Table S5. Climate-only annual atmospheric-growth diagnostic. The response is Global Carbon Budget atmospheric growth in GtC yr⁻¹ for 1979–2024. HAC covariance used two annual lags; tests are two-sided.

Term	Coefficient	HAC s.e.	t	p
Intercept	2.0680	0.3309	6.25	< 0.001
Temperature anomaly	3.6937	0.4215	8.76	< 0.001
MEI.v2	0.2715	0.1784	1.52	0.128

Table 6: Supplementary Table S6. Nonlinear hierarchical curvature and identifiability diagnostics. LR denotes likelihood ratio; VIF, variance inflation factor.

Diagnostic	Value
Profiled $\hat{\gamma}$	0.021584
95% LR interval	[0.019668, 0.023517]
Curvature-based s.e.	0.000987
Log-likelihood gain versus $\gamma = 0.011518$	53.538
Standardized fixed-design condition number	9.107
R^2 from regressing $q(t; \hat{\gamma})$ on other fixed terms	0.9529
VIF for $q(t; \hat{\gamma})$	21.237
$\text{corr}(t, q)$	0.9762
$\text{corr}(q, \partial q / \partial \gamma)$	0.9995

Table 7: Supplementary Table S7. Shared fixed effects in the nonlinear hierarchical model. Intervals are 95% Wald intervals conditional on the profiled curvature estimate.

Term	Estimate	s.e.	z	95% interval
Intercept	313.478	0.480	653.36	[312.537, 314.418]
t	0.9909	0.0129	76.82	[0.9657, 1.0162]
$q(t; \hat{\gamma})$	25.823	0.139	186.01	[25.551, 26.095]
$\sin(2\pi p)$	2.811	0.258	10.89	[2.305, 3.316]
$\cos(2\pi p)$	1.704	0.284	5.99	[1.146, 2.261]
$\sin(4\pi p)$	-0.981	0.014	-70.17	[-1.009, -0.954]
$\cos(4\pi p)$	0.249	0.014	17.84	[0.222, 0.277]

Table 8: Supplementary Table S8. Random-effect and residual standard deviations. Station slope deviations are reported per decade and per year.

Component	Standard deviation
Station intercept deviation	2.481 ppm
Station slope deviation	1.008 ppm decade ⁻¹ = 0.101 ppm yr ⁻¹
Station annual sine deviation	2.245 ppm
Station annual cosine deviation	2.476 ppm
Observation residual	1.572 ppm

Table 9: Supplementary Table S9. CarbonTracker CT2022 annual flux and anomaly diagnostics for 2000–2020. Positive land sink denotes uptake.

Diagnostic	Value
Mean fossil-fuel flux	8.714 Pg C yr ⁻¹
Mean wildfire flux	2.298 Pg C yr ⁻¹
Mean ocean uptake	2.916 Pg C yr ⁻¹
Mean land-biosphere uptake	3.544 Pg C yr ⁻¹
Largest positive land-sink anomaly	2011, +1.053 Pg C
Weakest land-sink anomaly	2016, -1.454 Pg C
Largest wildfire anomaly	2003, +0.341 Pg C
Monthly grid-state rows	252

Table 10: Supplementary Table S10. Nearest-cell station diagnostic. This local-grid join is descriptive and does not replace the transport-aware station operator.

Model	Observations	Stations	RMSE	R^2
Station effects plus nonlinear trend and seasonality	11,637	54	3.050	0.9502
Add nearest-cell fluxes and global anomaly terms	11,637	54	2.785	0.9585

Table 11: Supplementary Table S11. Selected standardized component-meteorology coefficients. Values are conditional transport-space corrections in ppm and are not causal flux sensitivities. Continuous predictors were standardized except the intercept.

Term	Estimate
Intercept	388.365
Fossil station component	8.115
First-guess forecast component	4.267
Pressure station field	4.408
Wildfire station component	3.475
Biosphere station component	3.990
Ocean station component	2.329
Model-data mismatch s.d.	0.751
Wind-speed station field	1.376
Specific-humidity station field	-3.873
Pressure monthly anomaly	-6.547

Table 12: Supplementary Table S12. Comparison with external concentration references. The 2026 model values are continuation nowcasts; the 2100 SSP values define a scenario range rather than a model confidence interval.

Reference	Model	Year	Model ppm	External value
Met Office Mauna Loa forecast	Pure exponential	2026	429.86	429.4 ± 0.6 ppm
Met Office Mauna Loa forecast	Corrected exponential	2026	429.52	429.4 ± 0.6 ppm
Met Office Mauna Loa forecast	Hierarchy	2026	430.77	429.4 ± 0.6 ppm
Met Office Mauna Loa forecast	Transport process	2026	430.25	429.4 ± 0.6 ppm
SSP concentration range	Pure exponential	2100	833.53	393–1135 ppm
SSP concentration range	Corrected exponential	2100	779.83	393–1135 ppm
SSP concentration range	Hierarchy	2100	908.04	393–1135 ppm
SSP concentration range	Transport process	2100	903.65	393–1135 ppm

Table 13: Supplementary Table S13. Complete model-to-task comparison. RMSE values come from different panels unless rows explicitly share the CT2022 matched panel; cross-panel values must not be treated as a pooled leaderboard.

Model	Evaluation panel	RMSE	Interpretation
Pure exponential	Mauna Loa 2016–2026	0.762	Best compact single-site benchmark
Corrected exponential	Mauna Loa 2016–2026	1.877	Interpretable acceleration estimator
Fixed-only station curve	NOAA stations 2016–2024	3.985	Pooled observation baseline
Nonlinear hierarchy	NOAA stations 2016–2024	1.850	Conditional prediction at represented stations
Nearest-cell diagnostic	Local-grid station panel	2.785	Descriptive grid-cell benchmark
Raw CT2022 posterior simulation	Matched panel 2016–2020	1.888	Physics-based posterior station simulation
CT2022 plus station bias	Matched panel 2016–2020	1.754	Persistent site-offset correction
Components plus meteorology and station effects	Matched panel 2016–2020	1.698	Best common-panel residual correction
Residual-state transition plus station operator	Matched panel 2016–2020	1.756	Competitive reduced state model

Supplementary Figures

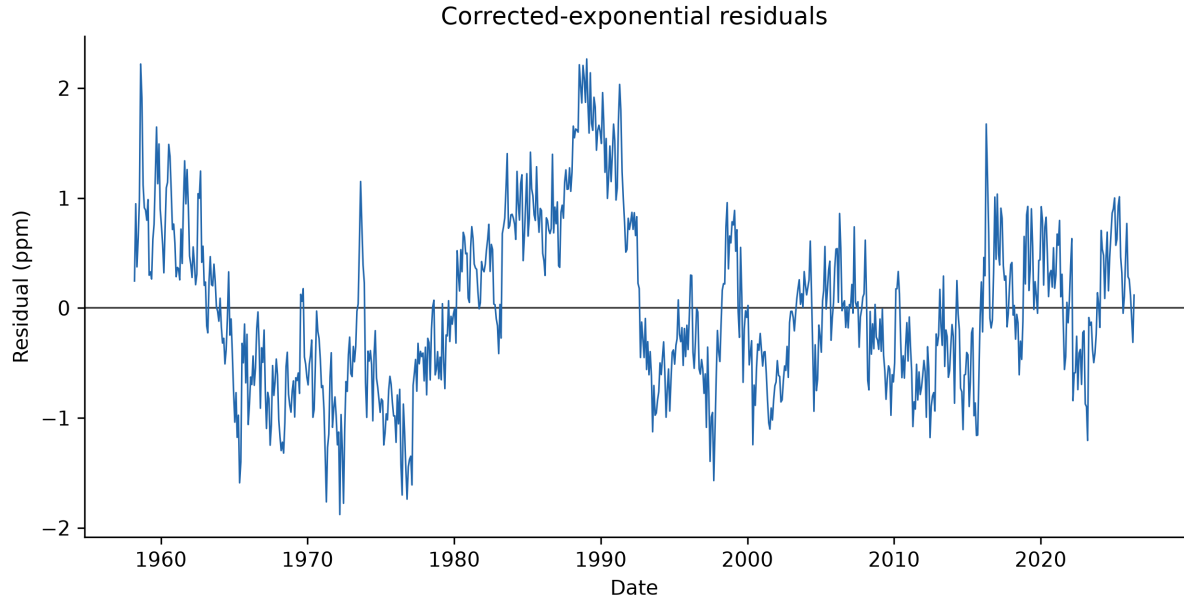


Figure 1: Supplementary Figure S1. Corrected-exponential residuals. Monthly Mauna Loa residuals retain low-frequency and episodic structure after removal of the fitted trend and seasonal harmonics. The horizontal line denotes zero ppm.

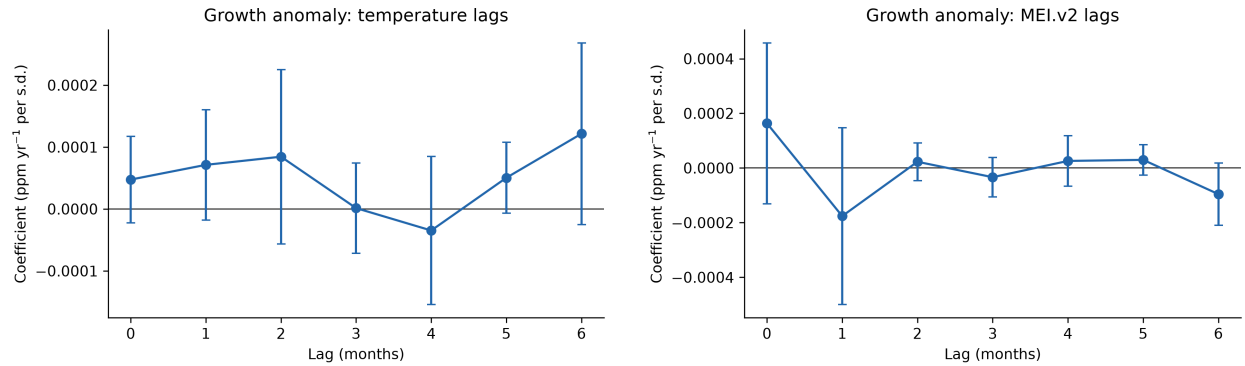


Figure 2: Supplementary Figure S2. Distributed-lag growth-anomaly coefficients. Coefficients for standardized temperature (left) and MEI.v2 (right) at lags 0–6 months. Values are ppm yr^{-1} per one-standard-deviation covariate increase; the horizontal line denotes zero.

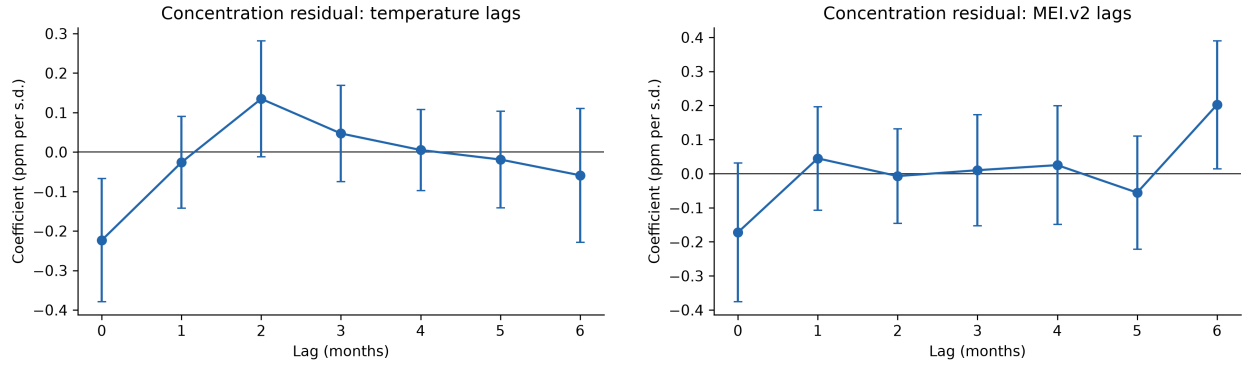


Figure 3: Supplementary Figure S3. Distributed-lag concentration-residual coefficients. Coefficients for standardized temperature (left) and MEI.v2 (right) at lags 0–6 months. Values are ppm per one-standard-deviation covariate increase; the horizontal line denotes zero.

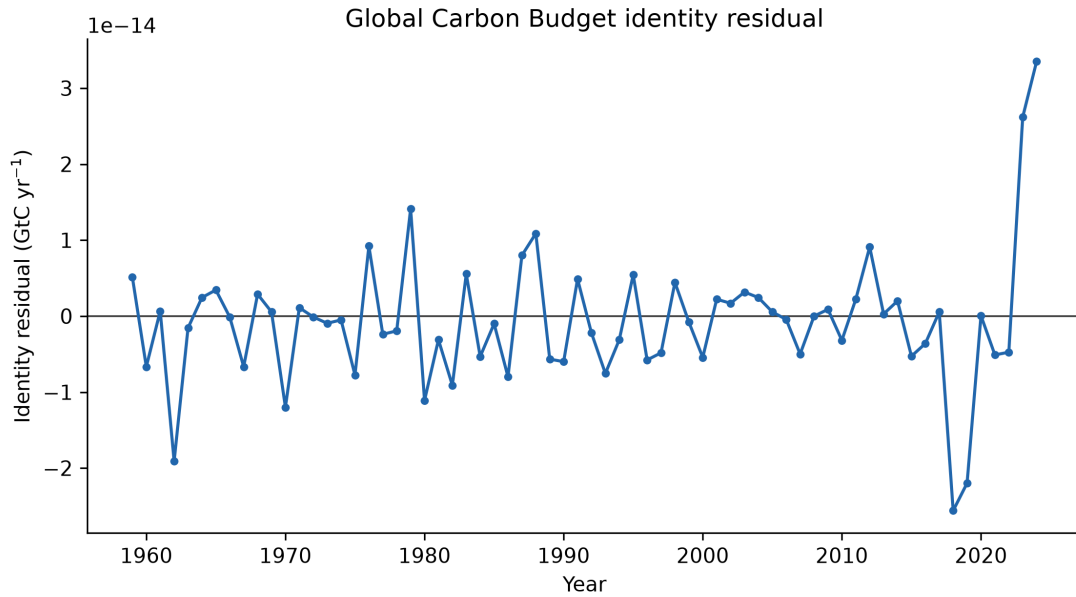


Figure 4: Supplementary Figure S4. Global Carbon Budget identity residual. Residuals are GtC yr⁻¹ and reflect numerical roundoff after all accounting terms, including cement carbonation, are included.

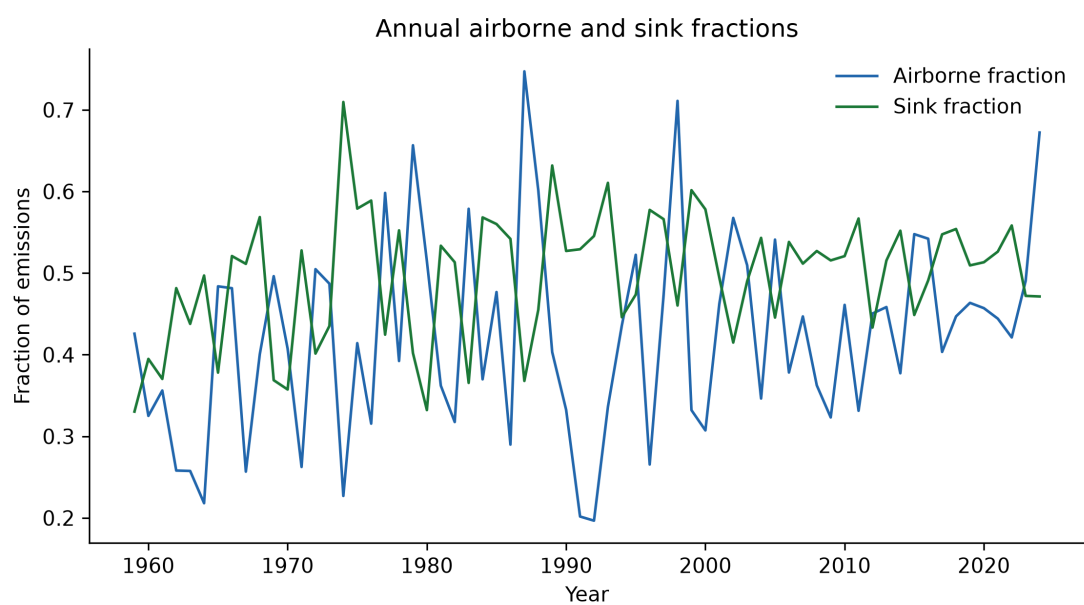


Figure 5: Supplementary Figure S5. Annual airborne and sink fractions. Fractions were calculated from Global Carbon Budget 2025 emissions, atmospheric growth and sink components.

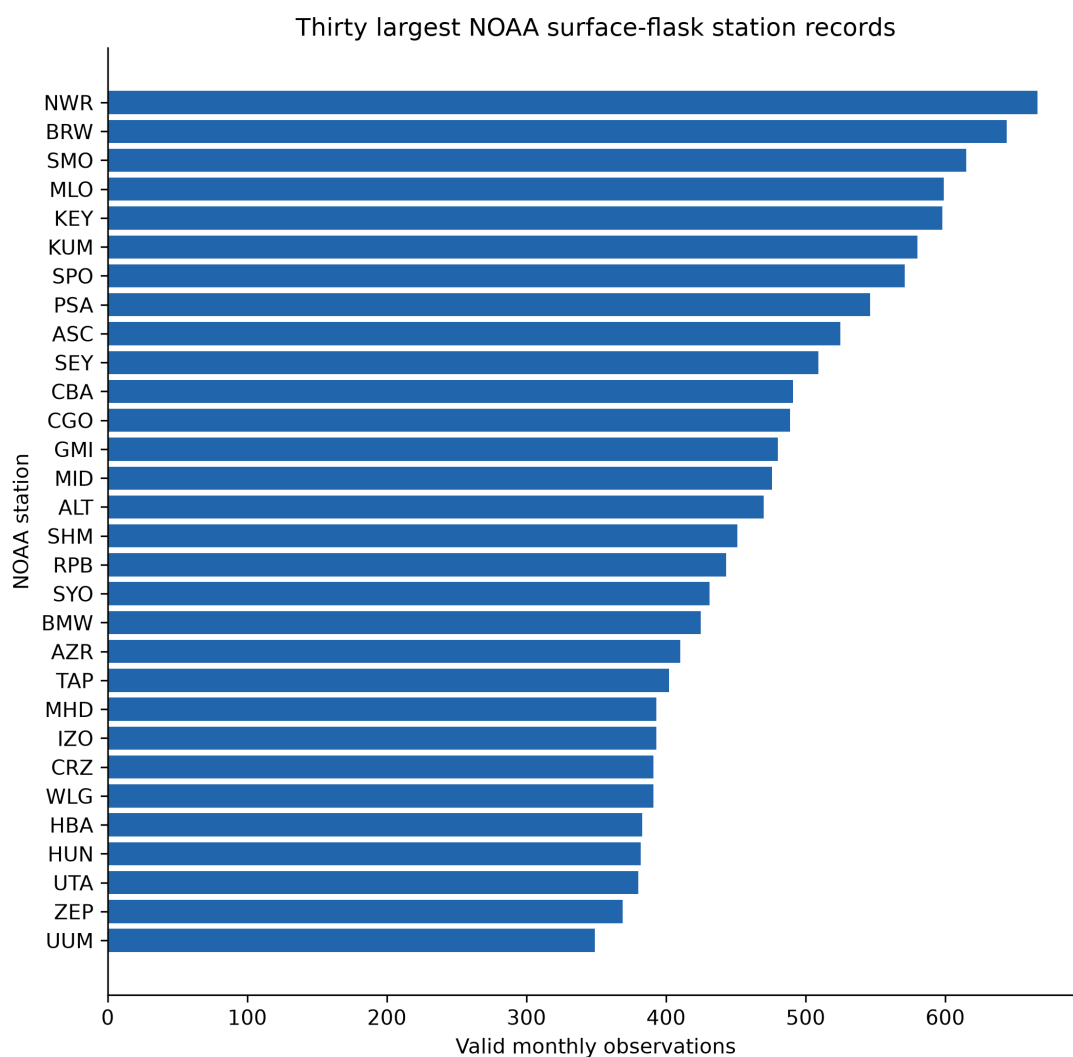


Figure 6: Supplementary Figure S6. NOAA station-record coverage. The thirty largest valid monthly station records before the 120-observation modeling threshold was applied. Station labels are NOAA GML codes.

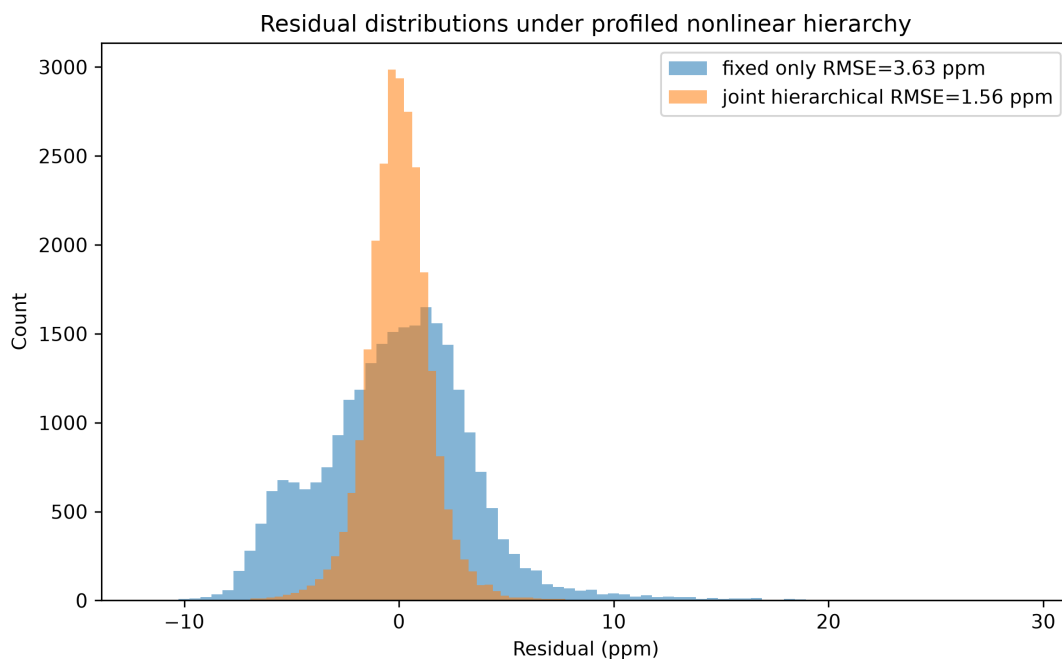


Figure 7: Supplementary Figure S7. Residual distributions before and after partial pooling. Fixed-only and conditional hierarchical residual counts are shown on the same ppm scale.

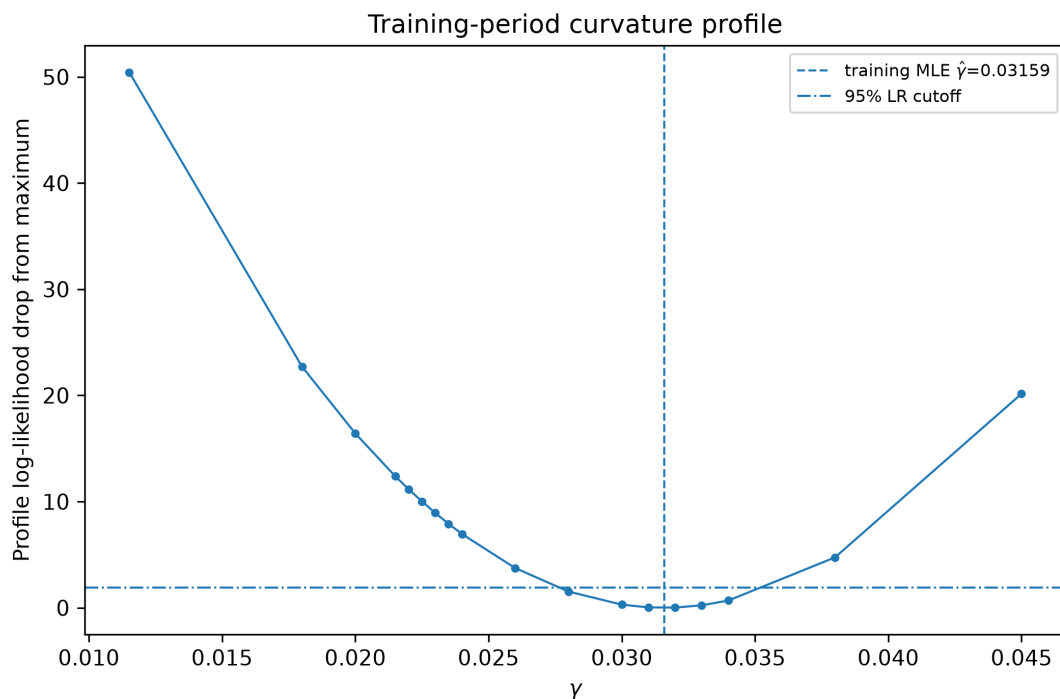


Figure 8: Supplementary Figure S8. Training-period curvature profile. Profile likelihood estimated using pre-2016 observations for the time-ordered station validation. Dashed lines mark the optimum and likelihood-ratio cutoff.

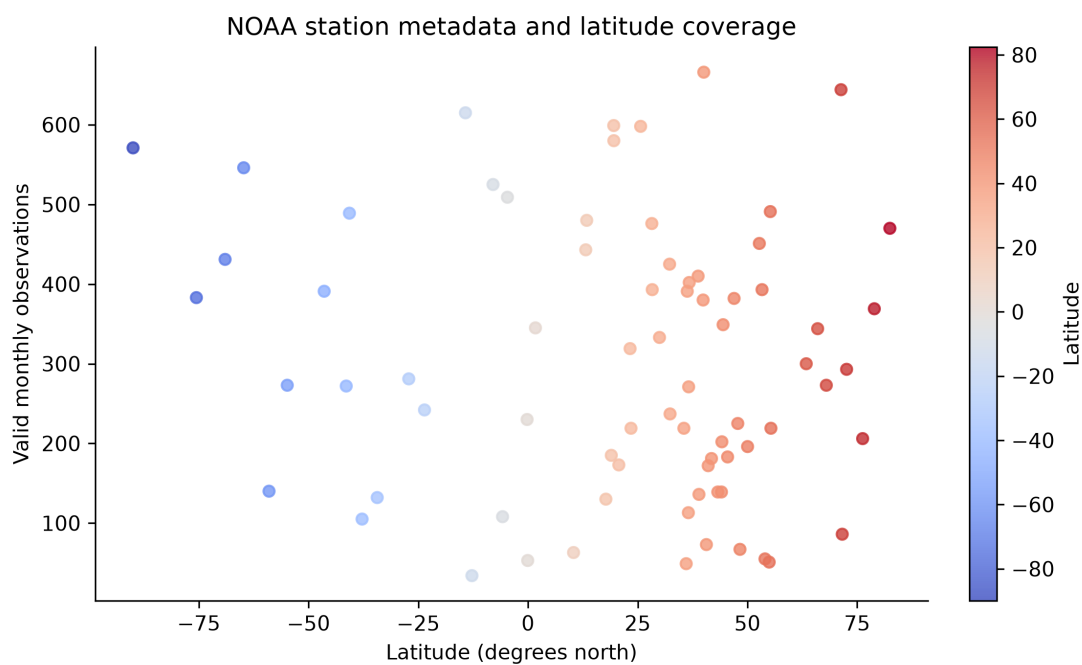


Figure 9: Supplementary Figure S9. Station metadata coverage by latitude. Each point is a geolocated station record; the vertical axis is valid monthly coverage.

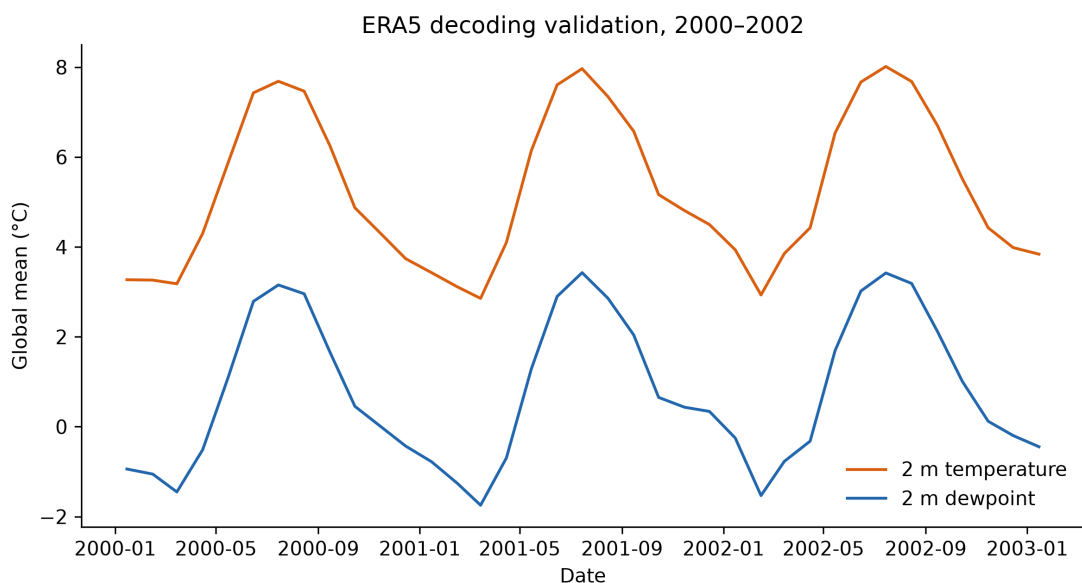


Figure 10: Supplementary Figure S10. ERA5 decoding validation. Global monthly mean 2 m temperature and 2 m dewpoint temperature for complete decoded months from January 2000 through December 2002. Values are plotted in degrees Celsius after conversion from Kelvin.

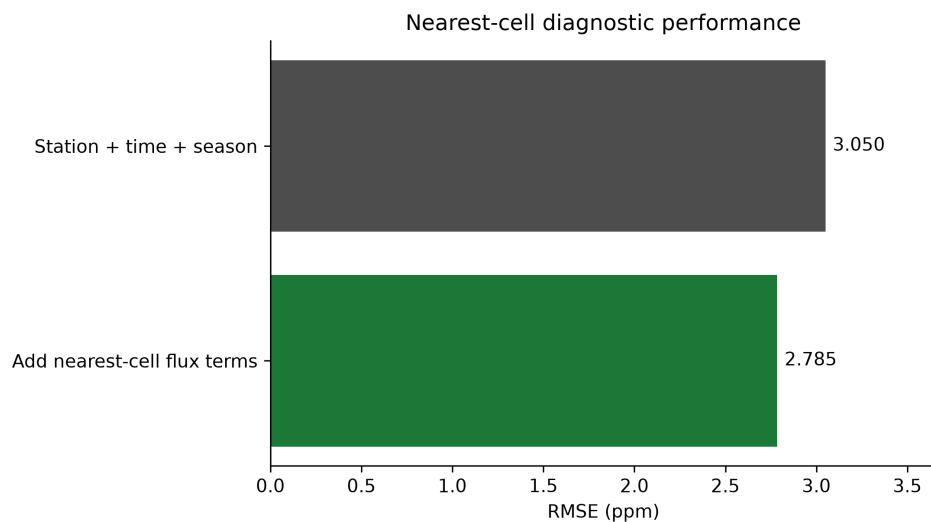


Figure 11: Supplementary Figure S11. Nearest-cell diagnostic RMSE. Adding local CarbonTracker flux and global anomaly terms reduced descriptive station-panel RMSE from 3.050 to 2.785 ppm.

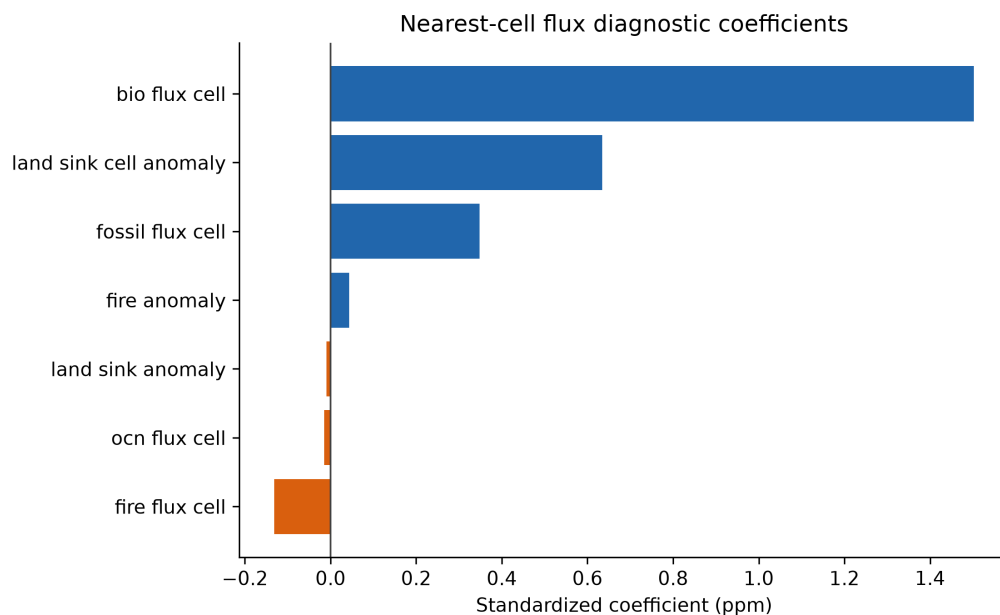


Figure 12: Supplementary Figure S12. Nearest-cell standardized coefficients. Coefficients are ppm per one-standard-deviation covariate change after station effects, nonlinear trend and seasonality. They are descriptive associations and not transport sensitivities.

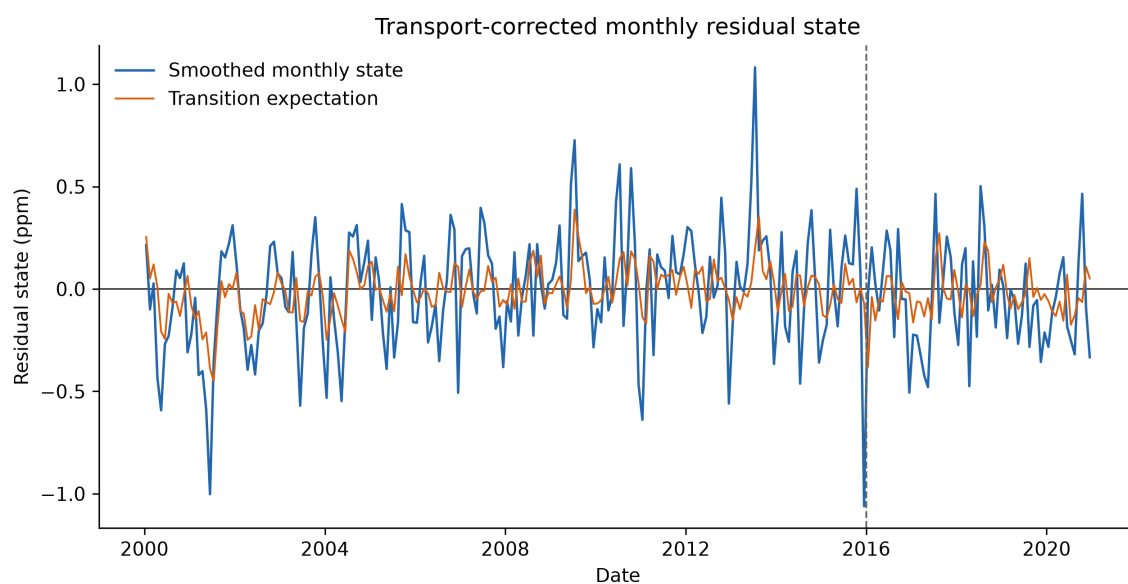


Figure 13: Supplementary Figure S13. Monthly transport-corrected residual state. The residual state follows subtraction of CT2022 posterior station concentration and persistent station bias. The transition expectation was fitted before 2016 and applied through 2020; the vertical dashed line marks the evaluation boundary.